The Pennsylvania State University
The Graduate School
Graduate Program in Cell and Development Biology

REVEALING THE TRANSCRIPTIONAL REGULATORY MECHANISMS
CONTROLLING FLORAL DEVELOPMENT

A Dissertation in
Cell and Developmental Biology

by
Xuan Ma

© 2012 Xuan Ma

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2012
The dissertation of Xuan Ma was reviewed and approved* by the following:

Hong Ma
Distinguished Professor of Biology
Dissertation Advisor
Chair of Committee

Teh-hui Kao
Professor of Biochemistry and Molecular Biology

Claude W. dePamphilis
Professor of Biology

Yinong Yang
Associate Professor of Plant Pathology

Naomi Altman
Professor of Statistics

Zhichun Lai
Chair of the Graduate Program in Cell and Developmental Biology

*Signatures are on file in the Graduate School
ABSTRACT

Flower development, and therefore reproductive fertility, is precisely controlled by a transcriptional and post-transcriptional regulation network. In this network, many pathways, both promoting and inhibiting ones, are responsive to external (environmental) and internal (developmental) signals. Recent advances in microarray technologies have made it possible to decipher mechanisms regulating floral development on a genome-wide scale.

Floral development has been elaborated as a canonical “ABC model”. Previous studies have revealed the molecular mechanisms underlying floral development of many angiosperms, especially in core eudicots and monocots. However, the understanding of basal eudicots, which can serve as an evolutionary link between core eudicots and monocots, remains limited. The microarray experiments on the basal eudicot *Eschscholzia californica* (California poppy) have made it possible to characterize its floral transcriptome. The interwoven-loop design of oligonucleotide microarray allowed the transcriptome of eight tissues (flower buds with pre-meiotic and meiotic cells, developing fruits, leaves and four floral organs at pre-anthesis stages: sepals, petals, stamens and carpels) to be compared simultaneously. Several gene families of transcription factors with putative regulatory function in flower development have been studied, such as MADS-box, AGO, MYB, ZF-HD, ARF, bZIP and bHLH. In addition, comparison between whorl-specific gene expression pattern of *E. californica* and other species suggested its unique evolutionary position.

Genes contributing to flower development have been studied for decades, especially those involved in the reproductive organ development. Several studies had focused on the male reproductive organs by analyzing transcriptome of male sterile mutants. Despite the increasing knowledge, the regulatory mechanisms regulating early anther development are still unclear. Microarray experiments on early anthers of wild type anther and sterile mutant, *aborted*
microspores (ams), have revealed 1368 genes that are differentially expressed in ams during a narrow time period around meiosis. The transcriptome of ams suggested the possible functions of AMS in metabolism, transportation, ubiquitination and stress response. Moreover, comparison between the transcriptome of early anther and those of vegetative tissues identified more anther specific genes which might have important roles in anther development. Combining with previous studies on two other mutants, SPOROCYTELESS (SPL)/NOZZLE (NZZ) and EXCESS MALE SPOROCYTES1 (EMS1), an expanded regulatory network of early anther development has been established.

Though developmental signaling pathways have been studied for years in Arabidopsis, the mechanisms in which environmental signals impact the development of reproductive organs mostly remain unclear. Among those factors affecting reproduction thus production, drought is a very, if not most, prominent one. The yield loss in crops as a result of drought prompted many researchers to elucidate the pathways in which plants adapt to water deficiency. Phenotypical analyses of wild type Arabidopsis in response to different drought conditions have shown that the severity of water deficiency is correlated with the yield loss. And the transcriptome analyses of inflorescences with different water supply at a series of time points provided more information about genes responsive to drought during the process at the molecular level. Moreover, on the basis of known cis-regulatory elements, a possible regulatory network of floral development integrating the external and internal signals has been proposed, providing a series of hypotheses for future experimentation.
TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................... viii

LIST OF TABLES ............................................................................................................... x

ACKNOWLEDGEMENTS ............................................................................................... xi

CHAPTER 1 INTRODUCTION AND LITERATURE REVIEW .............................................. 1

1. Mechanisms Regulating Flower Development ............................................................... 2
   1.1 The ABC model for control of flower development in Arabidopsis ..................... 2
   1.2 The important roles of MADS box genes in regulating flower development
       in various angiosperm plants ................................................................................... 6

2. Anther Development .................................................................................................... 10
   2.1 The anther development stages .............................................................................. 10
   2.2 Transcriptional regulations controlling anther development .................................. 11

3. The Plant Responses to Drought Stress ...................................................................... 19
   3.1 The response of plants under drought stress in Arabidopsis ............................... 19
   3.2 The ABA-dependent transcriptional regulation in response to drought stress
       in Arabidopsis ........................................................................................................ 19
   3.3 The ABA-independent pathways regulating drought response in
       Arabidopsis ............................................................................................................ 21
   3.4 The crosstalk between different pathways in Arabidopsis ................................... 23

4. Recent Progress in Transcriptome Analyses in Plant .............................................. 26
   4.1 Technologies and methods in transcriptomic studies ............................................. 26
   4.2 Application of expression profiling using microarray in plant ......................... 28
   4.3 Tools for microarray analyses .............................................................................. 31

5. Reference ...................................................................................................................... 34

CHAPTER 2 COMPARATIVE TRANSCRIPTOMICS AMONG FLORAL ORGANS
OF THE BASAL EUDICOT ESCHSCHOLZIA CALIFORNICA AS REFERENCE
FOR FLORAL EVOLUTIONARY DEVELOPMENTAL STUDIES ................................ 45

1. Abstract ....................................................................................................................... 46

2. Introduction ................................................................................................................ 47

3. Results and Discussion ............................................................................................. 50
   3.1 Construction and use of a microarray chip for E. californica ............................. 50
   3.2 An overview of differential expression profiling of floral development ............. 53
   3.3 Similar expression pattern of vegetative preferential genes in E. californica
       and in Arabidopsis ............................................................................................... 54
   3.4 Comparing transcriptome profiles at crucial stages of floral development in
       E. californica and in Arabidopsis ........................................................................ 55
   3.5 Identification of putative genes under control of certain genes in ABC
       model ................................................................................................................... 57
   3.6 Expression profiles of members of regulatory gene families .............................. 61

4. Conclusions ............................................................................................................... 65

5. Materials and Methods ............................................................................................. 77
   5.1 Tissue collection and RNA isolation .................................................................... 77
CHAPTER 3 AMS-DEPENDENT AND INDEPENDENT REGULATION OF ANOTHER TRANSCRIPTOME AND COMPARISON WITH THOSE AFFECTED BY OTHER ARABIDOPSIS ANOTHER GENES .................................................. 90

1. Abstract ........................................................................................................... 91
2. Introduction ...................................................................................................... 92
3. Results and Discussion .................................................................................... 96
   3.1 Identification of genes regulated by AMS .................................................... 96
   3.2 AMS affects genes with putative functions in phosphorylation, exocytosis, stress-response and ubiquitin-proteasome pathways during male reproduction .................................................................................. 98
   3.3 Anther-specific or preferential genes were over-represented among genes differentially expressed in the ams mutant ........................................................................................................... 100
   3.4 Genome-wide analysis of gene expression during early anther development by comparing anther transcriptomes of male sterile mutants, spl, ems1, and ams ........................................................................................................ 102
   3.5 SPL and EMS1 might control tapetum development by activating AMS-dependent gene expression ................................................................. 104
   3.6 SPL and EMS1 can regulate early anther development by AMS-independent pathways .................................................................................................................. 105
   3.7 AMS-dependent and independent anther expression of genes encoding transcription factors .............................................................................................................................. 107
   3.8 Transcriptional regulatory network for anther development ......................... 111
4. Conclusion .......................................................................................................... 124
5. Materials and Methods .................................................................................... 125
   5.1 Plant materials .............................................................................................. 125
   5.2 Microarray experiment ................................................................................ 125
   5.3 Microarray analysis to identify differentially expressed genes in anther of mutants ................................................................................................. 125
   5.4 Cis-regulatory element analysis ................................................................... 127
   5.5 Real-time PCR experiments ......................................................................... 127
6. Acknowledgements ........................................................................................... 128
7. References ........................................................................................................ 129

CHAPTER 4 IMPACTS OF DIFFERENT DROUGHT SEVERITIES ON REPRODUCTIVE ORGANS IN ARABIDOPSIS ................................................................ 134

1. Abstract .......................................................................................................... 135
2. Introduction ...................................................................................................... 136
3. Results and Discussion .................................................................................... 139
3.1 Morphological changes in response to different drought severities in *Arabidopsis* ................................................................. 139
3.2 Transcriptome analyses of inflorescences under moderate drought condition ......................................................................... 140
3.3 Comparison between moderate-drought and severe-drought responsive genes ........................................................................ 143
4. Conclusions ................................................................................................................................. 165
5. Materials and Methods .............................................................................................................. 165
5.1 Plant material and stress treatments ..................................................................................... 165
5.2 RNA isolation and hybridization to the ATH1 GeneChip ..................................................... 166
5.3 Microarray Analysis ............................................................................................................... 166
5.4 *Cis*-regulatory element analysis and GO analysis ............................................................. 167
6. Acknowledgements .................................................................................................................... 167
7. References .................................................................................................................................... 168
LIST OF FIGURES

Figure 1-1. The ABC and ABCE model that specify organ identity. ........................................4

Figure 1-2. Transcriptional regulatory network of anther development in Arabidopsis.........15

Figure 1-3. The key LRR-RLKs regulating cell differentiation during early anther development at different anther stages.................................................................18

Figure 1-4. Transcriptional regulatory networks of transcription factors involved in drought response in Arabidopsis. .................................................................25

Figure 2-1. An angiosperm phylogram with illustration of flower structures and the loop design of the E. californica microarray experiments.........................................68

Figure 2-2. Correlation coefficients between signal intensities from four biological replicates of the small floral buds. .................................................................69

Figure 2-3. Venn Diagrams of Genes expressed in reproductive tissues (the same abbreviations were used as in tables). .................................................................70

Figure 2-4. Heat maps and GO annotation pie chart of genes differentially expressed between any two tissues.................................................................71

Figure 2-5. Heat maps of genes preferentially expressed in different tissues. Red color represents high expression while green color represents low expression. .................72

Figure 2-6. The expression levels of MADS transcription factors families ....................73

Figure 2-7. The expression levels of ARGONAUTE, MYB, Zinc-finger, Homeodomain, ARF, bZIP and bHLH families. .................................................................74

Figure 3-1. The expression of genes differentially expressed in ams anthers..............112

Figure 3-2. A pie graph of GO categorization of genes differentially expressed in the ams mutant. .................................................................113

Figure 3-3. Marginal plot of fold changes of expression and number of E-box..............114

Figure 3-4. Venn diagrams of microarray results and previous related study. ..............115

Figure 3-5. Expression distribution of all genes differentially expressed and specific gene families ....................................................................................................116

Figure 3-6. Comparisons between transcriptome information from three mutants ........117

Figure 3-7. AMS-dependent or -independent regulatory model during anther development..118

Figure 3-8. Gene regulatory network of anther development during early stages ..........119
Figure 4-1. Time course of soil moisture during the drought treatment ........................................ 151

Figure 4-2. The whole plants and inflorescences after ten days treatment under six conditions ........................................................................................................................................ 152

Figure 4-3. The accumulated flower numbers were affected by drought stress ......................... 153

Figure 4-4. Drought stress affected the number of siliques ........................................................ 154

Figure 4-5. Drought stress affected the number of siliques ........................................................ 155

Figure 4-6. Venn diagram illustrating the number of genes that are induced under different condition at certain time point ........................................................................................................ 156

Figure 4-7. GO category of biological process for genes up-regulated due to drought stress ........................................................................................................................................ 156

Figure 4-8. GO category of biological process for genes down-regulated due to drought stress ........................................................................................................................................ 157

Figure 4-9. Genes differentially expressed under mild drought condition ................................. 158

Figure 4-10. Hierarchical clustering of NF-Y gene family under drought condition ............... 159

Figure 4-11. Percent of genes in each GO category of molecular functions for genes differentially expressed under both moderate and severe drought condition .................... 160

Figure 4-12. Percent of genes in each GO category of molecular functions for genes differentially expressed only under severe but not moderate drought condition ............ 161

Figure 4-13. Signaling pathways in response to drought stress .................................................. 162
LIST OF TABLES

Table 2-1. Genes preferentially expressed at pre-meiotic, meiotic stage and fruit in poppy................................................................. 75

Table 2-2. Expression levels of putative ABC genes in poppy........................................ 76

Table 3-1. Expression of genes known as anther development related genes. ................. 120

Table 3-2. Genes significantly down-regulated in all three mutants are involved in metabolism of pollen wall formation, including lipid, pectin, lignin and exine. .............. 120

Table 3-3. Genes related to endomembrane system affected by SPL, EMS1 and AMS........ 120

Table 3-4. Transcription factors in SEA-L and SE-L cluster with known or putative function in anther development...................................................... 120

Table 4-1. Expression of genes known as transcription factors........................................ 163

Table 4-2. Expression of genes involved in transportation.............................................. 163

Table 4-3. Expression of genes involved in embryogenesis and reproductive development.............................................................. 164

Table 4-4. Expression of genes involved in known stress responsive genes..................... 164
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Hong Ma, for his constant support and guidance. As the chair of Cell and Developmental Biology, he provided me an invaluable opportunity of studying in the Ma lab at Pennsylvania State University, which will continuously benefit my academic pursuit after graduation. He has also provided a great environment to me to learn not only knowledge and experiment expertise but also his attitude towards work and life. In addition, I want to acknowledge his tireless efforts to help me to accomplish my dissertation projects. It has been a great honor to work with him at both professional and social levels.

I would also like to thank Drs. Teh-hui Kao, Claude dePamphilis, Yinong Yang and Naomi Altman for serving on my dissertation committee and for their insightful, stimulating comments and advices. I am grateful for their knowledge and experience, which has aided my research. In addition, they have always been very positive and I greatly appreciate their contribution on my dissertation.

I want to extend my deepest gratitude to all the present and past members of the Ma lab for their friendship and help, in particular: Zhao Su, Baomin Feng, Dihong Lu, Pingli Lu and Bin Guo, who guided me through the road of molecular biology; Xiaofan Zhou, Xinwei Han, Yazhou Sun and Liye Zhang helped me with my numerous questions in different aspects of bioinformatics. In addition, I would like thank Yi Hu, Yiben Peng, Zhengu Lin, Li Quan, Huihong Guo, Jiange Yang and Jiong Wang for their efforts to make the lab a family.
In addition, I want to thank all the wonderful people, my classmates, Zhen Ren; my dearest lab neighbors, Zi Shi, Yuannian Jiao, Xiaoying Meng and Penglin Sun; and my friends, Luyan He, Nan Zhang, Ying-Chi Chen and Kehui Zhang. Without their supports, it is hard to imagine such a meaningful and pleasant life during these years in PSU. Their friendship is one of the most valuable things that I am taking with me. I also want to express my appreciation to The Huck Institutes secretaries and staff, the people working in the DNA facility. They have supported me with lots of patience and I am very thankful for their incomparable proficiency at what they do.

Most of all I want to thank my family. I am grateful for their love. They have faith in me and support my choice in every situation. I also want to thank my fiancé, Eliseu. He has been actively supportive of my pursuits. I am very grateful to have him in my life.
CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW
1. Mechanisms Regulating Flower Development

1.1 The ABC model for control of flower development in *Arabidopsis*

1.1.1 The proposal of ABC model

In recent years, a lot of attention has been focused on floral organ identity (Coen and Meyerowitz, 1991). It is widely known that the floral meristem in eudicot model plants, after initiation, develops into four whorls, outmost sepals (Whorl 1), petals (Whorl 2), male reproductive stamens (Whorl 3) and innermost female reproductive carpels (Whorl 4), and that this process is mainly regulated by a few key genes (Coen and Meyerowitz, 1991; Litt and Kramer, 2010). These genes, well-known in the “ABC model” of floral organ development, were first identified decades ago in *Arabidopsis* (Figure 1-1A) (Causier et al., 2010). Based on their expression patterns and the organs that they specify, these genes were divided into three main classes: A, B and C function genes. Each class of genes is expressed in the adjacent whorls, and thus controls the development of flower organs by regulating the expression of other genes downstream. The A-function genes are required to properly specify the sepal identity when expressed alone, and to specify petal identity if expressed together with B-function genes. C-function genes alone specify carpel identity, and they specify stamen identity if expressed along with B-function genes (Weigel and Meyerowitz, 1994). This model has been tested by several ectopic expression experiments with different combination of MADS box genes in abnormal domains of the flowers in both *Arabidopsis* and *Antirrhinum*.

In *Arabidopsis*, there are two A-function genes, *APETAL1 (AP1)* and *APETAL2 (AP2)* (Coen and Meyerowitz, 1991; Jofuku et al., 1994), the lack of which leads to defects in the first and second whorls (sepal and petal domains). Moreover, the A-function genes are also
responsible for the specification of the floral meristem. The *PISTILATA (PI)* and *APETALA3 (AP3)* were defined as B-function genes, regulating the development of the second and third whorls (petal and stamen domains) (Krizek and Meyerowitz, 1996; Weigel and Meyerowitz, 1994). Till now, only one gene, *AGAMOUS (AG)*, was characterized as C-function gene in *Arabidopsis*, controlling the third and fourth whorls (stamen and carpel domains) (Mizukami and Ma, 1992). The proteins encoded by these genes were then revealed as MADS-box transcription factors with the exception of AP2. It was subsequently found that there are more than one hundred MADS box genes in *Arabidopsis*, which have been shown to be important regulators in plant development (Parenicova et al., 2003a). Previous studies also revealed that the plant MADS-box proteins can bind to DNA as homo- or hetero-dimers or in higher-order complexes, raising the possibility of interactions between the proteins in the ABC model (Riechmann et al., 1996a).

In *Antirrhinum*, the flowers also have the canonical four-whorl structure but with different unit numbers (five sepals, five petals, four stamens, and two carpels) from *Arabidopsis* (two sepals, four petals, six stamens and one carpel). And mutations of the key genes (*OVU* as A function; *DEF* and *GLO* as B function; *PLE* as C function) lead to similar conversion of flower organ identity as the phenotypes observed in *Arabidopsis* (Coen and Meyerowitz, 1991). The discovery of the shared mechanisms orchestrating flower structures in two highly divergent species suggests that the ABC model might be highly conserved regulating reproductive organ identity through evolution since the emergence of flowering plants.
Figure 0-1. The ABC and ABCE model that specify organ identity.

(A) The ABC model. (B) The ABCE model. The E-function (yellow) specifies the floral domains where the ABC genes function. (C) (A)BC model.
1.1.2 Problems in ABC model and the updated ABCE model

After the ABC model was proposed, it has been shown that the A-, B- and C-function genes are necessary but not sufficient to convert the identity of vegetative organs into floral organs, suggesting that additional genes are required to establish the floral architecture (Krizek and Meyerowitz, 1996; Mizukami and Ma, 1992). The protein-protein interaction experiments that showed the lack of interaction between the B- and C- function gene products also indicated that other proteins might be involved in the processes (Riechmann et al., 1996a; Riechmann et al., 1996b). Therefore, a higher-order protein complex has been proposed and this complex has been demonstrated by in vitro binding experiments, first in Antirrhinum (Egea-Cortines et al., 1999) and then in Arabidopsis (Pelaz et al., 2000). And the identification of the mediators has led to an important modification of the previous model.

In Arabidopsis, genes involved in forming the protein complex other than the ABC genes were then identified as E function genes whose expression is essential for the initiation of the floral meristem organ identity genes (Figure 1-1B) (Pelaz et al., 2000). The E-function genes, which also belong to the MADS box family, can be expressed in all four whorls. Their functional analysis showed that the combination of the A-, B- and E function genes can convert Arabidopsis rosette leaves into petaloid structure, while the combination of the B-, C- and E-function genes can convert them into stamen-like organ (Honma and Goto, 2001; Pelaz et al., 2000; Pelaz et al., 2001). In Arabidopsis, there are four E-function genes known as SEPALLATA1 (SEP1), SEP2, SEP3 and SEP4 (Pelaz et al., 2000; Zahn et al., 2005a). The SEPs can mediate the interaction between products of the A-, B- and C- function genes, in order to form higher-order complexes as tetramers to specify different organ identity; this is called the ABCE model. This updated model has been supported by gel-shift assays which showed better stability of protein complexes and binding affinity to DNA probes with the MADS-box binding sequences (de Folter et al., 2005;
Melzer and Theissen, 2009). A recent study using yeast three-hybrid study in large scale suggested that the SEPs are involved in many protein complexes, especially in the MADS-box complex, functioning as a glue to regulate various developmental processes (Immink et al., 2009).

1.2 The important roles of MADS box genes in regulating flower development in various angiosperm plants

1.2.1 Conservation of genes controlling flower organ identity in angiosperms

Among the estimated 350,000 species of angiosperms, the morphologies of flowers are similar but different from each other in the types of organs (Litt and Kramer, 2010). Though most angiosperms have four whorls of floral organs as in Arabidopsis, some species, especially in basal eudicots and monocots, contain only three types of organs: undifferentiated sepals, stamens and carpels. Variations of flower morphologies can result from novel organs and the fusion of organs. It has been suggested that the duplication of the MADS-box genes played important roles in the origin and diversification of angiosperm flowers (Soltis et al., 2007). Thus, the investigation of floral organs in various species can shed some interesting lights on expending our knowledge of the ABC model.

The degree to which the ABC model is conserved has been assessed in a few flower plants, both in dicots and monocots, including economically important crops such as rice and maize (Ambrose et al., 2000; Causier et al., 2010; Fornara et al., 2003; Litt and Kramer, 2010; Parenicova et al., 2003a; Vandenbussche et al., 2003). Consistent with previous research in Arabidopsis, the homologs of AP3 and PI (B-function genes) are necessary to regulate the development of stamens not only in core eudicots, but also in basal angiosperms, magnoliids and monocots; and homologs of AG (C-function gene) are required for both stamens and carpels.
(Parenicova et al., 2003a; Soltis et al., 2007). The high degree of similarity among species allowed researchers to uncover the secrets of floral development in spite of the extreme variation of flower morphologies. In many species, more than one homolog could be identified and they might function redundantly or differently (Melzer et al., 2010; Zahn et al., 2005a). However, no counterparts of AP1 or AP2 (A-function genes) have been identified in other species even in other core eudicots, suggesting that the A-function genes are less conserved compared with the B- and C-function genes. It is hypothesized that the importance of the A-function genes is limited to a few core eudicots.

1.2.2 The War of whorls

APETALA1(AP1) and APETALA2(AP2) are usually defined as the A-function genes in Arabidopsis; however, mutations in AP1 and AP2 result in various phenotypes besides the predictable loss of sepal and petal, suggesting their multiple functions other than just defining floral organ identity (Litt, 2007). The abnormal development of lateral organs and the existence of novel flowers in the outer whorls indicated the importance of the A-function genes in the establishment of floral meristem identity. This phenomenon is also observed in other species (Kater et al., 2006), especially in Antirrhinum, where the mutant of the orthologue of AP1 results in replacement of flowers by new inflorescences bearing no buds (Huijser et al., 1992). The finding that AP2 plays important roles in both floral and non-floral development also suggested multiple roles of A-function genes.

Moreover, subsequent experiments showed that in the absence of the A-function genes, petals could still be produced in the lack of C-function gene (AGAMOUS) (Bowman et al., 1993) or inflorescence meristem identity gene AGL24 (Yu et al., 2004) or with ectopic expression of SEP3 (Castillejo et al., 2005). In addition, the loss of A-function in outer whorls of Arabidopsis
results in the expansion of the C-function genes; this suggests the antagonism of the A-function genes against the C-function gene (Litt, 2007). The studies in other species are in agreement with the observation in Arabidopsis (Keck et al., 2003).

The spatial partitioning of the A- and C-function genes is vital for normal floral development. To control the expression of the AP2 gene in Arabidopsis, miRNA172 acts as a translational repressor and restrict AP2 expression in the stamen and carpel domain (Chen, 2004). Similar to Arabidopsis, other species also use miRNA to specify organ identity. The mutants of the miRNA-encoding homolog genes, MIRBL of Petunia hybrid and MIRFIS of Antirrhinum majus showed similar homeotic phenotypes as mutant in A-function, suggesting their roles in restricting C-function. Given the ubiquitous expression of these negative regulators of the C-function, scientists strived to explain how the flower specifies the unique domain expression pattern by MIRFIS and MIRBL genes in the inner whorls (Sridhar et al., 2006a; Zahn et al., 2006a). The C-function is repressed in all floral domains (Cartolano et al., 2007; Sridhar et al., 2006b). Though there are differences in models of various species, it is clear that the partition of the A-, B- and C-function is tightly regulated.

1.2.3 The (A)BC model

Because of the lack of known A-function genes in other angiosperm species, a new model has been proposed: only the B- and C-functions are essential to the establishment of organ identity, consistent with their conserved expression patterns (Schwarzsommer et al., 1990). In addition, the B- and C- function genes are expressed within two adjacent whorls and exert their effects after the initiation of floral meristem to determine organ identity. Taken all the factors mentioned above into consideration, it is reasonable to separate the A-function genes from the B- and C-function.
Recently, the “BC” model has been expanded to the “(A)BC” model which incorporated a novel defined (A)-function (Figure 1-1C). The (A)-function plays several important roles in establishing the floral meristem identity, activating and facilitating the B- and C-function genes at later stages and regulating the B- and C-function by establishing their boundaries (Causier et al., 2010). The new definition of the (A)-function includes the original E function genes, which enable the B-and C-functions. The (A)-function might contribute to both floral meristem identity and later floral organ identity. Moreover, there is no requirement for a specific expression pattern of the new (A)-function gene. Thus the (A)-function provide a more general definition and the applicability of the “(A)BC” model has extended to more angiosperm species and can accommodate newly added genes in this flexible model.
2. Anther Development

2.1 The anther development stages

On the basis of the ABC model, many studies have been carried out about the floral organ identity (Irish, 2010). Many genes have been identified to be involved in this process, especially in the development of reproductive organs. In the past decades, many studies have lead to exciting progress on the male reproductive organ, the stamen.

In flowering plants, the stamen consists of a vascular filament and an anther (Ma, 2005). The anther contains somatic and meiotic cells, and the latter ones then produce male haploid spores (microspores) through meiosis. The microspores develop into mature pollen grains (microsporogenesis) and then undergo mitotic cell divisions to form sperm cells (microgametogenesis) (Scott et al., 2004). This process requires support from the adjacent non-reproductive cells, tapetum, middle layer and endothecium, which provide and protect the developing reproductive cells. Therefore, anther development is critical to achieve the success of fertilization (Goldberg et al., 1993a; Ma, 2005; McCormick, 1993).

According to the morphological features, anther development can be grouped into two phases and then be further divided into 14 anther stages (Ma, 2005; Paul M. Sanders, 1999). By the end of the first phase, cell differentiation and meiosis should be completed (anther stages 1 to 8) and the pollen be released during the second phase (anther stages 9-14). At anther stage 1, the stamen primordium has 3 layers, L1-L3 from surface to interior. The L1 cells later become epidermis and the L3 cells give rise to the vascular and connective tissues. The L2 cells develop into archesporial cells at stage 2 and then differentiate into parietal cells and primary sporogenous cells. The former ones then divide into two layers of secondary parietal cells, while the latter develop into microsporocytes. Through multiple rounds of divisions and differentiations, the
anther establishes a characteristic four-lobed structure at stage 5. Each lobe consists of three layers, endothecium, middle layer and tapetum, derived from parietal cells surrounding reproductive cells. Upon the establishment of anther morphology, primary sporogenous cells (reproductive cells) undergo meiosis at around stage 5-7. After meiosis, tetrads containing four microspores are formed at stage 7. Then the callose walls of the tetrads degenerate and microspores are released. In phase 2, the microspores undergo mitosis and develop into mature tricellular pollen grains by the end of stages 12. Meanwhile, pollen wall materials are deposited from the tapetum layer. After the degeneration of tapetum, the mature pollen is released and is able to start fertilization.

2.2 Transcriptional regulations controlling anther development

Recent studies at the molecular level suggested that thousands of genes are expressed in anther, many of which are proposed as anther-specific genes (Goldberg et al., 1993b; Scott et al., 2004). It is of interest to understand the pathways through which these genes are regulated. The male reproductive development is predominantly determined by transcriptional regulations and cell-cell communications at early stages (Coen, 2001b; Ge et al., 2010; Ma, 2005; Paul M. Sanders, 1999). Many gene families, including MADS, MYB, bHLH, LRR-RLK gene families, are involved in the process, controlling stamen initiation and identity, anther cell layer formation and function, and meiosis at different stages.

2.2.1 Transcription factors involved in anther cell layer specification

The SPOROCYTELESS (SPL)/NOZZLE (NZZ) gene is one of the earliest genes that determine anther cell fate (Schiefthaler et al., 1999b; Yang et al., 1999b), directly activated by
AG, a C function gene in the ABC model, through its binding consensus downstream of the SPL coding region (Fig. 2) (Ito et al., 2004a; Jack, 2004; Yanofsky et al., 1990a). SPL/NZZ is expressed during anther stage 3-5 and the loss of SPL leads to the failure of differentiation of parietal and sporogenous cells, and consequently the inability to form the anther wall or microsporocytes (Yang et al., 2003b). On the basis of SPL/NZZ protein sequence and its subcellular localization, SPL/NZZ is predicted to be a key transcription factor which controls male fertility. Further study suggested that it is sufficient to induce sporogenesis in the petaloid floral organs in the absence of the C-function gene, AG. The transcriptome analyses of spl/nzz anthers suggested that many genes are regulated by SPL/NZZ directly or indirectly, and many of them are transcription factors essential for later anther development, such as DYT1, MYB33 and MYB65 (Figure 1-2). Till now, the binding site of SPL/NZZ remains to be elucidated.

### 2.2.2 Transcription factors involved in tapetal development

After the establishment of the anther wall, DYSFUNCTIONAL TAPETUM1 (DYT1), which encodes a bHLH transcription factor, controls the functions of tapetum at later stages (Sorensen et al., 2003b; Zhang et al., 2006a). In dt1, tapetum cells harbor enlarged vacuoles but lack cytoplasm. The meiocytes have comparatively thinner callose walls and fail to complete cytokinesis. RNA in situ hybridization experiments showed that DYT1 reaches its peak expression at anther stages 5 to 6 in tapetum cells (Zhang et al., 2006a). Because of the reduction of the DYT1 expression level in spl/nzz, it is hypothesized that DYT1 is downstream of SPL/NZZ (Figure 1-2). Previous studies in genes belonging to the bHLH family have shown that the basic region of the bHLH domain can specifically recognize a consensus E-box, CAXXTG. Many genes involved in late anther development with E-box in their putative promoter sequences have
been studied and hypothesized as DYT1 targets, and some of them have been confirmed by in vitro binding experiment, such as Male Sterile 1 (MS1) (Figure 1-2).

In addition to bHLHs, transcription factors encoded by other gene families that are involved in anther development have also been identified, e.g., the MYB family (Mandaokar et al., 2006; Millar and Gubler, 2005; Preston et al., 2004; Zhang et al., 2007; Zhu et al., 2008). These transcription factors can regulate many aspects of anther development, especially those related to pollen wall composition. MYB33 and MYB65 act redundantly to facilitate anther development, especially under adverse growth conditions such as cold (Figure 1-2) (Millar and Gubler, 2005).

TAPETAL DEVELOPMENT AND FUNCTION1 (TDF1) is another transcription factor that functions at later stages of anther development than DYT1. TDF1, also known as MYB35 in the R2R3 MYB family, regulates callose dissolution. Its expression in tapetal cells, meiocytes and microspores suggested that it functions around meiosis. In tdf1 mutants, the tapetum does not complete cell division (Zhu et al., 2008). The lack of TDF1 can also result in the abnormal expression of other genes encoding transcription factors, such as AMS, MS1 and MYB103 (Figure 1-2) (Zhu et al., 2011).

ABORTED MICROSPORES (AMS) function after meiosis, because the microsporocytes can accomplish meiosis in ams mutants. However, the tapetum cells in ams are abnormal and microspores are degraded before mitosis (Sorensen et al., 2003). In the absence of AMS, many genes involved in lipid metabolic, signal transduction, exocytosis and transportation pathways showed alteration in their expression levels (Xu et al., 2010). In addition, the yeast-two-hybrid experiments also suggested that AMS can interact with many other bHLHs (such as DYT1, bHLH89 and bHLH91) and also a SET-domain protein ASHR3, indicating its indispensable roles in anther development by interacting with different regulatory partners (Figure 1-2) (Thorstensen et al., 2008).
MYB103 (or MYB80/MS188) is also required for tapetum development, callose dissolution and normal exine deposition (Higginson et al., 2003; Zhang et al., 2007b). The myb103 mutant showed distorted pollen and enlarged vacuoles in the tapetal cells. Transcriptome analysis of myb103 mutants suggested that MYB103 impacts cell wall modification, lipid metabolism and signal transduction (Zhu et al., 2010). It is expressed as early as stage 5 and through stage 9, and it functions upstream of MS1 (Fig. 2) (Zhu et al., 2011).

MALE STERILITY 1 (MS1), as a Plant Homeodomain (PHD) transcription factor, regulates the callose dissolution, and mutations in the ms1 gene result in male sterile phenotype. Many genes differentially expressed in the ms1 mutant play important roles in secretion and programmed cell death (PCD) in tapetal cells (Ito et al., 2007; Ito and Shinozaki, 2002). Using inducible promoter driven MS1 constructs in the presence of protein synthesis inhibitor cycloheximide (CHX), Ito et al. identified the direct target of MS1, MYB99, the lack of which results in reduced fertility with an observation of thin tapetum (Figure 1-2) (Alves-Ferreira et al., 2007).

MADS, bHLHs, PHD, and MYBs have been shown to have crucial roles in anther development. They might function independently targeting certain cis-regulatory element, or interact with each other to bind to different motifs. For example, bHLHs can form homo- and hetero-dimers to regulate their direct targets (Skinner et al., 2010). Recently, the in vitro binding experiments showed physical interaction between particular sub-families of bHLH and MYB, and their regulatory synergy indicated their combinatorial regulation in plant development (Feller et al., 2011). Wilson et al. then hypothesized that MYB33 and MYB65 might interact with DYT1 to regulate early anther development (Fig. 2) (Wilson and Zhang, 2009). It would be important to understand the complicated regulatory network of anther development given the ever-increasing knowledge at molecular and cellular levels.
Figure 0-2. Transcriptional regulatory network of anther development in *Arabidopsis*.

Solid arrows represent for direct regulation and dotted arrows indirect. Arrows with round end represent protein-protein interaction. (AG, AGAMOUS; AMS, ABORTED MICROSPORE; DYT1, DYSFUNCTIONAL TAPETUM 1; MS1, MALE STERILITY 1; NZZ/SPL, NOZZLE/SPOROCYTELESS).
2.3 cell-cell communication and other genes important for anther development

Besides the transcription factors controlling archesporial specification, there are other factors that control meiosis, later pollen production, pollen wall and exine formation, tapetal programmed cell death and anther dehiscence (Wilson and Zhang, 2009), particularly a number of leucine-rich repeat receptor-like protein kinases (LRR-RLKs) (Becraft, 2002; Zhao, 2009). Among those, BARELY ANY MERISTEM1 & BAM2 (BAM1&BAM2), two redundant genes, specify the anther cell fate, defining identity of different somatic cell layers, including the endothecium, middle layer and tapetum (DeYoung et al., 2006; Hord et al., 2006). The double bam1 bam2 mutants can produce microsporocytes, but not somatic cells, suggesting their function at early stage 2 (Figure 1-3).

At later stages, EXCESS MALE SPOROCYTES1/ EXTRA SPOROGENOUS CELLS (EMS1/EXS) which belongs to LRR-RLKs is also essential to male reproduction, and TAPETUM DETERMINANT1 (TPD1) is suggested as its ligand (Figure 1-3) (Yang et al., 2005; Zhao et al., 2002). In both ems1 and tpd1 mutants, the anther produces more microsporocytes at the expense of tapetum, indicating a crosstalk between adjacent cell layers that determine the cell fate of archesporial cells (Jia et al., 2008). This hypothesis has been supported by the in situ hybridization experiments showing that the EMS1/EXS gene is primarily expressed in the tapetum while TPD1 is appeared in microsporocytes, and by the fact that TPD1 can induce phosphorylation of EMS1/EXS.

Besides EMS1 and TPD1, SOMATIC EMBRYOGENESIS RECEPTORLIKE KINASES1 and SOMATIC EMBRYOGENESIS RECEPTORLIKE KINASES 2 (SERK1&SERK2) regulate the differentiation of the tapetum redundantly (Figure 1-3) (Albrecht et al., 2005; Colcombet et al., 2005). They were proposed to be involved in the same signaling pathway as EMS1/EXS because the serk1 serk2 double mutant shares a similar phenotype as ems1/exs.
Recently, the *RECEPTORLIKE PROTEIN KINASE2* (*RPK2*) is also suggested to function in the differentiation of middle layer and maintaining tapetum (Figure 1-3) (Mizuno et al., 2007; Mizuno et al., 2007). In the absence of RPK2, anthers cannot produce the middle layer but can only produce hypertrophic tapetal cells. The expression of *RPK2* is not limited to the middle layer but again tapetum, indicating its function in regulating the differentiation of the middle layer and maintaining the tapetum.

In addition to those regulators mentioned above, a large number of other genes have also been found to be expressed in anthers, mutations of many of which lead to male sterility by affecting archesporial differentiation, tapetum formation, microsporocyte division or pollen maturation (Boavida et al., 2005; Hord et al., 2006b; Ma, 2005; Paul M. Sanders, 1999; Wilson et al., 2001; Zhang et al., 2007; Zhu et al., 2008d). However, the limitation of forward genetics impedes our exploration of genes function in fertility, mainly due to the existence of functionally redundant genes and the subtleties of the phenotypes of single-gene mutants (Cutler and McCourt, 2005). Therefore, the expression profiling as an alternative method has been applied to discover genes that may be involved in anther development.
Figure 0-3. The key LRR-RLKs regulating cell differentiation during early anther development at different anther stages.

BARELY ANY MERISTEM1 & BAM2, BAM1&BAM2; EXCESS MALE SPOROCYTES1/ EXTRA SPOROGENOUS CELLS, EMS1/EXS; TAPETUM DETERMINANT1, TPD1; SOMATIC EMBRYOGENESIS RECEPTORLIKE KINASES1 & SERK2, SERK1&SERK2; RECEPTORLIKE PROTEIN KINASE2, RPK2.
3. Plant Responses to Drought Stress

3.1 Response under drought stress in *Arabidopsis*

The importance of internal signals in floral development has been studied for decades. However, how external signals impact reproductive organ development after flower initiation largely remains unknown (Irish, 2010). Plant growth can be influenced by different stresses, such as drought, salinity, and high and low temperatures (Huang et al., 2011). Due to their sessile nature, plants have to adapt to the various environmental stress conditions in order to survive, especially in drought conditions, because drought, as one of common stresses, has obvious effects on plant growth in many aspects (Seki et al., 2007)

After the exposure to water-deficit stress, desiccation as a common phenomenon will lead to irreversible damage to plants and thus affect the productivity of plants (Postel et al., 1996). Studies of several crops had shown that the lack of water could severely affect reproductive organs and even lead to sterility (Lalonde et al., 1997; Sheoran and Saini, 1996). It has been shown that plants will go through many changes morphologically, physiologically and biochemically to respond to the severe drought stress. In the studies of the responsive mechanisms of plants, *Arabidopsis* as a model plant has been widely used (Yamaguchi-Shinozaki and Shinozaki, 2006; Zhang et al., 2004) by which the complexity of several signal pathways altering the expression of many genes has been revealed (Bray, 2004).

3.2 The ABA-dependent transcriptional regulation in response to drought stress in *Arabidopsis*

In *Arabidopsis*, abscisic acid (ABA) is the main plant hormone involved in the drought responsive pathways. ABA, as an intracellular messenger, can trigger root development, stomatal
closure and bud dormancy, and thus prevent further water loss under drought condition. Many genes which are induced by drought are also responsive to exogenous ABA treatment, and their induction is completely blocked in ABA-deficient mutants (Seki et al., 2007). The ABA-inducible genes include those encoding stress responsive proteins, transporters, enzymes, late embryogenesis abundant (LEA) proteins, and regulatory factors, such as transcription factors, protein kinases and phosphatases (Kempa et al., 2008). And those ABA regulated genes account for approximately 10% of the entire Arabidopsis coding-genes, more than those induced by the other known plant hormones (Nemhauser et al., 2006).

Drought stress triggers the synthesis of ABA, which can then alter the expression levels of many other genes by activating ABA responsive transcription factors (Choi et al., 2000). At initial stages of signaling, ABA is recognized by its receptors, PYRABACTIN RESISTANCE (PYR)/REGULATORY COMPONENT OF ABA RECEPTOR (RCAR) proteins (Figure 1-4). The PYR/RCAR protein changes its conformation upon ABA binding and forms a new interface on its surface. After ABA activation, nine of the fourteen PYR/RCARs have been identified to interact with PROTEIN PHOSPHATASE 2Cs (PP2Cs), which have been known as negative regulators of ABA signaling pathway (Nishimura et al., 2010). PP2Cs bind to another protein class, SNF1-RELATED PROTEIN KINASE 2s (SnRK2) (Figure 1-4), which serve as positive regulators. In the presence of ABA, PYR/RCARs interact with PP2Cs to inhibit their activity and release SnRK2 to activate downstream genes (Figure 1-4). Thus, SnRK2, together with PP2Cs, receive the signal from ABA and form the PYR/RCAR–PP2C–SnRK2 signal transduction pathway (Figure 1-4) (Hubbard et al., 2010).

Upon activation of SnRK, several ABA-dependent genes have been identified as its direct targets, including several ion channels, enzymes, and transcription factors (Hubbard et al., 2010b). The transcription factors were named ABRE-binding factors (ABFs) because they bind to ABRE (PyACGTGG/TC), a major cis-acting element and then activate ABA-dependent gene
expression (Shinozaki et al., 2003; Uno et al., 2000). However, a single copy of ABRE is not sufficient to respond to the ABA signal. Additional copies of ABRE or other binding elements (such as CE6) near ABRE are required for successful induction (Figure 1-4) (Hobo et al., 1999). For example, RD29B has two ABRE motifs in its promoter that are required for response to ABA treatment (Nakashima et al., 2006). Activation of ABF downstream genes also requires ABA-dependent posttranscriptional modifications (Furihata et al., 2006). Over-expression of AREB1, an ABF, does not induce downstream gene expression until ABF has been phosphorylated at its R-X-X-S/T sites in the conserved region, which is the common target of SnRK2s. This finding has further been confirmed by site-directed mutagenesis experiments, suggesting the vital role of the phosphorylation of AREB1. Besides AREB1, three other TFs have been identified as water stress responsive genes (Yoshida et al., 2010). All of these four ABFs (ABF1-4) belong to subgroup A of the bZIP protein family, which consist of 13 members; the basic region is responsible for the specific binding of the TFs to the target motif and the leucine zipper for the protein-protein interaction. They form hetero- or homodimers to function in the nucleus upon activation via SnRK2s (Yoshida et al., 2010).

There are other drought-inducible transcription factors involved in the ABA-dependent pathway, such as AtMYC2 and AtMYB2 (Abe et al., 2003). Over-expression of these two genes leads to an ABA-hypersensitive phenotype and increasing drought resistance (Hubbard et al., 2010a). As transcriptional regulators, MYC2 and MYB2 positively regulate RD22 expression by recognizing their cis-acting elements in the promoter region of RD22.

### 3.3 The ABA-independent pathways regulating drought response in Arabidopsis

Besides the ABA-dependent pathway, ABA-independent pathways are also important for plant to respond to drought (Liu et al., 1998). The ABA-independent genes can be activated by
drought in ABA deficient mutants, suggesting their independence of ABA. However, ABA can still trigger elevated expression levels of some of them, indicating crosstalk between the ABA-dependent and ABA-independent pathways.

The ABA-independent response to drought is regulated via dehydration responsive element (DRE) (A/GCCGAC), the target of DRE-binding proteins (DREB) transcription factors. These transcription factors belong to the ERF/AP2 family, which bind to the consensus binding motif. Among them, DREB1/CBF and DREB2 have been studied for years. DREB1/CBF factors are involved not only in drought response but also in cold response, whereas the DREB2 is specifically involved in drought response and activates many other functional genes as transcription factor by specifically binding to the DRE motif (Liu et al., 1998; Sakuma et al., 2002). Phylogenic studies showed that at least six other DREB2 homologues and two other DREB homologs are also induced by drought. However, the lack of phenotypes in their over-expression transgenic lines suggested that post-transcriptional regulation is needed (Sakuma et al., 2006). Moreover, the truncated DREB2A functions as a constitutively active form and results in significant drought tolerance.

On the basis of the identification of ABA-independent factors, the downstream genes involved in the pathways have been identified by microarray experiments and computational motif searching (Seki et al., 2002). Genes with a core DRE motif were then hypothesized as direct targets of DREB transcription factors (Shinozaki et al., 2003). Interestingly, many genes with differential expression levels have no known DREB binding site, indicating that other cis-regulatory elements responsive to drought are yet to be elucidated.
3.4 The crosstalk between different pathways in *Arabidopsis*

The crosstalk between the ABA-dependent and ABA-independent pathways has been suggested by genetic and molecular analyses (Seki et al., 2002). The promoter sequence analyses have shown that many drought inducible genes contain both DRE and ABRE motifs, indicating the cooperation of the two main pathways; this was later supported by the study of *RD29A* expression changes (Narusaka et al., 2003). In addition, other stresses, such as cold and high salinity also share some key regulators within the drought responsive network (Shinozaki et al., 2003).

Transcriptome analyses have identified many other genes that are triggered by drought stress via ABA-dependent or -independent pathways (Seki et al., 2007). A large number of functional genes differentially expressed under drought stress have provided more information about the drought-inducible metabolic pathways (Rizhsky et al., 2004; Valliyodan and Nguyen, 2006). Accumulation of several small molecules, such as sugars, sugar alcohols, amines, amino acids and secondary messenger calcium, has been revealed to be involved in drought tolerance in plants (Bartels and Sunkar, 2005; Kasukabe et al., 2004; Nanjo et al., 1999; Salekdeh et al., 2005; Seki et al., 2007; Taji et al., 2002). Another main category of genes induced by dehydration encode transcriptional regulators, including kinases, phosphatases and transcription factors. Several families of transcription factors have been identified to be drought-inducible, such as MYB, bZIP, NAC, bHLH and homeo-domain families (Yamaguchi-Shinozaki and Shinozaki, 2006). The immediate responsive genes regulate several transcription factors to achieve stress tolerance gradually.

Moreover, the protective proteins are also accumulated in drought treated plants, such as late embryogenesis-abundant proteins (LEAs), aquaporins, ion channels and heat shock protein (HSP). The abundance of the protective proteins correlates with ABA levels and drought
tolerance, facilitating the survival of vegetative tissues under drought condition. For example, the LEA proteins have been divided into different subgroups based on their conserved domains, and each subgroup may contribute to a unique function, such as maintaining cell structure, binding water molecules, protecting cells from oxidative stress or maintaining normal protein structure (Dure et al., 1989; Galau et al., 1986; Mowla et al., 2006; Olvera-Carrillo et al., 2010).

In summary, the plant’s response to water deficiency involves expression shifts in thousands of genes that are precisely regulated by both transcriptional and post-transcriptional regulators. Previous studies have shown that alteration of their expression levels could influence the ability of plants to adapt to drought. Future research in the direction of identifying immediate regulators or in searching for downstream functional genes can be alluring in order to improve drought tolerance.
Figure 0-4. Transcriptional regulatory networks of transcription factors involved in drought response in *Arabidopsis*.

Transcription factors are shown in ellipses; cis-regulatory elements in boxes and post-transcriptional regulation in small circles. PROTEIN PHOSPHATASE 2Cs, PP2Cs; PYRABACTIN RESISTANCE/REGULATORY COMPONENT OF ABA RECEPTOR, PYR/RCAR; SNF1-RELATED PROTEIN KINASE 2, SnRK2; ABRE-binding factor, ABF; DRE-binding protein, DREB.
4. Recent Progress in Transcriptome Analyses in Plant

4.1 Technologies and methods in transcriptomic studies

The transcriptome, the whole repertoire of RNA molecules, is the key link between the genetic information embedded in the whole genome and the proteome, and thus resulting in the phenome. The RNA transcripts include mRNA, rRNA, tRNA, and other non-coding RNA. Transcripts in a single cell reach up to 3 million molecules of RNA, and the copy number from each gene varies drastically (Velculescu et al., 1999). The large size of a transcriptome made sRNA profiling very time-consuming and expensive using traditional methods, such as Northern blots, reverse-transcription PCR, expressed sequence tags (ESTs), and serial analysis of gene expression (SAGE), till the invention of microarray analysis.

4.1.1 Invention of microarray

A microarray contains a set of oligonucleotide probes on a solid substrate, and the design of probes depends on the availability of genome sequence, ESTs or opens reading frames (ORF). After extraction from the samples, transcripts are labeled with fluorescent dyes, hybridized to the probes. Signals (fluorescent intensities) corresponding to specific genes are recorded as a measure of expression level. Both one-color and two-color arrays have been employed and the variance between labs has prompted the development of quality control standards to ensure the reliability of microarray experiments (Patterson et al., 2006; Shi et al., 2006).
The conventional microarray techniques are limited by the knowledge about expressed sequences that are used for probe design. One type of microarray chip, tiling array, has overcome this problem. Unlike the common probe on chips, probes of tiling array cover the whole genome contiguously. Tiling arrays can discover previously unknown expressed regions in the genome; however, they are much more expensive than conventional arrays and contain highly similar sequences that will cross-hybridize with different transcripts.

### 4.1.2 The innovation of RNA-Seq

The revolution of transcriptome analysis has occurred recently along with the massive parallel sequencing of the genome. Drastically contrast to the early genome project, it takes only a few days to sequence the whole genome nowadays using a new tool, known as deep sequencing (Carrier et al., 2011; Mardis, 2008). The novel high-throughput sequencing method has extended to RNA, known as RNA-Seq using Illumina, HiSeq or SOLiD instruments, and become rapid-evolving alternative for transcriptome profiling. Instead of hybridizing the transcripts and probes, RNA-Seq directly sequences the transcripts and maps them back to the reference genome (Martin and Wang, 2011; Surget-Groba and Montoya-Burgos, 2010). Different from microarray, RNA-Seq has direct access to sequence. Because of this advantage, it can detect novel transcripts, alternative splicing, polymorphisms and even transcriptome without genome information.

### 4.1.3 Comparison between Microarray and RNA-Seq

It is obvious that RNA-Seq boasts some advantages that microarray cannot achieve. However, we still need to take some practical factors into consideration before using the new method. First, the low cost of commercial microarray chips, only about one tenth of RNA-Seq per
sample, is very competitive (Malone and Oliver, 2011). Second, the maturation of the microarray methodology minimizes the biases in the expression data (Aittokallio et al., 2003; Do and Choi, 2006) while the exploration of sequencing analyses is still evolving (Aittokallio et al., 2003). Moreover, RNA-Seq failed to provide adequate coverage of low copy genes in some studies, raising questions about the accuracy of the results (Auer and Doerge, 2010). In addition, it also allows biases in the expression due to heterogeneity in coverage. Last, but not the least, analyses of RNA-Seq require mastering a high-level computational language. It intimidates many biologists with little knowledge of programming (such as perl, Python, Unix and so on), which is essential for managing the large datasets generated by RNA-Seq.

Both microarray and RNA-Seq have been tested by traditional methods, such as RNA \textit{in situ} hybridization, northern blot or real-time PCR experiments (VanGuilder et al., 2008; Wijeratne et al., 2007; Zahn et al., 2010b). The strong correlation of the expression levels of transcripts from high-throughput and the conventional methods suggests that the high-throughput expression profiling provides reasonably reliable results in both plant and animal studies (Han et al., 2009; Malone and Oliver, 2011; Yang et al., 2011; Zahn et al., 2010b; Zhang et al., 2005). Given the advantages of both methods, it is highly possible that both microarray and RNA-Seq will remain to be commonly used in transcriptomic studies, regardless of the fact that RNA-Seq might be a better choice for specific purposes, such as detection of novel transcripts and isoforms.

\subsection*{4.2 Application of expression profiling using microarray in plant}

The increasing genome information and the commercialized microarray chips have facilitated the extensive applications of microarray. The transcriptome from specific cell types (Tang et al., 2011; Yang et al., 2011) and tissues (Zhang et al., 2005; Zhao et al., 2011) across varies developmental stages (Galla et al., 2009; Vodkin et al., 2004) under multiple
environmental conditions or chemical treatment (Guo and Gan, 2011; Umezawa, 2011) in different genotypes (Alves-Ferreira et al., 2007; Wijeratne et al., 2007) and species (Zahn et al., 2010) have been obtained.

4.2.1 The wide application of microarray in various species

In the latest decade, microarray have been commonly utilized for diverse investigations in plants, from specific gene function to global expression shifts under various conditions and even between various species. Till now, microarray technology has not been restricted to Arabidopsis, but applied to many important economic crops, including rice (Oryza sativa L.) (Chauhan et al., 2011), tomato (Solanum lycopersicum) (Liu et al., 2011), potato (Solanum tuberosum) (van Dijk et al., 2010), maize (Zea mays) (Gomez-Anduro et al., 2011), wheat (Triticum aestivum L.) (Laudencia-Chingcuanco et al., 2011), wine grape (Vitis vinifera L.) (Tillett et al., 2011), peanut (Arachis hypogaea L.) (Bi et al., 2010), soybean (Glycine max) (Duressa et al., 2010), cotton (Gossypium hirsutum L.) (Yao et al., 2011), and poppy (Eschscholzia californica) (Zahn et al., 2010). The transcriptome analyses from these species in different circumstances would be useful for food safety assessment (van Dijk et al., 2010). Moreover, the transcriptome of genes at interesting positions of phylogenetic trees might also shed some interesting lights on the evolutionary history (Chanderbali et al., 2010).

4.2.2 The application of microarray for different research interests

Mutants with obvious phenotypes during plant development have been studied by analyzing their transcriptomic differences with those of the wild type. Those studies have involved genes encoding transcription factors (Wijeratne et al., 2007), proteasome proteins (Ma et
al., 2003), transporters (Conn et al., 2011; Liu et al., 2010; Liu et al., 2011), phytochromes (Wang et al., 2002), etc. The transcriptome information from mutants not only provided more information to understand the function of the mutated gene at the molecular level but also enabled the global identification of genes that are involved in a particular transcriptional regulatory pathway.

Microarrays have also been widely applied to study the influence of environmental signals on plants. The transcriptome profiling of circadian clock, for example, showed that 6% of the genome in *Arabidopsis* is under clock regulation and thus elaborated the key transcriptional pathways in this process (Harmer et al., 2000). The transcriptome analyses in response to diverse stress, biotic or abiotic, are popular topics which have attracted attention of many outstanding scientists. The adaptive shifts of the transcriptome due to the alteration of nutrients, including carbon (Price et al., 2004), nitrogen (Scheible et al., 2004; Wang et al., 2003) and inorganic phosphate (Nilsson et al., 2010), have been studied for years. In addition to deficiency in nutrients, other abiotic stresses threaten the sustainability of agricultural growth even more severely (Ahuja et al., 2010). Microarray, as a rapid and effective tool, has been extensively used to understand the effects of the abiotic stresses, such as heat (Kotak et al., 2007), cold (Lissarre et al., 2010), drought (Ahuja et al., 2010), salt (Sahi et al., 2006), CO₂ and ozone (Ludwikow and Sadowski, 2008). On the other hand, microarrays are also good approaches to study the mechanisms by which plants respond to the biotic stress (Sahi et al., 2006). The integration of all above applications of microarrays in plant studies can enable us to develop a more sophisticated transcriptional regulatory network in plant.
4.3 Tools for microarray analyses

A series of large-scale microarray experiments in *Arabidopsis thaliana* have brought about tremendous accumulation of expression data in these years. However, how to analyze the huge amount of data becomes a challenge for scientists. Recently, many biologists together with specialists in bio-statistics or computer sciences have strived to scrutinize the data, and many tools became available to expand our understanding of the data.

4.3.1 Normalization and Clustering

The first step before deep investigation of the data is to control the variance by normalizing the raw data using statistical methods, such as Robust Multichip Average (RMA) (Bolstad et al., 2003). After verifying microarray data, people tend to focus on the genes differentially expressed among different samples, and the criteria for selection normally need statistical support (t-test, ANOVA, Significance Analysis of Microarrays (Tusher et al., 2001), or false discovery rate) besides applying an arbitrary cut-off (Zhang et al., 2005).

The genes with similar expression shifts might share similar function or be co-regulated. The similarity between genes that have similar expression shifts could be tested by clustering, which classifies data into groups. Generally, methods of clustering can be divided into two groups, hierarchical and non-hierarchical. Hierarchical clustering can be further divided into two types, agglomerative and divisive clustering. Many types of agglomerative clustering methods to calculate distance between clusters could be used. Single, complete, average and centroid linkage clustering are the most commonly used ones. Divisive hierarchical clustering is not as commonly used in microarray analysis. Though hierarchical clustering is widely used in microarray analysis, no compelling evidence shows that it is the best structure for expression profiling. As an
alternative, non-hierarchical methods (K-mean or Self Organizing Maps) are also widely applied to microarray analysis.

### 4.3.2 Construction of regulatory network

On the basis of clustering results, it is reasonable to hypothesize that genes within the same cluster (sharing similar expression patterns) might be regulated by the same set of regulatory factors. Therefore, they should have similar binding sites of these transcription factors in their putative promoter (normally within 2.0 kb upstream of the coding sequence). This hypothesis has been tested by various research groups in different organisms, such as yeast (Brazma et al., 1997; Bussemaker et al., 2001; Vilo et al., 2000), animals (Furman and Pilpel, 2006), and plants (Higo et al., 1998; Higo et al., 1999). The occurrence of sequence patterns of putative binding sites in promoter regions has been studied in the co-expressed genes following three steps: 1) identify genes sharing similar expression patterns; 2) Collect putative sequences upstream of coding region of these genes; 3) Search for statistically enriched motifs in those sequences and evaluate the significance of the results.

Besides, many studies have integrated the transcriptome information with external information, for instance protein conservation (determined by phylogenetic trees), to predict protein-protein interaction and furthermore their functions. Previous studies have suggested that co-expressed genes tend to encode proteins that interact with each other (Gerstein and Jansen, 2000), and function in the same pathway (van Noort et al., 2003). In this kind of study, scientists should first find co-expression pattern in two organisms, and then search for orthologous proteins with similar expression pattern in both organisms, and finally integrate the known metabolic pathway information to make a hypothesis on the interaction or function.
4.3.3 Software and tools for microarray

The expression profiling can infer regulatory relationships and construct a network (Segal et al., 2003). Both probabilistic model (Segal et al., 2003) (Gardner et al., 2003) have been used in generating model networks. In Arabidopsis, many websites have been designed to facilitate microarray analysis. All the sequences could be downloaded from TAIR (http://www.arabidopsis.org/) or NCBI (http://www.ncbi.nlm.nih.gov/). In order to compare with previous microarray data, many databases provide microarray datasets, such as GEO (www.ncbi.nlm.nih.gov/geo), SIGnAL (http://signal.salk.edu/cgi-bin/atta), Stanford Microarray Database (smd.stanford.edu), EBI (www.ebi.ac.uk/arrayexpress). There is even more software available for clustering and pathway finding. The website of gene ontology provided a collection of updated tools for data analysis (www.geneontology.org/GO.tools.microarray.shtml). Besides the software available online, there is also commercial software designed for data analyses, such as SAM (www-stat.stanford.edu/~tibs/SAM/), Me.V (www.tm4.org/) and ArrayStar (www.dnastar.com/t-products-arraystar.aspx).

Recent advances in analyzing tools of microarray have made it possible to understand regulatory pathways on genome-wide scale without too complicated training in statistics or programming. Therefore it has allowed more and more people to explore the vast information from microarray and to propose testable hypotheses.
5. Reference


CHAPTER 3

COMPARATIVE TRANSCRIPTOMICS AMONG FLORAL ORGANS OF THE BASAL EUDICOT ESCHSCHOLZIA CALIFORNICA AS REFERENCE FOR FLORAL EVOLUTIONARY DEVELOPMENTAL STUDIES

The work described in this chapter has been published in Zahn and Ma, Genome Biology 2010; 11: R101

Laura Zahn, a former post-doctoral fellow in the Ma lab, Hong Ma, Naomi Altman and Claude dePamphilis designed this study. She and Donglan Tian, a former technician in the Ma lab, performed tissue collection, RNA isolation, preliminary test with RT-PCR and microarray experiments. Cynthia Gibas and Raad Gharaibeh designed the oligonucleotides for the probes on the microarray chip. Laura performed first round of statistical analysis of the microarray data after normalization by Naomi Altman and Qing Zhang, with additional assistance from Kerr Wall, a student in the dePamphilis lab, and wrote a draft manuscript. Subsequently, Naomi repeated the normalization without some of the probes, and I re-analyzed all of the gene expression results that are presented in the paper. I also performed real-time RT-PCR experiments for a supplemental figure. I wrote a new draft based on re-analyzed results, made all of the figures, except Figures 1B and 2, in the published paper. Naomi Altman made Figures 1B and 2.
1. Abstract

Molecular genetic studies of floral development have concentrated in several core eudicots and grasses (monocots), which have canalized floral forms. Basal eudicots possess a wider range of floral morphologies than the core eudicots and grasses and can serve as an evolutionary link between core eudicots and monocots, and provide a reference for studies of other basal angiosperms. Recent advances in genomics have enabled researchers to profile gene activities during floral development, primarily in the eudicot *Arabidopsis thaliana* and the monocots rice and maize. However, our understanding of floral developmental processes among the basal eudicots remains limited.

Using a recently generated EST (expressed sequence tag) set, we have designed an oligonucleotide microarray for the basal eudicot *Eschscholzia californica* (California poppy). We performed microarray experiments with an interwoven-loop design in order to characterize the *E. californica* floral transcriptome and to identify differentially expressed genes in flower buds with pre-meiotic and meiotic cells, four floral organs at pre-anthesis stages (sepals, petals, stamens and carpels), developing fruits, and leaves.

Our results provide a foundation for comparative gene expression studies between eudicots and basal angiosperms. We identified whorl-specific gene expression patterns in *E. californica* and examined the floral expression of several gene families. Interestingly, most *E. californica* homologs of *Arabidopsis* genes important for flower development, except for genes encoding MADS-box transcription factors, show different expression patterns between the two species. Our comparative transcriptomics study highlights the unique evolutionary position of *E. californica* compared with basal angiosperms and core eudicots.
2. Introduction

The eudicots are believed to have originated approximately 130 million years ago (Friis et al., 2006). They include about 70% of all flowering plant species and are comprised of core eudicots (Anderson et al., 2005; Leebens-Mack et al., 2005; Moore et al., 2010), which include the groups containing *Arabidopsis thaliana* and *Antirrhinum majus*, respectively, and species that branched earlier from these groups and are at basal positions within the eudicot clade. The earliest branching lineage of the eudicots, the Ranunculales, contains the Papaveraceae (poppy) family, of which *Eschscholzia californica* (California poppy) is a member (Cajsa Lisa Anderson, 2005). The core eudicots commonly have stable (i.e., canalized) flower architecture (Figure 2-1A); by contrast, the basal eudicots exhibit a wider range of floral patterns (Endress and Doyle, 2007) (see examples in Figure 2-1A). Comparing the morphology and the underlying mechanisms of flower development between the core and basal eudicots may help us better understand the evolution of flower structures and development.

Molecular genetic studies in *Arabidopsis*, *Antirrhinum* and other core eudicots have uncovered the functions of many genes involved in regulating flowering time and floral organ identity and development (Baek et al., 2008; Liu et al., 2009; Putterill et al., 2004). In particular, it is known that a number of MADS-box genes are required to control flowering time and floral organ identities, as well as anther, ovule and fruit development. These include the well-known ABC genes *APETALA1* (A function), *APETALA3* and *PISTILLATA* (B function), and *AGAMOUS* (C function) from *Arabidopsis*, and their respective orthologs from *Antirrhinum* (*SQUAMOSA, DEFICIENS, GLOBOSA*, and *PLENA*) (Soltis et al., 2007; Zahn et al., 2005a; Zahn et al., 2005b). Comparative studies of core eudicots suggest that homologs of B and C function genes have relatively conserved functions, although some divergences have also been observed. Putative orthologs of these MADS-box genes may have diverged expression patterns in different species.
and the expression difference between recent duplicates is often associated with subfunctionalization (Zahn et al., 2005a; Zahn et al., 2005b). In addition, several MADS-box genes have been found to be important for floral organ identities in the monocots (Ma and dePamphilis, 2000; Paolacci et al., 2007; Theissen et al., 2000; Yamaguchi and Hirano, 2006). However, both the long evolutionary distance and the highly diverged flower architectures between monocots and core eudicots have made it difficult to study the evolution of floral gene function.

The investigation of floral gene function in the basal eudicots serves to bridge the gap between core eudicots and monocots. Molecular and expression studies of floral genes have been reported for some basal eudicots, providing informative initial knowledge of the conservation and divergence of floral gene activities among eudicots (Cui et al., 2006; Orashakova et al., 2009; Soltis et al., 2008). Molecular evolutionary studies of several MADS-box subfamilies, complemented by expression analyses, support that some of the MADS-box genes have maintained conserved functions throughout angiosperm evolution (Carlson et al., 2006; Jaramillo and Kramer, 2007; Kramer et al., 1998; Nam et al., 2004; Zahn et al., 2005). For example, expression studies of floral MADS-box genes in *E. californica* demonstrated that genes in the *AGAMOUS, GLOBOSA* and *SEPALLATA* subfamilies are highly conserved between basal and core eudicots (Carlson et al., 2006a; Zahn et al., 2005a; Zahn et al., 2005b). Additionally, in other ranunculids, expression divergences have also been observed between recently duplicated MADS-box genes (Zahn et al., 2005a; Zahn et al., 2005b).

High-throughput technologies, including microarrays, can be used to analyze transcriptomes of individual floral organs at specific developmental stages. Transcriptome studies have been performed extensively for *Arabidopsis* and, to a lesser extent, several other highly derived core eudicots (Galbraith, 2006; Ma, 2005; Orashakova et al., 2009; Tung et al., 2005; Wellmer et al., 2004; Yang et al., 2003; Zhang et al., 2005). Among basal eudicots, such studies
have only been carried out recently in the basal eudicot *Aquilegia*, which represents a different ranunculid lineage than *E. californica* (Voelckel et al., 2010). *E. californica* is a potential model organism because (1) it has a relatively small plant size, many seeds per fruit and a short generation time, which facilitate genetic studies; (2) it does not have determinate flowering and produces multiple flowers over its lifespan, providing easy access to floral materials (Becker et al., 2005); (3) it has a relatively small genome; and (4) it both has an efficient system for virally induced gene silencing and is transformable (Becker and Lange, 2010; Bennett and Leitch, 2005; Carlson et al., 2006b; Park and Facchini, 2000; Wege et al., 2007). Previous gene expression studies in *E. californica* showed that there is very good correlation between regions of gene expression and domains of gene function (Becker and Lange, 2010; Liscombe et al., 2009; Orashakova et al., 2009; Wall et al., 2009). An *E. californica* EST collection of over 6000 unigenes was constructed from a pre-meiotic floral cDNA library (Carlson et al., 2006), which provides gene sequence information for microarray analysis of *E. californica* leaf and floral transcriptomes. A transcriptome-level analysis facilitates our understanding of floral development in basal eudicots and sheds light on potential floral regulatory genes in *E. californica*.

In this study, we used microarray technology to investigate transcriptomes in *E. californica* and to identify differentially expressed genes in developing leaves and floral buds at pre-meiotic (small buds) and meiotic (medium buds) stages. Additionally, we examined the transcriptomes of developing fruits and four types of floral organs (sepals, petals, stamens, and carpels) at the pre-anthesis stage. We identified genes that are significantly differentially expressed in different floral organs or at different floral stages, in comparison with developing fruit and leaf tissues. We also analyzed the expression of genes in several regulatory gene families, some of which contain homologs of known floral genes from other organisms. Finally, we compared our results with similar studies in *Arabidopsis* and recent studies (Moore et al., 2007; Voelckel et al., 2010) in *Aquilegia* and *Persea* (avocado), a basal angiosperm related to
magnolia, to assess conservation and divergence in gene expression and discuss their implications for evolution of floral development in the eudicots.

3. Results and Discussion

3.1 Construction and use of a microarray chip for *E. californica*

To investigate the leaf and reproductive transcriptomes of *E. californica*, we generated a custom Agilent microarray chip with features for 6446 unigenes from the *E. californica* EST collection (Carlson et al., 2006) (see Materials and Methods for additional information). The oligonucleotide sequences for the probes were selected using available sequence information from *E. californica* ESTs, as well as other public sequence information, avoiding non-specific hybridization as much as possible. Additional criteria were used to consider potential secondary structure and hybridization temperature (see Materials and Methods).

A primary objective was to obtain expression profiles with the power to detect differential expression between vegetative (leaves) and reproductive organs, between different floral stages, and between different floral organs. Therefore, we sampled the *E. californica* plants for the following eight representative organs and stages (for convenience, referred to generally as tissues hereafter): leaves, early floral buds, medium floral buds, four floral organs (sepals, petals, stamens, and carpels) at pre-anthesis, and young fruits. Four sets of plants were sampled at the same time daily (8:30-10:30AM) to minimize variation due to circadian rhythms, yielding four biological replicates. RNAs from these 32 samples were used to generate cDNAs and labeled with Cy3 and Cy5 dyes, for two-channel microarray experiments. Finally, we used an interwoven loop design (Figure 2-1B) to maximize the comparative statistical power using a limited number of hybridizations (Sacan et al., 2009).
In an interwoven loop design, differences in gene expression can be estimated for all pairs of tissues with a relatively small number of hybridizations (Altman and Hua, 2006). Each of the 8 tissues was directly compared on the same slide with one of four other tissues, with one biological replicate for each comparison, resulting in a total of 16 hybridizations. The comparison of the two tissues on the same arrays allowed more precise results than those compared indirectly via other tissues. The specific pairings on the same array were chosen to optimize precision of comparisons for biologically important comparisons, while keeping the precision of different comparisons as similar as possible. Because our EST library was constructed with floral bud mRNAs, we compared developing floral buds at different stages with each of the four floral organs, and compared each of these tissues with leaves, the only vegetative organ in this study, and developing fruits. The comparison between small buds and leaves was aimed at identifying differentially expressed genes at early reproductive stages. We hypothesized that the sepal should be the most leaf-like tissue among all floral organs; whereas previous studies [25] suggest that the stamens might have the most complex transcriptome among the four major floral organs (Wellmer et al., 2004a). In this study, the fruit tissue represents the only post-anthesis tissue. We also considered the ABC model, which posits that sepals and petals both require A function genes, petals and stamens both need B function genes, and stamens and carpels both depend on C function genes. In addition, carpels and fruits were developmentally related tissues, with small and medium buds representing two consecutive stages in floral development.

After microarray hybridizations, we tested the quality of the microarray experiments. We assessed the reproducibility of the microarray hybridizations by determining the Pearson’s correlation coefficients between the biological replicates for each of the eight tissues (see Figure 2-2 for an example). As shown in Fig. 2, the Pearson’s correlation coefficients between any pair of the four biological replicates of small buds, one of the most complex tissues in this study,
ranged from 0.94 to 0.97. The high correlation values indicate that our results were highly reproducible.

In addition, we examined signal intensities. Because the EST library used for the probe design was constructed from mRNAs of flower buds, we assumed that expression of most genes should be detected in our microarray experiments from mostly flower-related tissues. The value of 5.41 for log2 of hybridization intensity (10% quantile of all genes on the chip) was selected as a cut-off to identify “present” signal similar to previous microarray experiments in *Arabidopsis* (Zhang et al., 2005). For the 10% quantile, we identified the number of genes detected in leaves (5905), small buds (5906), medium buds (5876), sepals (5876), petals (5870), stamens (5877), carpels (5851) and fruits (5881). These results were not surprising because the unigenes were derived from EST data, which tend to favor genes that are expressed at relatively high levels. Therefore, our microarray chip and hybridization experiments were able to detect the expression of several thousand genes in eight major tissues of *E. californica*. Of the genes examined, the majority of genes present in leaf were also observed in small buds and medium buds (Figure 2-3A). In addition, most genes expressed in sepal were also expressed in petal (Figure 2-3B), suggesting similar gene expression levels between these two tissues. There was significant overlap of genes expressed in petal and/or sepal with genes expressed in carpel and stamen (Figure 2-3C). Similarly, there was considerable overlap of expressed genes between the carpel and fruit (Figure 2-3D); this is not surprising since fruit is derived from the ovary containing large carpel tissues. Using the same cut-off for detection of expression, 5554 genes were expressed in all 8 tissues. We then examined GO categorization of all 5554 genes and found that the “unknown” genes (homolog of genes annotated as unknown in *Arabidopsis*) were under-represented while some specific functional categories were slightly over-represented, including transferase and protein binding group. The observation that most of the genes in this study were expressed in all tissues might be because our EST collection represented relatively abundant
genes, including most house-keeping genes. This might also explain why the “unknown” category was under-represented because widely expressed genes tend to have known annotations.

To verify our microarray results, real-time reverse-transcription polymerase chain reaction (RT-PCR) was performed using RNAs from the same 8 tissues as those in microarray experiments. Nine representative genes were examined relative to our reference gene, including three MADS-box genes, EScAGL2 (87251), EScAGL6 (86583), EScDEF1 (83744) (Zahn et al., 2005a). The other genes were homologs of a transcription factor MYB35 (86850), a gamma-tip protein (84392), a putative ferrodoxin (85140), a transducin family/ WD-40 repeat family protein (84618), and homologs (86386 & 88941) of two Arabidopsis genes encoding different “expressed proteins” without a known function. The real time RT-PCR results indicate that the gene expression patterns were generally supportive of the microarray results, and were also consistent with previous in RNA situ hybridization experiments (Drea et al., 2007; Kim et al., 2005; Zahn et al., 2005; Zahn et al., 2005).

3.2 An overview of differential expression profiling of floral development

Although the E. californica ESTs were obtained from a cDNA library that was constructed with mRNAs from multiple stages of floral development (Carlson et al., 2006), many of the corresponding genes were also expressed in leaves, different stages and various organs of the flower, as well as fruits. To determine additional transcriptome characteristics, we investigated whether specific genes were expressed similarly or differentially in the tissues tested. Of the 6446 unigenes examined, most genes (4513/6446) were not significantly differentially expressed with more than a two-fold change between any two of the eight tissues (with P-value < 0.05).
Nevertheless, 1933 genes were found to be differentially expressed between at least two tissues; however, most of these 1933 genes showed similar expression levels in the other tissues (Figure 2-4A). Not surprisingly, carpel and fruit, as well as small and medium buds, showed the most similar expression patterns at sequential development stages. Leaf, the only vegetative organ in our study, had similar expression patterns to those of the green organs (carpel and fruit), which may be due to shared high expression of photosynthesis-related genes (see below). Interestingly, stamen had the most different expression profile, suggesting a distinct developmental process relative to the other floral organs.

To obtain additional insights into functions of those differentially expressed genes, we examined the GO categorization for the most similar Arabidopsis homologs of each poppy gene using functions within TAIR website. Genes encoding proteins categorized as “other enzyme activity” (chi-square test with $p$-value < 0.01) and “structural molecule” ($p$-value < 0.001) were enriched among those genes differentially expressed between at least two tissues (Figure 2-4C) relative to the control group of all genes on the microarray chip (Figure 2-4B). These results suggested that variation in the expression of metabolic genes across those tissues might be in part responsible for their morphological and/or physiological differences in *E. californica*.

### 3.3 Similar expression pattern of vegetative preferential genes in *E. californica* and in *Arabidopsis*

To identify genes with greater expression in either vegetative or reproductive tissues, we performed pairwise comparisons among all tissues as well as groups of floral organs and/or stages. Only one gene, 90036 (with no significant BLASTX hits to *Arabidopsis* predicted proteome, nor the NCBI NR database), was significantly two fold greater in all reproductive tissues and through all stages including fruit, compared to leaf tissue. However, 65 genes were
expressed significantly higher in leaves compared to all floral tissues and stages. To obtain overall expression patterns of vegetative genes, we constructed a heat-map (Figure 2-5A) resulting in two main clusters. In the first cluster, most genes that were highly expressed in leaves were also highly expressed in floral tissues except stamens. In the second cluster, most genes were highly expressed in leaves but not in the other tissues.

To compare gene expression pattern of leaf-preferential genes in *E. californica* and their homologs in *Arabidopsis*, we used BLAST to search the *E. californica* EST sequences against the *Arabidopsis* genome. Our BLAST results (with 10E⁻¹⁰ as cut-off) indicate that 58 out of the 65 leaf-preferential genes have identifiable homologs in *Arabidopsis*. On the basis of previous microarray data, of these 58 genes all but one (*RBCS1A*) of their *Arabidopsis* homologs were also preferentially expressed in leaves (Schmid et al., 2005). According to TAIR9 annotation, most of these genes encode for proteins that are localized in the chloroplast. GO categorization on the basis of gene function (methods) indicate that most of these genes are likely to be involved in photosynthesis, encoding homologs of protochlorophyllide reductases, photosystem I reaction center subunits and oxygen-evolving enhancer proteins.

### 3.4 Comparing transcriptome profiles at crucial stages of floral development in *E. californica* and in *Arabidopsis*

To identify developmental stage-specific genes in *E. californica* flowers, we examined the expression patterns of genes in the pre-meiotic (small buds), meiotic (medium buds) and pre-anthesis stages (four floral organs: sepals, petals, stamens and carpels). Pre-meiotic buds (small buds < 5mm) had 49 differentially expressed genes in comparison with any other tissues examined (*p*-value <0.05 and two-fold cut-off). Among these genes, 30 had identifiable *Arabidopsis* homologs, 24 of which have expression data available. Unlike leaf-preferential
genes, only seven of these 24 genes showed expression peaks in early *Arabidopsis* flower buds while the rest were predominately expressed in specific floral organs at higher levels than in leaves. The proteins encoded by these seven genes include two transcription factors, one oxidoreductase, two peroxidases, one electron carrier and one gene of unknown function (table 2-1, genes and annotation with peak expression in small floral buds, information obtained from Markus’ results (Schmid et al., 2005)). The *Arabidopsis* homologs for two transcription factors, *MYB35*, which regulates anther cell layer formation at early stages and a bHLH gene that has not been fully studied (Zhang et al., 2006; Zhu et al., 2008), were also preferentially expressed in anthers (unpublished data, X.M. and Baomin Feng). However, the corresponding *E. californica* genes were expressed at low levels in the pre-anthesis stamens, possibly because either these genes are not highly expressed in *E. californica* stamens or that our stamen expression data from pre-anthesis stamens was too late relative to the stages of highest expression in *Arabidopsis*, which may be during earlier anther developmental stages.

In medium buds (which span the meiotic stage), we found eight genes that were expressed two fold significantly higher and none that were significantly down regulated compared with any of the other tissues examined (table 2-1). All of these genes have homologs in *Arabidopsis* and most encode proteins that may have enzymatic activities (table 2-1). However, none of the *Arabidopsis* homologs of those genes show expression peaks in the equivalent stages as our medium buds in *Arabidopsis* (Schmid et al., 2005) (table 2-1). Interestingly, the homolog of *E. californica* gene 88096 in *Arabidopsis* (AT3G11450) encodes a DnaJ heat shock protein proposed to be involved in either mitosis or meiosis. The expression pattern of these homologs differs in that in *Arabidopsis* it is highly expressed in both vegetative and reproductive tissues. It is possible that the gene function might have diverged after the separation of basal eudicots from core eudicots.
In fruits, 9 genes were expressed significantly two fold higher than the other tissues in *E. californica* (table 2-1). None of their homologs showed an expression peak in the *Arabidopsis* fruit. Among the genes of particular interest, the *Arabidopsis* homolog of 86118 (At5g62200, MMI9) plays an important role in embryo development (Ascencio-Ibanez et al., 2008), and given its high expression in the fruits, suggesting that its *E. californica* homolog might have a similar function.

### 3.5 Identification of putative genes under control of certain genes in ABC model

According to the ABC model, the A-function genes are transcription factors that are required to properly specify the sepal (alone) and petal (along with the B-function genes) identities, with B-function genes specifying the stamens (along with the C function genes), and C function specifying the carpels. Thus, genes expressed in the sepals and petals (regions encompassing the A-domain) are called A-domain genes, genes expressed in the petals and stamens are called B-domain genes, and genes expressed in the stamens and carpels are called C-domain genes. Although the homologs of *Arabidopsis* A function genes (such as AP1 and AP2) might not have conserved functions in other eudicots [47-49], because of the distinct sepals and petals in *E. californica*, we tried to identify putative A-function genes on the basis of regulatory genes expressed in the A-domain, hypothesizing that they may function in specifying the sepal and petal identities in *E. californica*.

From our hypothesis that A-domain genes should be more highly expressed in the sepals and possibly in the petals, than in the other floral organs we compared them with three tissues: leaf, stamen and carpel collected approximately 1 day pre-anthesis. We found significantly greater expression of 64 genes in sepals over each of the above three tissues and 49 genes in
petals over each of the three tissues, respectively. When compared with all seven other tissues, 34 genes in sepals and 29 genes in petals were significantly preferentially expressed. Whereas genes highly expressed in sepals or petals tended to be expressed in all tissues at moderately high levels (Figure 2-5B & C), genes with lower expression in sepals and/or petals were scarcely expressed in other tissues. On the basis of comparisons of petals and sepals with leaves, stamens and carpels, only 5 genes were expressed two fold greater in tissues controlled by A-function genes (table 2). Interestingly, two of these genes are members of the MADS-box family. However, the expression of their closest *Arabidopsis* homologs, *AGL2/SEP1* and *AGL6*, is not sepal, petal or even floral specific (Figure 2-6D&F). *SEP1* is an E-function gene (Cui et al., 2009; Ma et al., 1991), and is involved in the development of all floral organs in *Arabidopsis*. A homolog of *SEP1* in soybean (*GmSEP1*) is expressed in reproductive development especially in petals and seed coats (Huang et al., 2009). *AGL6* and its homologs have been shown to function in flower development not only in eudicots, like *Arabidopsis* and *Petunia*, but also in orchid, rice, and other monocots. In the grasses, *AGL6* has high expression in paleas, lodicules, carpels and ovule integuments, as well as the receptacle (Hsu et al., 2003; Li et al., 2010; Ohmori et al., 2009; Reinheimer and Kellogg, 2009; Rijpkema et al., 2009). We hypothesize that other MADS genes, possibly *SEP* homologs, may serve as A-function genes in *E. californica* instead of *AP1* and *AP2* in *Arabidopsis*, in part because the *API* subfamily is closely related to the *AGL6* and *SEP* subfamilies (Litt and Irish, 2003).

B-function genes, such as the *Arabidopsis* *APETALA3* and *PISTILLATA* genes, are required for the identities of petals and stamens (Kanno et al., 2007; Soltis et al., 2007; Zahn et al., 2005). In monocots like tulip, homologs of *AP3* and *PI* are expressed in the tepals (petal-like organs found in the outer two whorls). We searched for putative B-domain genes on the basis of their expression patterns in *E. californica* and found that 60 genes in petals and 180 genes in stamens were expressed significantly higher in these organs than sepals, carpels and leaves. And
94 genes were 2 fold significantly greater in stamens than all the other organs. The large number of genes with stamen-preferential expression patterns suggested that the development of stamen requires more specialized genes. Alternatively, the larger number of stamen-preferential genes identified here may be explained by the fact that stamens comprise much of the biomass of developing *E. californica* buds, relative to other developing floral organs (Figure 2-5D).

We combined the expression data from petals and stamens to represent the B-domain group and compared their expression levels with those of leaves, sepals, carpels and fruits (table 2-2), identifying 13 genes as preferentially expressed in the B-domain organs. A homolog of *PI* (87167) and two homologs of *AP3* (83744 and 87005) were identified in this group (Zahn et al., 2005b). Since *PI* and *AP3* are B-function genes in *Arabidopsis* and other species, such as lily (Riechmann and Meyerowitz, 1997; Schultz et al., 1991; Tzeng et al., 2009), it is possible that their homologs in *E. californica* function in a similar manner. It should also be noted that *in situ* analysis showed that the *AP3* homologs are also expressed in ovules in *E. californica* (Zahn et al., 2005b), suggesting that they may have roles outside of B-function.

Of the genes preferentially expressed in the B-domain, one is a homolog of the *AtbZIP61* gene, which encodes a putative transcription factor and is expressed in *Arabidopsis* flowers, with especially high expression in petals. It is not known whether *AtbZIP61* regulates floral development in *Arabidopsis*. However, on the basis of its expression pattern and that of its homolog in *E. californica* we speculate that its function is downstream of the B function genes to regulate petal development.

In *Arabidopsis* C-function is controlled by *AGAMOUS*, which specifies the stamens and carpels. When compared with leaves, sepals and petals, 26 genes were preferentially expressed in carpels (compared to 168 genes in stamens). We searched for C-domain genes and found that 7 genes were expressed 2-fold significantly greater in stamens and carpels than in leaves, sepals and petals. Among them was a homolog of the *Arabidopsis* C-function gene *AG* (Riechmann and
Meyerowitz, 1997). Since both monocots (rice) and other eudicots have $AG$ homologs functioning in stamen and carpel development, we hypothesize that the $AG$ homolog in *E. californica* has similar functions (Kyozuka and Shimamoto, 2002; Lee et al., 2004b; Zahn et al., 2005a). It has been proposed that D-domain genes are required for ovule development, but only one *E. californica* gene (88769) was expressed in carpels two fold significantly higher over all other tissues. This EST did not have an identifiable *Arabidopsis* homolog.

To uncover additional candidates of A-, B- or C-domain genes, we used less stringent criteria and selected genes with expression levels at least two fold higher in each pre-anthesis reproductive tissue than in leaves (with FDR <0.05) (Figure 2-3E&F). We found that most of these genes were expressed in a whorl-specific manner and only a small numbers of genes were co-upregulated in sepals and petals, in petals and stamens, or in stamens and carpels. Furthermore, the overlap of A/B-domain and that of B/C-domain genes were even smaller (Figure 2-3E& 3F). Unlike studies in *Persea* and *Aquilegia*, whose floral transcriptomes were interpreted as support for a “fading borders” model of floral organ identity (Chanderbali et al., 2009; Voelckel et al., 2010), the *E. californica* floral transcriptomes were rather distinctive, providing a molecular explanation for the morphologically different sepals and petals. Therefore, *E. californica* might have adopted an ABC model with relatively sharp borders, similar to those found in core eudicots. Because *E. californica* is basal to *Aquilegia* within the Ranunculales, as determined by phylogenetic analyses (Moore et al., 2007), it may be that sharply defined floral organ borders represent an ancestral state for all eudicots, but has been lost in some more derived lineages.
3.6 Expression profiles of members of regulatory gene families

To gain further insights into the transcriptional activities of putative regulatory genes in floral development, we focused on gene families that are homologous to known regulators of plant development, particularly those encoding known or putative transcription factors. For convenience, we will refer to their predicted functions without using the words putative or predicted.

MADS: Genes containing a MADS-box DNA binding domain represent the best-studied floral gene family of which multiple members are crucial for floral development. In *E. californica* the expression of *EscaAG1* (84248), *EscaAG2* (86612), *EScaAGL2* (87251), *EScaAGL9* (87125), *EScaAGL11* (89484), *EScaGLO* (87167), *EScaDEF1* (83744) and *EScaDEF2* (87005) have been studied with *in situ* hybridization ([Drea et al., 2007; Kim et al., 2005; Zahn et al., 2005a; Zahn et al., 2005b]). Additionally, MADS-box genes homologous to those lacking characterized functions in *Arabidopsis* were included on our array, such as *EScaAGL54* (87912). Expression of *EScaAGL54* was highest in small buds, but showed similar levels in all the other tissues, suggesting a putative function in early floral stages.

To further understand the expression of the *E. californica* MADS-box genes, we plotted *E. californica* unigene expression profiles in comparison to the closest *Arabidopsis* homologs ([Zahn et al., 2005b]). Expression patterns were largely similar between the two species, but there were some interesting differences (Figure 2-6). Both of the *E. californica* AP3 homologs showed similar expression patterns to AP3, differing only in that 87005 (*EscaDEF2*) showed lower expression in all tissues relative to 83744 (*EscaDEF1*) or AP3 in *Arabidopsis* (Figure 2-6A). At the same time, 87167 (*EscaGLO*), a homolog of PI, showed similar expression to PI in *Arabidopsis* (Figure 2-6B). Additionally, the *E. californica* homologs of the *Arabidopsis* C-function gene AG both showed similar expression to that of AG, again with one expressed lower
than the other (Figure 2-6C). Besides those key MADS box genes regulating floral development, we found that *E. californica* homologs of E-function genes also have similar expression pattern as E-function genes in *Arabidopsis* (Figure 2-6D, E).

Homologs of other MADS box genes demonstrated different expression patterns. Unigene 84248 (*EscaAG1*, an AG homolog (Zahn et al., 2006b)) was highly expressed in stamens and carpels as expected, while 86612 (*EscaAG2*, a second AG homolog (Zahn et al., 2006b)) exhibited similar levels of expression in all floral tissues, suggesting a divergent function for this gene in *E. californica* flower development (Figure 2-6C). Also, the homolog of AGL6 (86583) also showed a higher expression in sepals and petals (Figure 2-6F), in contrast to the low expression of the *Arabidopsis* AGL6 gene in sepals on the basis of microarray expression (Schmid et al., 2005b). Since a homolog of A-function gene has not been found in *E. californica*, it is possible that 86583 may function in the outer two whorls as an A-function gene (Figure 2-6F).

AGO: The ARGONAUTE family is involved in RNA post-transcriptional regulation (Chen, 2009). In *Arabidopsis*, members of the AGO family are involved in floral development, most likely through miRNA and siRNA silencing. Our microarray included 10 members of the AGO family, all of which were differentially expressed in at least one tissue (Figure 2-7A & B). Among those genes, there was an interesting pattern, which identified three genes that were generally highly expressed in all organs while the remaining seven genes were expressed at a moderate/low level.

Among the genes examined in this study, three AGO1 homologs (1 in the high expression group and two in the low expression group) shared similar expression patterns: two fold higher expression in petals, pre-meiotic and meiotic buds than in sepals. The AGO genes in *Arabidopsis* encode proteins with a PAZ domain (with nucleic acid binding activity (Lingel et al., 2003)) and are expressed at similar levels in different tissues except PAZ-1, which was preferentially
expressed in carpel, pre-meiotic and meiotic buds compared with sepal with more than two fold changes.

**MYB:** MYB transcription factors contain DNA binding domains and some have been identified as flower developmental regulators (Martin et al., 2002; Peng, 2009b). Eleven *E. californica* MYB genes were included on our microarray. Most MYB genes showed dramatic differential expression among tissues, but two of them were not differentially expressed among any of the tissues tested (Figure 2-7C). One homolog of At4g32730 (*MYB1*) was expressed at higher levels in mature petals and stamens, suggesting that this gene may have a role in B-function. A homolog of At4g32730 (*AtMYB3R1*) was significantly preferentially expressed (> two fold higher) in the pre-meiotic bud in comparison with sepals, petals, and stamens and carpels. A homolog of At3g28470 (*AtMYB35*) was also preferentially expressed in pre-meiotic buds compared with all seven other tissues. A homolog of At4g01680 (*AtMYB55*) was significantly preferentially expressed in fruit in comparison with leaves, pre-meiotic and meiotic buds, petals, sepals and stamens. An At2g37630 (*AtMYB91/AS1*) homolog was more varied in expression but generally showed lower expression in stamens than in carpels, fruits, leaves, pre-meiotic and meiotic buds and lesser down regulation in petals relative to carpels, leaves and pre-meiotic buds. Last but not the least, a homolog of At3g61250 (*AtMYB17*) was expressed two fold significantly higher in meiotic buds compared with fruits.

**ZF-HD:** Zinc finger homeodomains are expressed during floral development in *Arabidopsis* (Tan and Irish, 2006). Our microarray contained four genes in this family. Two homologs of At1g75240 (*ATHB33*) were expressed without significant difference across all tissues. Of these two genes, one (88691) was expressed highly in both vegetative and reproductive organs while the other was barely expressed in all tissues, suggesting a functional divergence between these two paralogs (Figure 2-7D).
ARF: Auxin-response factors (ARFs) are believed to regulate auxin responsive genes (Ellis et al., 2005; Goetz et al., 2007). This family contains *ETTIN* (At2G33860), a developmental regulatory gene that acts on regional identity in the perianth, stamens and carpels (Sessions et al., 1997). Most of the poppy ARF genes that were included on our microarray showed no differential expression among the tissues examined (Figure 2-7E). Only one gene, a homolog of the At5g62000 (ARF2, 84471), showed two fold significantly differential expression: two fold lower in stamens when compared with all tissues but sepal; and two fold lower in sepals compared with carpels, fruits and pre-meiotic buds.

bZIP: The bZIP protein family contains the *Arabidopsis* *FD* (At4G35900, FD-1) and *PERIANTHIA* (At1G68640) genes, which are involved in flower development and the *HY5* (At5G11260) gene involved in root development. Our array contained 12 members of this family, one of which was not differentially expressed among all tissues examined (Figure 2-7F). From our microarray results, most of these genes showed only slightly different expression levels except the homologs of *bZIP7* (83748) and *bZIP8* (87035), both of which were expressed highly in stamens, with *bZIP8* also highly expressed in petals. Previous studies of genes in bZIP family suggested that some of them may act downstream of B-function genes to regulate floral development (Das et al., 2009; Gibalova et al., 2009; Jakoby et al., 2002). Because the homolog of *bZIP8* was co-expressed with B function genes, we speculate that this gene might have a function similar to that of the *Arabidopsis* homolog. In addition, a homolog of At4g38900 is expressed at a level two fold higher in sepals than in stamens.

bHLH: The basic helix-loop-helix family contains several *Arabidopsis* genes regulating flower development including *SPATULA*, which controls the development of the carpel margins (Heisler et al., 2001). Eleven members of this family were included on our microarray, seven of which showed no significant differential expression (Figure 2-7G). The other four genes demonstrated two fold differential expressions among tissues examined. A homolog of
At2g31210 (bHLH91, 89282) was most highly expressed in pre-meiotic buds and the expression level was at least two fold higher than in all the other tissues; and its expression level in meiotic buds was at least two fold higher than any other floral organs. Since At2g31210 has an important role in anther development in *Arabidopsis* (Zhang et al., 2006), its homolog in *E. californica* may function in a similar manner. Another gene, a homolog of At5g09460 (bHLH143), was also expressed at a higher level in the pre-meiotic buds than in sepals, petals and stamens and in meiotic buds. Additionally, this gene was expressed two fold higher in carpels and fruits than in stamens. A homolog of At1g26260 (bHLH76, CIB5) was expressed in pre-meiotic buds significantly two fold higher than in fruits and stamens. A homolog of At3g26744 (bHLH116/ICE1) was significantly down regulated by two fold in stamens relative to carpels, fruits, leaves, meiotic buds and petals. This gene was also significantly more highly expressed by two fold in petals over sepals. The expression patterns of bHLH genes suggest that they might regulate several aspects of floral development and/or physiology, but not necessarily associated with ABC functions. Further study of bHLH genes, and indeed many of the floral gene families examined here, in *Arabidopsis* and other species, including *E. californica*, may uncover their functions and reveal possible functional conservation among the eudicots.

4. Conclusions

We examined transcriptome landscapes from eight tissues of the basal eudicot *E. californica* and identified preferentially expressed genes within and among floral developmental tissues, fruits and leaves. By comparing genes showing tissue-preferential expression patterns in *E. californica*, we found that genes preferentially expressed in specific reproductive organs or at certain stages tended to have less conserved expression levels compared with *Arabidopsis* than
those preferentially expressed in leaves (table 2-1 & 2). One possible explanation is that most of the leaf-preferential genes encode highly conserved chloroplast proteins.

We also identified the co-expressed and tissue-specific floral genes and characterized the signature of ABC domain genes. Our comparison of the gene expression patterns in *E. californica, Aquilegia, Persea* and *Arabidopsis* showed that the *E. californica* results support a “sharp border” model, similar to that for core eudicot such as *Arabidopsis*, rather than the “fading border” model in other basal angiosperms (Chanderbali et al., 2009; Voelckel et al., 2010). This is consistent with the clear morphological distinction of sepals and petals, and the lack of intermediate floral organs such as staminodes in *E. californica* flowers. In contrast, *Aquilegia* flowers have similar outer perianth organs and a distinct type of floral organ between stamens and the carpels, which is in good agreement with the microarray results of the floral organs (Voelckel et al., 2010). Therefore, although both *E. californica* and *Aquilegia* are basal eudicots, the morphological and expression characteristics strongly suggested that they have divergent developmental programs, with *E. californica* more similar to core eudicots and *Aquilegia* resembling basal angiosperms. Our analysis of *E. californica* further suggested that flowers with distinct perianth organs might have originated at an earlier time than the ancestor of core eudicots. This study along with other works (Nam et al., 2004; Voelckel et al., 2010) highlight the importance of careful analysis of basal eudicots as an intermediate group of flowering plants to provide crucial information to bridge the gap between highly canalized core eudicots and morphological flexible basal angiosperms.

Our data also provided an overview of divergence and conservation between different species. The highly similar expression patterns of B- and C-function genes compared with the varied expression levels of other MADS box genes in *Arabidopsis* and *E. californica* suggested that the conserved expression of only a few key genes may result in the high similarity of flower morphology between *Arabidopsis* and *E. californica*. The transcriptome analysis of other
families with known functions in floral development indicates their possible roles in *E. californica*. Recent study of protein-protein interaction in basal eudicots (*Eupeletia pleiospermum, Akebia trifoliata* and *Pachysandra terminalis*) suggested that MADS box genes that interact with each other have co-evolved. This is most likely due to the fact that the majority of the protein-protein interactions are expected to be conserved to some extent to orchestrate floral architecture (Davies et al., 1996). However, Zhao et al. [73] showed the AP1 lineage had a distinct interaction pattern; this together with our results that *AGL6* and *SEP* homologs are expressed in the A-domain support that A-function genes show less conservation [58]. In *Arabidopsis*, *AP1* not only regulates the development of sepal and petal, but also integrates growth, patterning and hormonal pathways (Kaufmann et al., 2010). This dual function of *AP1* observed in the core eudicots might be a more recent innovation that evolved since the divergence of the core from the basal eudicots.

Many of the genes showing tissue specific expression noted in this study have homologs in *Arabidopsis* that are currently lacking in functional analyses. This study, when compared with similar studies in *Arabidopsis* and other species should help us identify genes of interest that may play important, conserved, roles in floral development (Voelckel et al., 2010; Wellmer et al., 2004a; Zhang et al., 2005c). We have identified a number of candidate genes that share similar expression patterns between *E. californica* and *Arabidopsis*, but have not been functionally characterized. Our results suggest that *E. californica* has a similar floral program to the core eudicots, despite a mostly divergent set of genes outside of the MADS-box family. These results not only indicate that different regulatory machinery may operate among basal eudicots, but that canalized floral development might have originated prior to the core eudicots. Our findings also allow for informative comparisons with other species, allowing hypotheses formulation and stimulating further experimentation in model organisms, which now includes *E. californica*. 
Figure 0-1. An angiosperm phylogram with illustration of flower structures and the loop design of the E. californica microarray experiments.

(A) A phylogram of angiosperms with flower architectures for several representative species. S represents sepal; P, petal; St, stamen; Std, staminodia; C, carpel; Ot, outer tepals; It, inner tepals. (B) We sampled from eight different tissues, including leaves, small floral buds, medium floral buds, four floral organs (sepals, petals, stamens, and pistils) at anthesis, and young fruits (four replicates for each tissue, 32 in total). Each line connects samples from two tissues in one microarray hybridization reaction, and four different colors represent four replicates of each tissue. The arrows point to the samples labeled with Cy5 dyes while base point to the samples labeled with Cy3 dyes.
Figure 0-2. Correlation coefficients between signal intensities from four biological replicates of the small floral buds.

Pearson’s correlation coefficients were between 0.94 and 0.97 between any pair of the four biological replicates, indicating that the results were highly reproducible.
Figure 0-3. Venn Diagrams of Genes expressed in reproductive tissues (the same abbreviations were used as in tables).

(A) - (D) genes expressed in different tissues and their intersections. (E) - (F) genes significantly preferentially expressed compared with leaf with more than two-fold differences and their intersections.
Figure 0-4. Heat maps and GO annotation pie chart of genes differentially expressed between any two tissues.

(A) Heat map for the mRNA profiles of 1921 genes differentially expressed between any two tissues. Red color represents high expression while green color represents low expression. HCL clustering was performed on transcripts ratios of all tissues across tissues and genes. Two major clusters had been identified as C1 and C2. L represents leaf; C represents carpel; F represents fruit; P represents petal; SB represents small bud; MB represents medium bud; S represents sepal; ST represents stamen. (B) GO categorization of all Arabidopsis homologs of Poppy genes included in our chip as control. (C) GO categorization of all Arabidopsis homologs of poppy genes that were statistically significantly differentially expressed.
Figure 0-5. Heat maps of genes preferentially expressed in different tissues. Red color represents high expression while green color represents low expression.

(A) Heat map of genes preferentially expressed in (A) leaf compared with all the other tissues, (B) sepal compared with all the other tissues, and (C) petal compared with all the other tissues. (D) stamen compared with all the other tissues. See definition of abbreviation in figure 4A.
Figure 0-6. The expression levels of MADS transcription factors families

All the expression values are log2 ratio. L represents leaf; C represents carpel; F represents fruit; P represents petal; SB represents small bud; MB represents medium bud; S represents sepal; ST represents stamen. The y-axis is the log2 ratio of gene expression levels.
Figure 0-7. The expression levels of ARGONAUTE, MYB, Zinc-finger, Homeodomain, ARF, bZIP and bHLH families.

All the expression values are log2 ratio. The same abbreviations of different tissues were used as in Figure 2-5.
Table 0-1. Genes preferentially expressed at pre-meiotic, meiotic stage and fruit in poppy.

<table>
<thead>
<tr>
<th>gene</th>
<th>BestATHit</th>
<th>L</th>
<th>SB</th>
<th>MB</th>
<th>S</th>
<th>P</th>
<th>ST</th>
<th>C</th>
<th>F</th>
<th>annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>89282</td>
<td>AT2G31210.1</td>
<td>5.3</td>
<td>9.0</td>
<td>7.1</td>
<td>5.6</td>
<td>5.3</td>
<td>5.3</td>
<td>5.3</td>
<td>5.1</td>
<td>bHLH</td>
</tr>
<tr>
<td>83967</td>
<td>AT5G16920.1</td>
<td>7.1</td>
<td>9.9</td>
<td>8.5</td>
<td>7.3</td>
<td>7.0</td>
<td>6.9</td>
<td>6.9</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td>84082</td>
<td>AT1G68540.1</td>
<td>6.8</td>
<td>10.2</td>
<td>8.9</td>
<td>6.8</td>
<td>6.9</td>
<td>6.7</td>
<td>6.5</td>
<td>6.2</td>
<td>oxidoreductase</td>
</tr>
<tr>
<td>87393</td>
<td>AT1G44970.1</td>
<td>5.1</td>
<td>7.9</td>
<td>5.9</td>
<td>5.1</td>
<td>5.0</td>
<td>5.0</td>
<td>5.6</td>
<td>5.2</td>
<td>peroxidase</td>
</tr>
<tr>
<td>86946</td>
<td>AT4G33870.1</td>
<td>7.8</td>
<td>9.5</td>
<td>8.1</td>
<td>8.0</td>
<td>7.8</td>
<td>7.9</td>
<td>7.8</td>
<td>7.8</td>
<td>peroxidase</td>
</tr>
<tr>
<td>86850</td>
<td>AT3G28470.1</td>
<td>6.2</td>
<td>7.5</td>
<td>6.4</td>
<td>6.1</td>
<td>6.1</td>
<td>6.1</td>
<td>6.1</td>
<td>6.0</td>
<td>ATMYB35</td>
</tr>
<tr>
<td>85123</td>
<td>AT5G09970.1</td>
<td>5.9</td>
<td>9.5</td>
<td>7.6</td>
<td>5.4</td>
<td>5.4</td>
<td>5.1</td>
<td>6.5</td>
<td>7.3</td>
<td>CYP78A7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Preferentially expressed in meiotic buds</td>
</tr>
<tr>
<td>84975</td>
<td>AT5G35630.2</td>
<td>6.9</td>
<td>6.7</td>
<td>8.5</td>
<td>6.8</td>
<td>6.6</td>
<td>6.7</td>
<td>6.6</td>
<td>6.9</td>
<td>GS2</td>
</tr>
<tr>
<td>85233</td>
<td>AT1G11910.1</td>
<td>5.6</td>
<td>7.4</td>
<td>10.2</td>
<td>9.1</td>
<td>6.1</td>
<td>8.5</td>
<td>6.1</td>
<td>8.4</td>
<td>aspartyl protease</td>
</tr>
<tr>
<td>86094</td>
<td>AT1G54220.1</td>
<td>6.8</td>
<td>7.8</td>
<td>9.9</td>
<td>7.5</td>
<td>7.3</td>
<td>8.6</td>
<td>7.0</td>
<td>7.2</td>
<td>dihydrolipoamide S-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>acetyltransferase</td>
</tr>
<tr>
<td>88004</td>
<td>AT4G16260.1</td>
<td>5.7</td>
<td>7.5</td>
<td>9.7</td>
<td>6.0</td>
<td>5.9</td>
<td>6.1</td>
<td>5.4</td>
<td>5.8</td>
<td>hydrolase</td>
</tr>
<tr>
<td>88092</td>
<td>AT4G12910.1</td>
<td>9.1</td>
<td>9.3</td>
<td>10.9</td>
<td>8.9</td>
<td>8.5</td>
<td>8.4</td>
<td>9.0</td>
<td>9.4</td>
<td>scpl20</td>
</tr>
<tr>
<td>88096</td>
<td>AT3G11450.1</td>
<td>7.8</td>
<td>8.2</td>
<td>9.9</td>
<td>7.8</td>
<td>7.8</td>
<td>8.2</td>
<td>7.9</td>
<td>7.9</td>
<td>cell division protein-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>related</td>
</tr>
<tr>
<td>88675</td>
<td>AT4G35160.1</td>
<td>6.3</td>
<td>6.6</td>
<td>7.9</td>
<td>6.6</td>
<td>6.3</td>
<td>6.2</td>
<td>6.1</td>
<td>6.2</td>
<td>O-methyltransferase</td>
</tr>
<tr>
<td>89901</td>
<td>AT5G03880.1</td>
<td>7.6</td>
<td>7.6</td>
<td>8.7</td>
<td>7.7</td>
<td>7.4</td>
<td>7.6</td>
<td>7.3</td>
<td>7.5</td>
<td>electron carrier</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Preferentially expressed in fruits</td>
</tr>
<tr>
<td>83998</td>
<td></td>
<td>6.4</td>
<td>5.8</td>
<td>5.7</td>
<td>6.3</td>
<td>6.5</td>
<td>5.8</td>
<td>6.2</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td>84097</td>
<td>AT5G54160.1</td>
<td>9.4</td>
<td>9.1</td>
<td>10.0</td>
<td>9.1</td>
<td>8.6</td>
<td>8.1</td>
<td>9.0</td>
<td>11.1</td>
<td>ATOMT1</td>
</tr>
<tr>
<td>86118</td>
<td>AT5G62200.1</td>
<td>7.6</td>
<td>7.0</td>
<td>7.4</td>
<td>7.6</td>
<td>7.6</td>
<td>7.7</td>
<td>7.3</td>
<td>9.3</td>
<td>embryo-specific protein</td>
</tr>
<tr>
<td>86486</td>
<td>AT1G07080.1</td>
<td>6.5</td>
<td>6.6</td>
<td>6.9</td>
<td>6.8</td>
<td>6.3</td>
<td>6.8</td>
<td>6.6</td>
<td>10.1</td>
<td>GILT</td>
</tr>
<tr>
<td>87027</td>
<td></td>
<td>5.8</td>
<td>5.5</td>
<td>5.5</td>
<td>5.7</td>
<td>5.6</td>
<td>6.0</td>
<td>5.8</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td>87195</td>
<td>AT5G12380.1</td>
<td>6.6</td>
<td>6.2</td>
<td>6.5</td>
<td>6.7</td>
<td>6.4</td>
<td>6.5</td>
<td>7.2</td>
<td>9.6</td>
<td>annexin</td>
</tr>
<tr>
<td>87830</td>
<td>AT5G08260.1</td>
<td>6.0</td>
<td>5.9</td>
<td>6.2</td>
<td>6.0</td>
<td>6.1</td>
<td>5.9</td>
<td>6.1</td>
<td>7.4</td>
<td>scpl35</td>
</tr>
<tr>
<td>88106</td>
<td>AT1G20030.2</td>
<td>6.6</td>
<td>6.3</td>
<td>6.8</td>
<td>7.3</td>
<td>5.9</td>
<td>6.4</td>
<td>6.5</td>
<td>9.0</td>
<td>pathogenesis-related</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>thaumatin</td>
</tr>
<tr>
<td>89333</td>
<td></td>
<td>8.8</td>
<td>6.5</td>
<td>8.0</td>
<td>8.4</td>
<td>5.9</td>
<td>7.6</td>
<td>7.1</td>
<td>10.5</td>
<td></td>
</tr>
</tbody>
</table>

The first column is the gene number for gene represented by poppy ESTs. The second column is the closest *Arabidopsis* homolog of poppy genes. All expression values are log2 ratio. L represents leaf; C represents carpel; F represents fruit; P represents petal; SB represents small bud; MB represents medium bud; S represents sepal; ST represents stamen. Annotations are from TAIR version 9.
Table 0-2. Expression levels of putative ABC genes in poppy.

<table>
<thead>
<tr>
<th>gene</th>
<th>BestATHit</th>
<th>L</th>
<th>SB</th>
<th>MB</th>
<th>S</th>
<th>P</th>
<th>ST</th>
<th>C</th>
<th>F</th>
<th>annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A function genes</td>
</tr>
<tr>
<td>84392</td>
<td>AT2G36830.1</td>
<td>14.4</td>
<td>13.9</td>
<td>14.3</td>
<td>16.2</td>
<td>16.3</td>
<td>15.1</td>
<td>14.5</td>
<td>14.4</td>
<td>GAMMA-TIP</td>
</tr>
<tr>
<td>86583</td>
<td>AT2G45650.1</td>
<td>6.4</td>
<td>10.1</td>
<td>10.4</td>
<td>11.9</td>
<td>10.7</td>
<td>7.0</td>
<td>8.9</td>
<td>8.9</td>
<td>AGL6</td>
</tr>
<tr>
<td>87043</td>
<td>AT3G05490.1</td>
<td>8.9</td>
<td>9.1</td>
<td>9.8</td>
<td>10.6</td>
<td>10.9</td>
<td>9.6</td>
<td>9.2</td>
<td>9.3</td>
<td>RALFL22</td>
</tr>
<tr>
<td>87251</td>
<td>AT5G15800.1</td>
<td>6.3</td>
<td>9.2</td>
<td>9.4</td>
<td>10.5</td>
<td>9.5</td>
<td>6.9</td>
<td>8.4</td>
<td>8.3</td>
<td>SEP1, AGL2</td>
</tr>
<tr>
<td>85671</td>
<td></td>
<td>7.3</td>
<td>7.0</td>
<td>6.9</td>
<td>10.5</td>
<td>11.1</td>
<td>8.4</td>
<td>7.3</td>
<td>7.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B function genes</td>
</tr>
<tr>
<td>83744</td>
<td>AT3G54340.1</td>
<td>8.2</td>
<td>11.4</td>
<td>12</td>
<td>9.1</td>
<td>11.9</td>
<td>13.1</td>
<td>9.9</td>
<td>9.1</td>
<td>AP3</td>
</tr>
<tr>
<td>83763</td>
<td>AT1G69500.1</td>
<td>5.2</td>
<td>5.7</td>
<td>6.1</td>
<td>5.9</td>
<td>7.7</td>
<td>7.3</td>
<td>6.1</td>
<td>5.7</td>
<td>electron carrier</td>
</tr>
<tr>
<td>83991</td>
<td>AT5G19770.1</td>
<td>10.0</td>
<td>10.1</td>
<td>10.3</td>
<td>9.0</td>
<td>11</td>
<td>11.2</td>
<td>9.8</td>
<td>10.1</td>
<td>TUA3</td>
</tr>
<tr>
<td>84789</td>
<td>AT5G64250.2</td>
<td>11.9</td>
<td>11.4</td>
<td>13.2</td>
<td>13.8</td>
<td>15.6</td>
<td>15.0</td>
<td>13.9</td>
<td>13.2</td>
<td>2-nitropropane dioxygenase</td>
</tr>
<tr>
<td>85140</td>
<td>AT2G27510.1</td>
<td>9.2</td>
<td>11.1</td>
<td>12.2</td>
<td>10.8</td>
<td>13.8</td>
<td>11.9</td>
<td>10.5</td>
<td>10.0</td>
<td>ferredoxin 3</td>
</tr>
<tr>
<td>85166</td>
<td>AT5G62690.1</td>
<td>9.2</td>
<td>10.0</td>
<td>10.2</td>
<td>8.3</td>
<td>10.4</td>
<td>10.9</td>
<td>9.3</td>
<td>9.8</td>
<td>TUB2</td>
</tr>
<tr>
<td>85610</td>
<td>AT4G36250.1</td>
<td>5.4</td>
<td>6.5</td>
<td>8.5</td>
<td>6.4</td>
<td>7.8</td>
<td>7.6</td>
<td>6.1</td>
<td>5.5</td>
<td>Aldehyde Dehydrogenase 3F1</td>
</tr>
<tr>
<td>87005</td>
<td>AT3G54340.1</td>
<td>4.6</td>
<td>7.8</td>
<td>8.3</td>
<td>6.0</td>
<td>10.2</td>
<td>8.1</td>
<td>5.3</td>
<td>5.1</td>
<td>AP3</td>
</tr>
<tr>
<td>87035</td>
<td>AT3G58120.1</td>
<td>5.6</td>
<td>5.5</td>
<td>5.7</td>
<td>5.5</td>
<td>7.9</td>
<td>8.0</td>
<td>5.1</td>
<td>5.2</td>
<td>ATBZIP61</td>
</tr>
<tr>
<td>87167</td>
<td>AT5G20240.1</td>
<td>7.3</td>
<td>11.2</td>
<td>12.0</td>
<td>9.2</td>
<td>12.8</td>
<td>11.9</td>
<td>8.4</td>
<td>8.1</td>
<td>PI</td>
</tr>
<tr>
<td>87294</td>
<td>AT5G03690.2</td>
<td>8.0</td>
<td>9.3</td>
<td>10.0</td>
<td>7.8</td>
<td>9.8</td>
<td>10.0</td>
<td>8.7</td>
<td>9.2</td>
<td>fructose-bisphosphate aldolase</td>
</tr>
<tr>
<td>89750</td>
<td>AT4G37990.1</td>
<td>8.2</td>
<td>8.8</td>
<td>9.6</td>
<td>8.5</td>
<td>11.4</td>
<td>10.2</td>
<td>7.7</td>
<td>7.9</td>
<td>mannitol dehydrogenase</td>
</tr>
<tr>
<td>89805</td>
<td>AT5G66310.1</td>
<td>5.9</td>
<td>6.5</td>
<td>6.8</td>
<td>5.3</td>
<td>7.2</td>
<td>7.8</td>
<td>6.1</td>
<td>6.4</td>
<td>kinesin motor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C function genes</td>
</tr>
<tr>
<td>84248</td>
<td>AT4G18960.1</td>
<td>6.7</td>
<td>10.6</td>
<td>11.1</td>
<td>7.2</td>
<td>6.6</td>
<td>11.5</td>
<td>11.6</td>
<td>11.6</td>
<td>AG</td>
</tr>
<tr>
<td>84252</td>
<td>AT4G26220.1</td>
<td>7.3</td>
<td>10.9</td>
<td>10.9</td>
<td>7.0</td>
<td>7.2</td>
<td>10.9</td>
<td>10.1</td>
<td>6.6</td>
<td>caffeoyl-CoA 3-O-methyltransferase</td>
</tr>
<tr>
<td>84340</td>
<td>AT3G44260.1</td>
<td>7.9</td>
<td>8.4</td>
<td>8.3</td>
<td>7.8</td>
<td>8</td>
<td>9.9</td>
<td>9.2</td>
<td>8.6</td>
<td>CCR4-NOT transcription complex protein</td>
</tr>
<tr>
<td>84512</td>
<td>AT1G1910.1</td>
<td>7.2</td>
<td>9.5</td>
<td>10.1</td>
<td>7.0</td>
<td>7.2</td>
<td>8.8</td>
<td>9.1</td>
<td>9.1</td>
<td>aspartyl protease</td>
</tr>
<tr>
<td>84691</td>
<td>AT2G44480.1</td>
<td>9.1</td>
<td>12.4</td>
<td>12.8</td>
<td>8.7</td>
<td>8.6</td>
<td>12.6</td>
<td>12</td>
<td>12.9</td>
<td>BETA GLUCOSIDASE 17</td>
</tr>
<tr>
<td>89115</td>
<td>AT3G20240.1</td>
<td>6.4</td>
<td>7.4</td>
<td>7.3</td>
<td>6.3</td>
<td>6.2</td>
<td>8.2</td>
<td>7.6</td>
<td>7.0</td>
<td>mitochondrial substrate carrier</td>
</tr>
<tr>
<td>89980</td>
<td>AT1G35720.1</td>
<td>7.1</td>
<td>8.8</td>
<td>9.5</td>
<td>7.2</td>
<td>7.9</td>
<td>10.1</td>
<td>9.2</td>
<td>8.5</td>
<td>ANNEXIN ARABIDOPSIS 1</td>
</tr>
</tbody>
</table>

Column sequence, abbreviation and the version of annotation are as those used as in Table 2-1. All the expression values are log2 ratio.
5. Materials and Methods

5.1 Tissue collection and RNA isolation

Sixteen *E. californica* cv. ‘Aurantica Orange’ (J.L. Hudson Seedsman) plants were grown from seeds in a controlled greenhouse environment at the Pennsylvania State University (University Park, PA) under 16 hours light and watered and fertilized as needed. To avoid potential expression differences among collections due to circadian rhythms leaves, floral tissues were only collected from individual plants between 8:30-10:30 AM. Developing leaves of less than 5mm length, developing fruits, pre-meiotic (small) buds less than 5mm long, meiotic (medium) buds of 5-10mm length and pre-anthesis sepals, petals, stamens and carpels were collected from 16 plants, immediately placed in liquid nitrogen and stored in a -80°C freezer until RNA extraction. Tissues from a group of four plants were then pooled to create one biological replicate, for a total of four replicates.

5.2 Probe design for the *E. californica* transcriptome

To design oligonucleotide probes for *E. californica*, a two-stage pipeline for oligonucleotide probe design, Microarray Oligonucleotide Design and Integration Tool (MODIT) was used. Briefly, MODIT integrates two existing programs: Array Oligo Selector (6) (AOS) and OligoArray (8) (OA), with subsequent independent evaluation and optimization steps. The pipeline enables one to design a set of probes having well-defined sequence and thermodynamic properties by first taking advantage of the strict thermodynamic criteria of OA, to produce a partial set of optimized probes, and then fills in the set from among the large number of probes selected by AOS, after screening them for thermodynamic compatibility.
The MODIT pipeline screens candidate probes based on three parameters: high sequence specificity, appropriate melting temperature $T_m$, and lack of stable secondary structure. The first criterion, sequence specificity, was determined using BLAST and Smith-Waterman local alignment tools to eliminate probes having a match to any non-target sequences of more than 15 consecutive nucleotides, or an overall match of more than 30 nucleotides (Kane et al., 2000; Li and Stormo, 2001; Relogio et al., 2002). The second criterion was that the probe set should have very similar $T_m$. The MODIT user is informed of probes with $T_m$ outside a recommended range by flagging in the database, and she/he can decide whether to use such probes. A third criterion was the lack of stable secondary structure. MODIT allows values of probe $\Delta G_{SS}$ above -0.5 kcal.mol$^{-1}$, less than the energy of one hydrogen bond between bases (Bloomfield et al., 2000).

We use melting temperature to independently recalculate a consistent set of thermodynamic properties for the probes and check for consistency (Le Novere, 2001). The pipeline stores comprehensive information about probe thermodynamic properties and potential cross reactions in a MySQL database, so that they can subsequently be used in array data analysis.

The MODIT pipeline was used to generate one 60 base probe for each gene in the 6,846 $E. \text{californica}$ Unigene set (Wang et al., 2004; Wang et al., 2008b), after masking regions that were conserved in multigene families in Arabidopsis, rice ($Oryza$) and Populus. Unigenes were sorted into gene families using PlantTribes (Wall et al., 2008b) and conserved sites in the multiple sequence alignment were identified using the column score metric calculated by CLUSTAL (Thompson et al., 1997). Sodium concentration of 0.5 M was used in modeling of thermodynamic properties, following hybridization conditions recommended by Agilent for their 60-mer Arabidopsis Oligo Microarray Kit, and the conditions modeled by Lee et al. (Lee et al., 2004a). The probe concentration range that was used in the thermodynamics calculations is 2.44 mM following the calculations of Riccelli et al. and assuming the default 1 nM target recommended in (Chou et al., 2004; Riccelli et al., 2001). In the OA run, duplex melting
temperature $T_m$ was constrained above 70 °C, and the duplex $T_m$ for predicted cross-reactions and stable secondary structures was constrained below 60 °C. For the AOS run, the constraint on GC content was maintained around 52%. Duplex melting temperature was constrained to keep 20 °C separation between the upper and lower $T_m$ limits, to allow for selection of more candidate probes. The probe maximum and minimum match for non-target sequences were maintained at 15 and 10 nucleotides, respectively. When the two sets of probes were merged, the constraints applied to the merged set were: 80 °C ≤$T_m$ ≤ 90 °C, overall match with non-target as well as with consensus sequences should be less than 30 nucleotides and $\Delta G_{SS}$ above -0.5 kcal.mol$^{-1}$. Since one goal of this design was to obtain complete coverage of all target sequences, a selection of known suboptimal probes was added back to the final design, and their sequence and thermodynamic properties tracked in the MODIT database. The design results obtained using MODIT for the target sequences from *E. californica* are summarized in Table 3. No application, including MODIT, could provide 100% target coverage while satisfying all of the design criteria for each probe. However, MODIT improved on target coverage and significantly limited potential cross reactions relative to OA, while nearly eliminating probes which were predicted to form stable secondary structure.

Oligonucleotides of 60 base pair length were designed from 6446 *E. californica* unigenes obtained from a floral EST library (Carlson et al., 2006) and cell culture suspension library (Lin et al., 2009). Unigene builds were performed as described by Carlson et al (2006) and the sorted into putative gene families using the PlantTribes database (Carlson et al., 2006; Wall et al., 2008). Because the complete genome of *E. californica* is not yet sequenced, oligos were designed to specifically exclude conserved regions, when identified, so that expression analyses putatively represent single genes (see above). Oligonucleotide probes were arrayed on glass slides by Agilent (La Jolla, CA, USA).
5.3 RNA extraction, microarray hybridization and scanning

RNA was isolated from eight tissues examined each with four biological replicate pools and cleaned using the RNeasy plant mini Kit (Qiagen, USA) following Agilent’s instructions. RNA concentrations were quantified using an Agilent 2100 Bioanalyzer and stored at -80°C before use, with yields of 20-35 micrograms of total RNAs from ~100 mg of tissues. Approximately 400 ng of total RNAs were used for cRNA synthesis with Cyanine 3-dCTP and Cyanine5-dCTP (Perkin–Elmer Life Sciences, Inc., USA) incorporation, using the Agilent Low RNA Input Kit (Agilent, USA), according to the manufacturers’ protocol. Qiagen’s RNeasy mini spin columns (Qiagen, USA) were used to purify amplified cRNA samples. Sample concentrations and were quantified using a NanoDrop spectrometer (NanoDrop Technologies, USA). Hybridization was performed using the In situ Hybridization Kit (Agilent, USA) with 35ng of Cy3 and Cy5 labeled cRNA followed the instructions at 65°C for 17h. Prior to scanning, each slide was washed, rinsed and dried in Agilent’s Stabilization and Drying Solution, as directed. Scanning was performed using a Gene Pix 4000A scanner and the Gene Pix Pro 3.0.6 Software (Axon Instruments (now Molecular Devices, USA) to produce two TIFF images at 532nm and 635nm. The microarray data were submitted to the GEO database, with accession number GSE24237.

5.4 Statistical analyses of genes differentially expressed among tissues and developmental stages

Analyses were performed with the R programming language and the limma package Bioconductor. Arrays were background corrected and loess normalized within arrays and Aq normalized between arrays (Yang and Thorne, 2003). Agilent controls and other control probes were removed from the data. For the 93 E. californica oligos with multiple probes, we chose the
probe with the highest 75% quantile value among the normalized “A” intensities of all 16 arrays. A one-way single-channel empirical Bayes ANOVA was used to identify those genes (Smyth, 2004; Smyth, 2006) that were significantly differentially expressed among the seven floral RNAs and one leaf RNA examined, with an FDR of 0.05. Additionally, significant differences between combinations of more than one floral organ and leaf were also identified under the same parameters.

In order to identify those genes that were most likely to be organ/stage specific in *E. californica*, we examined those genes with a significantly (FDR=0.05) two fold greater expression in a single organ/stage relative to all other tissues stages examined. The expression of these genes was then compared to the expression, as determined by Affymetrix arrays (Zhang et al., 2005), to their closest identified *Arabidopsis* homolog based on a tribe-MCL analysis, when available, to determine which genes may have conserved expression profiles. We were able to directly compare expression in pre-meiotic and meiotic buds in *E. californica* vs. inflorescences containing stage 1-9 flowers in *Arabidopsis* (developing inflorescences), the *E. californica* fruit, capsules, vs. the *Arabidopsis* fruit, siliques, *Arabidopsis* flowers at stage twelve nearing pre-anthesis vs. sepals, petals stamens and carpels at anthesis in *E. californica* and genes preferentially expressed in leaves in both organisms.

### 5.5 Real-time PCR experiments

To test the reliability of our microarray hybridizations, nine genes and one reference were investigated using Quantitative Real-Time PCR. RNA (1ug) of each tissue was treated with DNase (Invitrogen, USA), followed by reverse transcription using the Superscript III reverse transcriptase (Invitrogen, USA). We then performed real time PCR using DyNAmo SYBR Green qPCR Kit from New England Biolabs (NEB, USA) under the following parameters: 95°C 10min,
40 cycle for 95°C 30 sec, 60°C for 1min. Fluorescence intensity was measured using ABI’s 7300 Sequence Detection System. Eca_2514 (Unigene84142) was chosen as the reference gene as it was not significantly differentially expressed among any of our examined tissues in the microarray experiments and it was expressed at a moderate level in all our tissues compared to all other genes. The relative amounts of cRNA converted from a messenger RNA was calculated using intensities corresponding to “experimental” genes relative to the reference gene. We performed triplicate reactions for all tissues with samples containing no reverse transcriptase and no RNA as negative controls.
6. Acknowledgements

We would like to acknowledge Philip Larkin and Toni Kutchin, for providing EST data included in the microarray designs and thank Xiaofan Zhou and Dihong Lu for comments on the manuscript. This work was supported by the Floral Genome Project (NSF NSF Plant Genome Award DBI-0115684) and Ancestral Angiosperm Genome Project (NSF Plant Genome Comparative Sequencing DEB-0638595) to CWD, HM, and JLM. RG was supported by an NIH grant (R01-GM072619) and HM was also supported by funds from Fudan University.


www.arabidopsis.org
www.bioconductor.org
www.r-project.org


CHAPTER 3

AMS-DEPENDENT AND INDEPENDENT REGULATION OF ANther TRANSCRIPTOME AND COMPARISON WITH THOSE AFFECTED BY OTHER ARABIDOPSIS ANther GENES

The work described in this chapter has been published in BMC Plant Biology 2012; 12:23.

Baomin Feng, a former PhD student in the Ma lab, performed tissue collection and RNA isolation. The Penn State University Microarray Facility carried out the microarray experiments. I performed all the data analysis, real-time RT-PCR experiments and wrote the manuscript.
1. Abstract

In flowering plants, the development of male reproductive organs is controlled precisely to achieve successful fertilization and reproduction. Despite the increasing knowledge of genes that contribute to anther development, the regulatory mechanisms controlling this process are still unclear.

In this study, we analyzed the transcriptome profiles of early anthers of sterile mutants aborted microspores (ams) and found that 1368 genes were differentially expressed in ams compared to wild type anthers, affecting metabolism, transportation, ubiquitination and stress response. Moreover, the lack of significant enrichment of potential AMS binding sites (E-box) in the promoters of differentially expressed genes suggests both direct and indirect regulation for AMS-dependent regulation of anther transcriptome involving other transcription factors. Combining ams transcriptome profiles with those of two other sterile mutants, spl/nzz and ems1/exs, expression of 3058 genes were altered in at least one mutant. Our investigation of expression patterns of major transcription factor families, such as bHLH, MYB and MADS, suggested that some closely related homologs of known anther developmental genes might also have similar functions. Additionally, comparison of expression levels of genes in different organs suggested that anther-preferential genes could play important roles in anther development.

Analysis of ams anther transcriptome and its comparison with those of spl/nzz and ems1/exs anthers uncovered overlapping and distinct sets of regulated genes, including those encoding transcription factors and other proteins. These results support an expanded regulatory network for early anther development, providing a series of hypotheses for future experimentation.
2. Introduction

In flowering plants, male reproductive organs are called stamens, each of which consists of a filament and an anther (Ge et al., 2010). Cells in the anther undergo meiosis to produce microspores, which further develop into mature pollen grains (Scott et al., 2004). Therefore, anther development is critical to achieve pollen formation and subsequent success of fertilization (Goldberg et al., 1993b; Ma, 2005; Ma and Sundaresan, 2010; McCormick, 1993). According to morphological features, anther development can be grouped into two phases and then be further divided into 14 anther stages (Chang et al., 2011; Ma, 2005; Sanders et al., 1999). At the beginning of phase 1 (anther stages 1 to 8), the stamen primordium has 3 layers, L1-L3 from surface to interior. The L1 cells later become the epidermis and the L3 cells give rise to the vascular and connective tissues. Some of the L2 cells develop into archesporial cells which then divide into parietal cells and primary sporogenous cells. Additional cell division and differentiation in the L2-lineage establish a characteristic four-lobed structure at anther stage 5. Each lobe consists of central pollen mother cells surrounded by outer endothecium, middle layer and inner tapetum. Pollen mother cells undergo meiosis at stage 5-6, producing tetrads at stage 7. Dissolution of the tetrad callose wall releases microspores at stage 8. In phase 2, the microspores undergo mitosis and develop into mature pollen grains during stages 9-12. Meanwhile, pollen wall materials are deposited from both the microspores and the tapetum layer. After the degeneration of tapetum, the mature pollen is released and is able to start pollination.

Previous studies indicated that early anther development depends on transcriptional regulation and cell-cell communication (Chang et al., 2011; Coen, 2001a; Ma, 2005; Sanders et al., 1999). The SPOROCYTELESS (SPL)/NOZZLE (NZZ) gene is one of the earliest genes that
regulate anther cell fate determination (Schieflhaler et al., 1999a; Yang et al., 1999a). SPL/NZZ is activated by AG, a C function gene in the ABC model (Ito et al., 2004b; Jack, 2004; Yanofsky et al., 1990b). SPL/NZZ is expressed as early as anther stage 2-5 and a mutation in SPL/NZZ leads to the failure of differentiation of parietal and sporogenous cells, and consequentially blocks the formation of anther wall and microsporocytes (Hord et al., 2006a; Yang et al., 2003a).

EXCESS MALE SPOROCYTES1 (EMS1) and TAPETUM DETERMINANT1 (TPD1) are also essential for male fertility with a later expression peak at stage 5 (Zhao et al., 2002). EMS1 is a leucine-rich repeat receptor-like protein kinase (LRR-RLKs) and TPD1 is likely its ligand (Jia et al., 2008; Yang et al., 2005; Yang et al., 2003a). In both ems1 and tpd1 mutants, anthers produce more microsporocytes at the expense of the tapetum, indicating that communication between adjacent cell layers determines the cell fate of archesporial cell progenies in order to form normal anther wall (Zhao et al., 2002). Besides EMS1 and TPD1, other cell-cell communication-related genes are also involved in anther development, such as SOMATIC EMBRYOGENESIS RECEPTORLIKE KINASES1/2 (SERK1/2), and RECEPTORLIKE PROTEIN KINASE2 (RPK2) (Colcombet et al., 2005a; Mizuno et al., 2007b).

Upon the formation of the anther lobes, DYSFUNCTIONAL TAPETUM1 (DYT1) and AMS, encoding two bHLH transcription factors, are required for tapetal functions at subsequent stages (Sorensen et al., 2003a; Zhang et al., 2006b). In dyt1, tapetum cells harbor enlarged vacuoles and reduced cytoplasm. The dyt1 meiocytes have comparatively thin callose walls, cannot complete cytokinesis and finally collapse. RNA in situ hybridization experiments showed that DYT1 reaches its peak expression at anther stage 5 to 6 (Zhang et al., 2006b). AMS functions near the time of meiosis, slightly later than that of DYT1. In the ams mutant, the microsporocytes can complete meiosis but the tapetum cells prematurely collapse and microspores are degraded.
before the first pollen mitosis (Sorensen et al., 2003a). Beside these regulators, a large number of other genes are also expressed in the anther, and mutations in some of them lead to male sterility by affecting early anther cell formation, tapetum formation, meiosis or pollen maturation (Boavida et al., 2005; Hord et al., 2006a; Ma, 2005; Sanders et al., 1999; Wilson et al., 2001a; Wilson and Zhang, 2009b; Zhang et al., 2007b; Zhu et al., 2008a).

However, due to the functional redundancy of members of many gene families, the subtleties of the phenotypes of single-gene mutants, and possible early phenotypes that obscure anther function, forward genetics has limitations in uncovering anther gene functions (Cutler and McCourt, 2005). Expression profiling has become increasingly informative and might circumvent the limitation of forward genetics. In recent years, global gene expression profiling by microarray has been used to detect floral gene expression and obtain clues for understanding reproductive development. However, most studies to investigate stamen expression profiles have been conducted by analyzing transcripts from the whole inflorescences of male sterile mutants (Alves-Ferreira et al., 2007b; Honys and Twell, 2003; Hsu et al., 2008; Wellmer et al., 2004b; Wijeratne et al., 2007b; Xu et al., 2010), rather than the anther itself (Wijeratne et al., 2007b). Little transcriptomic information about specific organs is currently available, especially for Arabidopsis whose male reproductive organs are quite tiny (Hirano et al., 2008; Honys and Twell, 2003; Wijeratne et al., 2007b). Thus the detection of anther-specific or preferential genes in mixed floral tissues might be hampered by the moderate detection sensitivity of microarray technology. As mentioned above, SPL, EMS1 and AMS have important functions at different stages of anther development, although they have temporal overlap of expression (Sorensen et al., 2003a; Yang et al., 1999a; Zhang et al., 2006b; Zhao et al., 2002). Therefore, analysis of their shared and distinct effects on the anther transcriptome can shed some light on gene regulatory networks (Aoki et al., 2007; Khandelwal et al., 2008; Mentzen and Wurtele, 2008).
To obtain more information on transcriptomes near the stage of meiosis, we collected anthers at stage 4 to 7 from *ams* mutants and wild-type *Arabidopsis*, even though it is time consuming and technically difficult to dissect developing anthers, because we wanted to identify the genes affected by the *ams* mutation that might be too diluted to detect using RNAs from whole-inflorescences. The *ams* transcriptome data and comparison with previous data from *spl* and *ems1* anthers (Wijeratne et al., 2007b) provide detailed information on early anther development. Additionally, with known information of other floral organs in *Arabidopsis*, we identified genes that function during early anther stage around meiosis. We found that many transcription factor genes were preferentially expressed during early anther development, such as *bHLH*, *MYB*, and *MADS*. Closely related homologs were hypothesized to have either redundant or divergent functions according to phylogenic studies (Li et al., 2006; Matus et al., 2008; Wu et al., 2005). Moreover, further investigation of organ-specific transcriptome revealed the importance of both anther-specific and non-specific transcription factors in early anther development. We propose an expanded gene regulatory network that contributes to the precise regulation of temporal and spatial events during early anther development.
3. Results and Discussion

3.1 Identification of genes regulated by AMS

To characterize genes involved in tapetum development and function near the time of meiosis, we isolated total RNA of stage 4-7 anthers from wild-type and the *ams* mutant plants for Affymatrix ATH1 microarray analysis. We included three biological replicates for each genotype and the results are highly reproducible (with correlation coefficients higher than 0.96). We identified 1,368 genes that were differentially expressed in *ams* compared with wild-type anthers with at least 2-fold differences (P<0.05) (Wijeratne et al., 2007b; Zhang et al., 2005a). The scatter-plot of the 1,368 genes shows that they include genes expressed at different levels (Fig. 3-1A, B); furthermore, genes with higher expression in *ams* than wild-type tend to have low wild type expression, whereas those with lower than normal expression in *ams* tend to be expressed at higher levels (Fig. 3-1B).

Recently, Xu et al. reported totally 549 genes that are differentially expressed in *ams* floral buds compared with wild-type buds, at four different stages using two color arrays, including 134 genes that were differentially expressed near the time of meiosis (Xu et al., 2010). Among the 1,368 genes identified in our study, 90 were also identified by Xu et al. in floral buds (Fig. 1C). Because *AMS* is expressed from near anther stage 6 (meiosis) through the formation of microspores, our samples from early stage anthers allowed an examination of the early AMS function in regulating transcriptome and sensitive detection of expression shifts without dilution by other floral tissues, resulting in the identification of additional 1,278 genes (478 down- and 800 up-regulated in the *ams* anthers) with differential expression between wild-type and *ams* anthers (Fig. 3-1C).
Nevertheless, our results and the previous study did both detect 90 genes that are significantly affected by the *ams* mutation (Fig. 3-1C) (Xu et al., 2010). Some of these genes show the same direction in expression shifts between the two studies; however, others had the opposite directions (Fig. 3-1C). Specifically, 34 genes with higher expression in the *ams* anther than the wild-type anther had reduced expression levels in the *ams* inflorescences compared with the wild-type inflorescence; 9 genes showed the opposite trend. These differences might be due to the difference of sampling anther vs. flower buds that included later stages, although other possibilities cannot be ruled out. We observed more similar expression pattern between our anther transcriptome and the published flower bud transcriptome at meiosis stage. 172 out of the 519 genes down-regulated in *ams* were expressed significantly higher in wild type anther than inflorescences, while 102 of the 849 up-regulated genes showed this pattern (P-value <0.05, Fig. 3-1D), suggesting that preferential anther expression contributed to the difference between the two studies. It is also possible that the loss of *AMS* function might affect other aspects of flower development than anther development, although not revealed by phenotypic changes.

The GO categorization analysis of our anther transcriptome results showed that categories of enzymes, transporters, structural and other molecular proteins were over-represented in the genes with reduced expression, and hydrolases in those with elevated expression levels in *ams* compared with wild-type (Fig. 3-2A-C). To further investigate the putative functions of genes with different expression patterns in the *ams* anther from inflorescences, we then applied GO categorization to all the newly found differentially expressed genes in the *ams* anther. We found that some categories were enriched in those with reduced expression levels in the *ams* anther, such as structural molecules, transporters, oxidoreductases. These categories are associated with metabolic activities that are very dynamic in tapetum, suggesting a positive role of AMS in regulating metabolic functions in the tapetum. Meanwhile, genes related to ion
binding, glycosyl-transferase and hydrolase activities were enriched among the genes activated in *ams*.

As a putative bHLH transcription factor, AMS has the ability to bind to the canonical bHLH binding site (E-box: CANNTG) *in vitro* and *in vivo* (Xu et al., 2010). In order to find candidate AMS target genes, we searched E-box elements within 1 kb upstream sequences of genes with statistically significant differential expression between *ams* and wild-type anthers (Fig. 3-3A&B). We did not find statistically significant interaction between the number of E-boxes in the putative promoter regions and the fold change in gene expression (compared with randomly selected not differentially expressed genes on the chip). It is possible that active AMS binding sites are located not just in the 1-kb regions being analyzed, but also in regions further upstream or even downstream of the coding region. It is also possible that a number of the genes affected in the *ams* anthers are indirectly regulated by AMS, hence not containing AMS-binding sites in their promoters.

3.2 AMS affects genes with putative functions in phosphorylation, exocytosis, stress-response and ubiquitin-proteasome pathways during male reproduction

Both somatic and reproductive cells are evidently affected in the *ams* mutant anther, morphologically and transcriptomically (Sorensen et al., 2003a; Xu et al., 2010). Specifically, the *ams* inflorescence showed reduced expression of genes predicted to be involved in metabolism, such as lipid synthesis-related genes (Xu et al., 2010). Our anther transcriptome data provided spatially more specific information for the expression patterns of metabolism-related genes and showed that the expression levels of genes involved in cell wall formation, lipid synthesis and secondary metabolism were obviously altered in the *ams* anther, consistent with morphological defects.
Interestingly, 32 genes located on chloroplast DNA were reduced in expression in the *ams* anthers whereas starch and sucrose related genes were increased. In addition, more metabolism-related genes were found with shifted expression, especially glycosyl-transferase (P<0.01). Besides, the expression levels of genes with putative regulatory functions were also changed, such as kinases and transcription factors. Interestingly, most of the genes encoding kinases with expression shifts were activated in the *ams* mutant, suggesting a putative negative regulatory role of AMS.

In addition, we found that genes likely involved in vesicular transport were up-regulated in *ams*, including genes encoding two SNARE proteins and others related to this process: 3 syntaxins, 3 myosin heavy chains and 2 clathrin proteins. Intracellular trafficking machinery such as SNARE complex is important in animal and plant development (Mohrmann et al., 2010; Zipfel, 2008); for example, one SNARE protein, SEC22, is preferentially expressed in the flower and essential for gametophyte development (Jurgens et al., 2011). Other vesicular transport genes, such as *ArVAM3* encoding a syntaxin-related protein, were shown to function in vacuolar assembly in *Arabidopsis* (Sato et al., 1997). It is possible that the higher than normal expression of genes for vesicular transport contributes to the abnormally vacuolated tapetal cells observed in the *ams* anther (Sorensen et al., 2003a).

We also found that the expression levels of stress-responsive genes were changed in *ams*, especially the increased expression of 10 disease resistance genes and two genes encoding respiratory burst oxidases. These findings are consistent with recent studies showing that multiple abiotic stresses can lead to male sterility, such as extreme temperatures and drought (Chen, 2006; Chow and McCourt, 2006; Yang et al., 2009). In addition, some stress-inducible and/or hormone-related genes were also found with expression alteration, including *RD22*, an ABA-inducible gene responsive to dehydration; *VSP1*, a JA-inducible gene; *EPS1*, a gene possibly act upstream of SA; *CCR1*, a cold inducible gene; four disease resistance genes encoding TIR-NBS-LRR class
proteins; and three heat-shock genes, suggesting complex interactions between internal and external signals regulating anther development and/or functions (Peng, 2009).

Another regulatory pathway activated in *ams* is the ubiquitin-proteasome pathway, with increased expression of genes encoding subunits of the E3 ubiquitin ligases (Moon et al., 2004). Previous studies demonstrated essential roles of the ubiquitin-proteasome pathway in embryogenesis, hormone signaling, light response, floral development, self-incompatibility, and senescence (Chow and McCourt, 2006; Meng et al., 2010; Moon et al., 2004). Our results suggested that this pathway may also be regulated by AMS. It is possible that AMS directly regulates the expression of some genes in the ubiquitin-proteasome pathway; alternatively, AMS could influence the expression of such genes indirectly either via AMS-target genes or possibly through the accumulation of damaged proteins which then induce the ubiquitin-proteasome pathway (van Wijk and Timmers, 2010). Further experiments are needed to test these hypotheses.

### 3.3 Anther-specific or preferential genes were over-represented among genes differentially expressed in the *ams* mutant

Differential expression patterns in vegetative and floral organs can provide clues about gene functions (Zhang et al., 2005a). To find out the relationship between the gene expression shifts in the *ams* mutant and their expression preferences in different organs, we compared our data from wild-type anther with previous microarray data from roots, stems, leaves, seedlings, siliques and inflorescences. The same RNA extraction method and ATH1 platform were applied in both studies so the datasets should be comparable (Zhang et al., 2005a). We defined as anther-specific (A-S) using these criteria: 1) the expression in anther is significantly higher than in any other tissue (with FDR<0.05); 2) the expression is present in anther but not in any other tissues according to two alternative methods (see materials and methods for details and explanations)
(Zhang et al., 2005a). Using the presence call of the MAS5 algorithm identified 124 A-S genes, 76 of which had at least two fold difference; using expression level of 50 as threshold identified 172 A-S genes, 146 of which had at least two fold difference. Because both methods for calling “presence” have limitations, only the 43 genes detected by both methods were discussed as A-S gene (this rule also applied to the two groups described below).

Genes were defined as anther-preferential (A-P) if the expression in the anther is: 1) significantly higher than those in any other tissue with FDR<0.05; and 2) present in anther according to the MAS5 algorithm or with expression level of at least 50. Therefore, A-P genes included A-S genes. In addition, those with statistically significantly higher expression levels in anther than in non-floral organs were called reproductive preferential (R-P) genes (see material and methods for detail). We performed real-time PCR for 6 of these genes and the results were consistent. In our result, 24 genes involved in male reproductive development were detected (Table 3-1). Consistent with previous studies, SPL were found in the A-P group and EMS1 in R-P group, while AMS as an A-S gene (Yang et al., 1999a; Zhao et al., 2002).

Recently, other studies were conducted to identify male reproductive development-related genes. Wellmer et al. identified genes expressed in stamen indirectly by comparing the inflorescence transcriptome of floral homeotic mutants lacking stamens with wild-type (Wellmer et al., 2004b). In another study, Honys et al. analyzed microspores/pollen from different stages and defined the male gametophytic transcriptome (Honys and Twell, 2003). A comparison of our A-P genes with these two previous gene lists (defined as stamen and pollen) revealed that only a small number of genes overlapped between the three male reproductive datasets (Fig. 3-4A). The differences in identified genes can be explained by the difference of samples used in different studies: our samples only included wild-type anthers at early stages (stage 4-7), whereas the pollen transcriptome data were from microspores and pollen at different stages; and stamen-specific genes was indirectly obtained by subtraction of mutant transcriptome from wild-type and
genes in this list might function earlier during organ specification. The dramatic differences between different samples suggest strongly that gene activities alter dramatically between different developmental stages of male reproductive organs (Honys and Twell, 2003).

We analyzed the GO categorization for possible enrichment of specific categories among the groups of differentially expressed genes and found that, among the 266 A-P genes, the over-represented GO categories were hydrolases, proteins with other binding activities, and other enzymes. No enrichment of other enzyme activity was detected in pollen-specific or stamen-specific datasets found previously (Honys and Twell, 2003), suggesting a specific expression profile of early anther development.

Among genes with differential expression in *ams*, the percentage of A-P genes (5%) is significantly higher than its percentage in the whole genome (1%) (Fig. 3-4B). The stamen-specific genes were also enriched among those differentially expressed in *ams* (9%) compared with whole genome data (5%) (Fig. 3-4C&D). The results were consistent with our hypothesis that AMS regulates genes with important functions in male-reproductive organ where they have higher expression levels (Chang et al., 2011; Ge et al., 2010).

### 3.4 Genome-wide analysis of gene expression during early anther development by comparing anther transcriptomes of male sterile mutants, *spl, ems1, and ams*

Previous studies revealed essential roles of *SPL* and *EMS1* in early anther development and ATH1 microarray data from anthers of these mutants at stage 4-6 were collected and analyzed (Wijeratne et al., 2007). To obtain a better overview of early anther development, we analyzed the anther transcriptome data from this study with those of *spl* and *ems1* (detailed methods applied to all microarray data is described in experimental procedures). 1813 and 802 genes were identified as differentially expressed in *spl* and *ems1*, respectively, contributing to a
total of 3058 genes that were differentially expressed by 2-fold or more between the wild-type anther and one or more of the *spl*, *ems1* and *ams* mutant anthers. Using the log₂ values of the ratio of expression of the differentially expressed genes, hierarchical clustering was carried out to obtain heat-maps (Fig. 3-5A). The patterns of *spl* and *ems1* were similar whereas *ams* had a different pattern, consistent with the fact that the tapetum layer is absent in both *spl* and *ems1* but is formed in the *ams* anther.

In addition, we compared the direction of differential gene expression by pair-wise comparison between different mutants, as shown in Venn diagrams (Fig. 3-6A-D) and found that many more genes showed changes in the same direction in all three mutants than genes with changes in the opposite direction, suggesting that the three transcription factors had similar effects on some of the target genes. We also found that the non-overlapping (differentially expressed in one mutant, but not in either of the other two) percentage of differentially expressed genes in *ams* (76%) is larger than those in *spl* and *ems1* (59% & 23%, respectively, Fig. 3-6D), providing strong evidence at the transcriptome level that the AMS function was distinct from those of SPL and EMS1 and likely regulates late gene expression in anther development, consistent with other studies (Sorensen et al., 2003).

Because the three mutants showed related but distinct phenotypes, we speculate that the functions of genes differentially expressed in these mutants might differ from each other. Thus we applied GO categorization of molecular function to genes up- or down-regulated in each mutant. First, genes annotated to have “other binding activities” and “other enzyme activities” were significantly enriched in categories with reduced expression in each mutant (P-value < 0.05), consistent with previous knowledge of dynamic metabolism in tapetum cells. In addition, genes encoding transcription factors and DNA binding proteins are enriched in categories with both up-and down-regulated genes in the *spl* mutant, suggesting that SPL control anther development at least in part by regulating genes encoding transcription factors. Furthermore, the
ams mutant showed reduced expression of many genes encoding structural proteins, which mainly contribute to cell structural integrity, suggesting that AMS might activate these genes to promote maturation of tapetum cells.

3.5 SPL and EMS1 might control tapetum development by activating AMS-dependent gene expression

To gain a better understanding of genes that may function together in anther development, we divided the 3058 genes into different clusters based on their expression patterns. Totally 136 genes had repressed expression in all three mutants. Since tapetum cells are either absent or dysfunctional in the mutants, we expected that the expression of tapetum-related genes would reduce significantly. Previous studies indicated that tapetal cells were primarily involved in nutrition and material provision for pollen maturation (Scott et al., 2004). Consistent with this notion, genes encoding enzymes in this group (21.8%) are obviously over-represented comparing with all genes on ATH1 chip (10.1%). Besides, genes belonging to the other binding category were also enriched in this group (16.4% V.s 9.9%).

Among these genes, many of them are involved in biosynthesis of pollen wall-related compounds, such as lipids, lignin and flavonoids. A recent study showed that the loss of acyl-CoA synthetase, GhACS1, which might be involved in biosynthesis and transfer of lipids, can lead to male sterility in cotton (Wang and Li, 2009). The expression levels of 8 Arabidopsis genes involved in the lipid metabolism pathway were significantly reduced in the three mutants, suggesting their potential roles in metabolism in tapetal cells (Table 3-2). Besides, we found that the expression levels of 30 genes involved in endomembrane system decreased in all mutants (Table 3). Recent studies in plants suggested that many endomembrane proteins might be involved in trafficking thus influencing signal transduction and development (Carter et al., 2004;
Jurgens, 2004; Surpin and Raikhel, 2004). Based on the observation of tapetum defects in all three mutants (Yang et al., 1999a; Zhang et al., 2006b; Zhao et al., 2002), we speculate that genes sharing similar expression patterns might be important for maintaining the tapetum identity.

In addition, five genes for potential transcription factors were also found in this category (Table 4). Among them, At5g58610 and AGL25/At5g10140 are A-P genes. At5g58610 has a putative function in pathogen defense reaction, uncovering a possible factor in both anther development and external biotic stress response pathways (Ascencio-Ibanez et al., 2008a; Libault et al., 2007). AGL25, also known as FLC, is a repressor of flowering and its expression is epigenetically regulated (Pien et al., 2008). However, its possible function in anther development is not known. Three others were AGL40/At4g36590, MYB80/At5g56110 and HAT9/At2g22800. AGL40 was found in the proliferative endosperm transcriptome and MYB80/At5g56110 in tapetum development (Zhang et al., 2007). These results suggested that normal tapetum functions might require multiple transcription factors preferentially expressed in the anther downstream of AMS.

3.6 SPL and EMS1 can regulate early anther development by AMS-independent pathways

Moreover, 354 genes showed reduced expression in spl and ems1 but not in ams, including the enrichment of the categories of hydrolase activity (15.5% vs. 8.4%), other binding activity (19.5% vs. 9.9%), and other enzyme activity (18.0% vs. 10.1%). Among the genes in this cluster, four genes: MS2, ACOS5, CYP703A2 and A7, were involved in sporopollenin monomer biosynthesis, the lack of which leads to male sterility (Table 3-1) (de Azevedo Souza et al., 2009; Rubinelli et al., 1998). Since these genes were not affected in the ams mutant, some lipid
metabolic genes might be activated independent of AMS and they might exert functions earlier than AMS or in parallel to AMS (Aarts et al., 1997; de Azevedo Souza et al., 2009).

Besides, several genes encoding putative transcription factors were found within this subset (Table 4). A-P genes with known functions, such as TDF1/At3g28470 (or MYB35) and bHLH89/ At1g06170, were also identified in this category (Wijeratne et al., 2007b; Xu et al., 2010; Zhu et al., 2008a). TDF1 is essential to the tapetum function controlling callose dissolution and acts downstream of SPL and upstream of AMS and MYB103 (Table 3-4) (Zhu et al., 2008a). Our data also support the regulatory hierarchy of SPL-TDF1-AMS.

The expression of AMS is significantly reduced in spl, therefore we assumed that genes down-regulated in ams should have similar reduction in spl. Interestingly, we found that 56 genes showed opposite expression changes in spl and in ams compared with wild type anther, and even larger proportion (1065 genes) only differentially expressed in ams (Fig. 3-6A). Another gene with reduced expression in spl and ems1 mutants is DYT1, which encodes a bHLH protein similar to AMS (Zhang et al., 2006b). It is possible that SPL might also regulate anther development through pathways independent of AMS, such as those requiring DYT1 function (Zhang et al., 2006). We speculate that SPL might activate other transcription factors that affect AMS-regulated genes in contrast to the function of AMS (represented by factors X and Y in Fig. 7). The effects of AMS reduction in spl might be outweighed by the loss of X or Y; such regulatory interactions would explain the opposite expression changes in spl and ams. The identification and understanding of the proposed factors will require further investigations.
3.7 AMS-dependent and independent anther expression of genes encoding transcription factors

3.7.1 Expression of bHLH genes during anther development

Since many transcription factors have been found to play key roles in regulating anther development, we analyzed our anther transcriptome profiles by focusing on transcription factor gene families (Feng and Dickinson, 2010; Ma, 2005; Wilson and Zhang, 2009). To identify additional candidate genes for anther development, we analyzed all 147 known bHLH genes in Arabidopsis (Li et al., 2006). For several clades according to the most recent phylogeny trees of bHLH family (Bailey et al., 2003; Li et al., 2006; Toledo-Ortiz et al., 2003), including the clade that includes AMS, all or most members of the same clade were expressed similarly in the anther, suggesting conserved functional roles in anther development. For example, bHLH91 and bHLH89 shared similar reductions in all three mutants, suggesting possible redundant functions in the anther (Fig. 3-5B).

In other cases, the closely related homologs did not share similar expression patterns in mutant vs. wild type anthers. For example, bHLH93 is a close homolog of AMS; but unlike AMS, it was preferentially expressed in the inflorescence compared with the anther. Also unlike AMS, it was elevated in expression in spl. It is possible that some compensatory mechanisms might act to increase transcription of bHLH93 when AMS is mutated (Fig. 3-5B).

In addition, some bHLH genes with known functions in other organs showed increased expression in the spl mutant, suggesting that SPL acts to maintain the identity of male reproductive organ by reducing the expression of genes needed for other organs. For example, ZCW32 (bHLH31) controls petal formation and was activated in the spl anther (Brioudes et al., 2009; Szecsi et al., 2006), suggesting that SPL can promote the normal anther development at an early stage by repressing some genes normally expressed in nearby whorls.
3.7.2 Possible role of MADS-box genes in anther development

Genes of the MADS-box family have been extensively studied in *Arabidopsis*, because they were first identified as flower homeotic genes that determine floral organ and meristem identities (Ma, 2009; Parenicova et al., 2003). Till now, more than one hundred MADS-box genes have been identified, 79 of which were found to be present in our anther microarray data but most were non-anther-specific (Parenicova et al., 2003). Except for *APETALA2 (AP2)*, majority of genes involved in the ABCDE model belong to the MADS family (Ma, 2009). They are mostly inflorescence-preferential rather than anther-specific genes from the comparison of microarray data as described above. *APETALAI (AP1)* is an A function gene controlling the first and second whorls and no expression shift was observed (Irish and Sussex, 1990). *APETALA3 (AP3)* and *PISTILLATA (PI)* are both B function genes, essential for the formation of petals and stamens (Goto and Meyerowitz, 1994; Irish and Sussex, 1990; Jack et al., 1992). Interestingly, their expression patterns were different. *PI* is an anther-preferential gene, but its expression level did not change in any mutant while *AP3* was obviously up-regulated in *spl*, suggesting that *AP3* is regulated more tightly than *PI* during anther development. *AG*, the C class gene controlling both stamen and carpel identities, shared similar expression patterns in the anther with *AP3* (Yanofsky et al., 1990b), supporting a role of *AG* in anther development after the specification of stamen identity (Fig. 5C).

Moreover, D class genes, including *STK/AT4g09960*, *SHP1/At3g58780* and *SHP2/At2g42830*, are important for ovule development (Favaro et al., 2003; Pinyopich et al., 2003). Although the expression of D class genes was relatively low, we observed increased expression of *SHP1* in the *ams* mutant, suggesting a possible negative regulatory role of AMS in ovule development. On the other hand, E class genes, *SEP1, SEP2, SEP3* and *SEP4*, which are
homologs that have redundant functions, had different expression pattern in the anthers. \textit{SEP1} and \textit{SEP2} were activated in \textit{spl} and \textit{ams}, whereas \textit{SEP3} and \textit{SEP4} did not change much (Fig. 5C).

Beside the ABCDE genes, some other MADS genes were also expressed in the anther. The expression levels of known flowering-time related genes (\textit{FLM}, \textit{AGL15}, \textit{AGL18} and \textit{AGL20}) (Adamczyk et al., 2007; Liu et al., 2008; Werner et al., 2005) were reduced in \textit{spl} and \textit{ems1} slightly. \textit{FUL} involved in fruit development (Gu et al., 1998) was up-regulated in the \textit{spl} and \textit{ems1} mutants, suggesting negative roles of SPL and EMS1 in whorl 4. \textit{AGL80}, important for central cell and endosperm formation in female gametophytes (Portereiko et al., 2006), was also reduced in all three mutants, suggesting a possible role in male gametophyte.

### 3.7.3 Differential expression of MYB genes in three mutants

In addition to the bHLH and MADS-box families, other gene families are also involved in anther development. As the largest \textit{Arabidopsis} transcription factor family, \textit{MYB} genes play important roles in controlling many cellular processes, such as secondary metabolism, morphogenesis, and signal transduction (Riechmann and Ratcliffe, 2000). Previous studies revealed a number of roles of MYB genes in early anther development (Fig. 3-5D). For example, \textit{GAMYB} in rice functions in anther development via GA signaling pathway (Aya et al., 2009). In \textit{Arabidopsis}, the GAMYB homologs \textit{MYB33} and \textit{MYB65} also share a redundant function regulating tapetum differentiation (Millar and Gubler, 2005a; Zhang et al., 2006; Zhang et al., 2007b; Zhong et al., 2007). Our microarray results indicated that expression of \textit{MYB33} and \textit{MYB65} was reduced only in \textit{spl}, not in the other two mutants, implying that the functions of \textit{MYB33} and \textit{MYB65} are independent of EMS1 or AMS.

In addition, \textit{MYB35/TDF1} and \textit{MYB80/MYB103} controlling callose dissolution and exine formation (Zhang et al., 2007) were reduced in \textit{spl} and \textit{ems1}, and \textit{MYB80} was also down-
regulated in *ams*, suggesting that it acts downstream of *AMS*. Moreover, the *MYB99* and *MYB101* genes that regulate phenylpropanoid metabolism (Alves-Ferreira et al., 2007) showed a similar expression pattern to that of *MYB35/TDF1*. *MYB26/MS35* and *MYB105* are closely related homologs; both were down-regulated in *spl* but up-regulated in *ams*. Previous study suggested that *MYB26* is required for endothecium thickening and anther dehiscence (Yang et al., 2007). RNA in situ hybridization revealed that *MYB105* as well as *MYB101* are expressed in late tapetum (Steiner-Lange et al., 2003; Yang et al., 2007), consistent with our findings of the changes of their expression in the mutant anthers.

### 3.7.4 Expression of WRKY, bZIP, AP2/ERF and NAC genes

The *WRKY* family contains at least 72 members in *Arabidopsis* (Wu et al., 2005) and has diverse functions, such as abiotic and biotic stress response, hormone signaling pathway, immune response and development in plants (Eulgem et al., 2000). However, it is not known whether *WRKY* genes are important for flower development. Here we compared the expression of all *WRKY* genes on the ATH1 chip and found that 29 of them were expressed in the anther, with the highly similar *WRKY2/At5g56270* and *WRKY32/At4g30935* (Eulgem et al., 2000) being anther-preferential. Moreover, *WRKY2* was down-regulated in *spl*, suggesting that it might function downstream of *SPL* in anther development.

We also analyzed bZIP, ERF and NAC families of transcription factors. Like the *WRKY* family, most genes in these families do not have known functions in reproductive development. However, we found several of them were differentially expressed in the anthers of male sterile mutants, suggesting they are components of a complex transcriptional network regulating anther development.
3.8 Transcriptional regulatory network for anther development

Genetic studies and our transcriptomic analyses reported here support an emerging transcriptional network (Fig. 3-8). Previous molecular genetic studies showed that SPL up-regulates the expression of *EMS1* and *DYT1*, which are upstream of *AMS* (Sorensen et al., 2003; Zhang et al., 2006), as well as other genes encoding transcription factors shown to be important in anther development (Chang et al., 2011). *SPL* also negatively regulates the expression of B and C function genes in the anther, as well as some genes that are normally expressed in petals and carpals, probably to prevent anther from developing traits of other floral organs. In addition, the key position of *SPL* in anther regulatory hierarchy as indicated by genetic studies is supported by its effects on the anther transcriptome (Fig. 3-8A).

*EMS1* also positively regulates the expression of *DYT1* (Wijeratne et al., 2007b). *EMS1* was shown to interact with its putative ligand TPD1 (Jia et al., 2008), thereby regulating genes essential for the differentiation of tapetum cells. In addition, some genes important for meiosis are also affected in the *ems1* mutant. For example, the MMD and ROXY2 genes that are important in anther lobe formation and meiosis, respectively, were significantly reduced in *ems1* (Fig. 3-8B).

*AMS* was down-regulated in *spl* and *ems1* according to the microarray data. Because DYT1 and AMS are related bHLH proteins, which are known to form homodimers or heterodimers with other bHLH proteins, we propose that they probably regulate the expression of different genes by forming different complex with other proteins. DYT1 is also a putative candidate that exerts opposite function as X and/or Y downstream of SPL in anther development by interact with different proteins (Fig. 3-8C). This proposed transcriptional regulatory network of anther development is based on information from genetics, transcriptomics, and phylogenetics studies (Fig. 3-8A-C). The hypothesized interactions, including the roles of some functionally redundant genes, could be tested by further experiments.
Figure 3-1. The expression of genes differentially expressed in *ams* anthers.

(A) A comparison between transcriptome data from *ams* and wild type anthers. All expression data were converted to logarithm base 2 ratio. (B) A histogram of genes with elevated or reduced expression levels between *ams* and wild type anthers. The y-axis the frequency of expression and the x-axis is the log 2 ratio of expression signals. (C) A comparison between *ams* transcriptome data from anthers and inflorescences. (D) A comparison between differentially expressed genes in *ams* anthers and those in wild type inflorescences compared with wild type anthers.
Figure 3-2. A pie graph of GO categorization of genes differentially expressed in the *ams* mutant.

(A-B) GO categorization of genes up- and down-regulated in *ams*, with enriched categories circled compared with all genes on ATH1. (C) GO categorization of all genes on the ATH1 chip.
Figure 3-3. Marginal plot of fold changes of expression and number of E-box.

(A) Genes differentially expressed in the *ams* anther compared with wild-type anther were selected (more than two fold changes with P<0.05). The logarithm 2 values of fold-changes and the number of AMS binding sites (E-box) within 1kb putative promoter sequence upstream of the start codon were plotted. (B) Randomly selected genes without differential expression in *ams* were plotted as control.
Figure 3-4. Venn diagrams of microarray results and previous related study.

(A) A comparison of anther preferential genes identified in our study with previously known pollen genes and stamen genes. (B-D) Comparisons between genes differentially expressed in the *ams* anther and those preferentially expressed in certain organ: anther preferential, stamen and pollen respectively.
Figure 3-5. Expression distribution of all genes differentially expressed and specific gene families.

(A) Hierarchical clustering of genes differentially expressed in at least one mutant. (B-D) Heat-map of bHLH, MADS, MYB genes with putative or known function in male-reproductive development. The number indicate logarithm ratio of the fold change in mutant compared with wild-type anther. “w” represents wild-type anther and “s”, “e” and “a” represents spl, ems1 and ams. Red color represents genes which have higher expression level in mutants and green indicates reduced expression.
(A-C) Genes differentially expressed in three mutants respectively compared with wild-type anther have been identified and then compared with each other pairwisely. (D) Comparison of genes differentially expressed in three mutants compared with wild-type.
Figure 3-7. AMS-dependent or -independent regulatory model during anther development

Genes with differential expression in the *ams* mutant is divided into eight groups A-H (plus represents for higher expression in mutant compared with wild-type and minus for lower). Negative regulation is shown by a T-bar and positive by an arrow, and stronger impact is shown by a thicker T-bar or arrow.
Gene regulation is represented by T-bars (negatively) and arrows (positively). The direct regulation confirmed by experiment is represented in bold line. Genes encoding proteins with interaction is represented by double arrows. Gene expression patterns in different tissues are shown by colors (blue for anther specific; red for anther-preferential; green for reproductive-preferential and yellow for genes not included in ATH1 chip). Gene function in tapetum formation is marked by an apostrophe; in pollen wall formation by an asterisk; in callose dissolution by double asterisks; in stamen and petal formation by the letter b; in stamen and carpel formation by the letter c.
Table 3-1. Expression of genes known as anther development related genes.

<table>
<thead>
<tr>
<th>AGI</th>
<th>Name</th>
<th>wt</th>
<th>s/w</th>
<th>e/w</th>
<th>a/w</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT2G16910</td>
<td>AMS</td>
<td>7.7</td>
<td>-3.4</td>
<td>-3.5</td>
<td>1.7</td>
<td>tapetum dev.</td>
<td>Sorensen et al., 2003</td>
</tr>
<tr>
<td>AT1G66170</td>
<td>MMD1</td>
<td>5.5</td>
<td>-1.8</td>
<td>-1.8</td>
<td>-0.1</td>
<td>male meiosis</td>
<td>Alves-Ferreira et al., 2007</td>
</tr>
<tr>
<td>AT1G01280</td>
<td>CYP703A2</td>
<td>9.8</td>
<td>-6.2</td>
<td>-6.2</td>
<td>-0.6</td>
<td>pollen dev. and sporopollenin biosynthesis</td>
<td>Souza et al., 2009</td>
</tr>
<tr>
<td>AT1G62940</td>
<td>ACO55</td>
<td>11.6</td>
<td>-6.4</td>
<td>-5.7</td>
<td>-1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G11980</td>
<td>MS2</td>
<td>10.6</td>
<td>-6.1</td>
<td>-5.9</td>
<td>-0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G28395</td>
<td>A7</td>
<td>9.6</td>
<td>-4.1</td>
<td>-3.9</td>
<td>0.5</td>
<td></td>
<td>Rubinelli et al., 1998</td>
</tr>
<tr>
<td>AT2G17950</td>
<td>WUSCHEL 1</td>
<td>7.7</td>
<td>-2.5</td>
<td>0.4</td>
<td>0.5</td>
<td>floral dev.</td>
<td>Ming et al., 2009</td>
</tr>
<tr>
<td>AT4G27330</td>
<td>NZZ/SPL</td>
<td>9.3</td>
<td>-3.9</td>
<td>-0.5</td>
<td>-0.5</td>
<td></td>
<td>Ito et al., 2004</td>
</tr>
<tr>
<td>AT5G14070</td>
<td>ROXY2</td>
<td>9.1</td>
<td>-3.2</td>
<td>-1.7</td>
<td>-0.2</td>
<td>early anther formation</td>
<td>Xing et al., 2008</td>
</tr>
<tr>
<td>AT3G11440</td>
<td>MYB65</td>
<td>8.9</td>
<td>-2.2</td>
<td>-0.5</td>
<td>-0.4</td>
<td></td>
<td>Millar et al., 2005</td>
</tr>
<tr>
<td>AT5G06100</td>
<td>MYB33</td>
<td>7.6</td>
<td>-1.1</td>
<td>0.1</td>
<td>-0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G42960</td>
<td>ATA1</td>
<td>11.9</td>
<td>-7.4</td>
<td>-6.7</td>
<td>-1.7</td>
<td>tapetum function</td>
<td>Lebel-Hardenack et al., 1997</td>
</tr>
<tr>
<td>AT3G51590</td>
<td>LTP12</td>
<td>10.4</td>
<td>-6.1</td>
<td>-5.2</td>
<td>0.9</td>
<td></td>
<td>Ariizumi et al., 2002</td>
</tr>
<tr>
<td>AT3G28470</td>
<td>MYB35</td>
<td>8.6</td>
<td>-4.9</td>
<td>-4.5</td>
<td>0.0</td>
<td></td>
<td>Zhu et al., 2008</td>
</tr>
<tr>
<td>AT1G69500</td>
<td>CYP704B1</td>
<td>11.5</td>
<td>-7.4</td>
<td>-7.0</td>
<td>-2.0</td>
<td>pollen dev. and sporopollenin biosynthesis</td>
<td>Souza et al., 2009</td>
</tr>
<tr>
<td>AT4G34850</td>
<td>LAP 5</td>
<td>11.4</td>
<td>-6.3</td>
<td>-5.8</td>
<td>-1.6</td>
<td></td>
<td>Dobritsa et al., 2010</td>
</tr>
<tr>
<td>AT4G35420</td>
<td>DRL1</td>
<td>11.9</td>
<td>-5.1</td>
<td>-4.6</td>
<td>-1.9</td>
<td></td>
<td>Tang et al., 2009</td>
</tr>
<tr>
<td>AT5G62080</td>
<td>MTG10</td>
<td>13.0</td>
<td>-7.4</td>
<td>-5.2</td>
<td>-4.2</td>
<td></td>
<td>Xing et al., 2007</td>
</tr>
<tr>
<td>AT3G22880</td>
<td>DMC1</td>
<td>10.7</td>
<td>-2.2</td>
<td>-0.1</td>
<td>0.0</td>
<td>male meiosis</td>
<td>Doutriaux et al., 1998</td>
</tr>
<tr>
<td>AT3G15400</td>
<td>ATA20</td>
<td>12.1</td>
<td>-6.6</td>
<td>-5.6</td>
<td>0.1</td>
<td>pollen wall</td>
<td>Rubinelli et al., 1998</td>
</tr>
</tbody>
</table>

R-P
<table>
<thead>
<tr>
<th>AGI</th>
<th>wt</th>
<th>s/w</th>
<th>e/w</th>
<th>a/w</th>
<th>Function</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>lipid related</td>
<td></td>
</tr>
<tr>
<td>At5g61320</td>
<td>8.7</td>
<td>-3.9</td>
<td>-3.6</td>
<td>-2.2</td>
<td>cytochrome P450 - like protein</td>
<td>A-S</td>
</tr>
<tr>
<td>At5g08250</td>
<td>8.9</td>
<td>-4.5</td>
<td>-4.1</td>
<td>-1.4</td>
<td>lipase-like protein</td>
<td></td>
</tr>
<tr>
<td>At1g06250</td>
<td>7.6</td>
<td>-3.2</td>
<td>-3.0</td>
<td>-1.8</td>
<td>putative pectate lyase</td>
<td>A-P</td>
</tr>
<tr>
<td>At5g62080</td>
<td>13</td>
<td>-7.4</td>
<td>-5.2</td>
<td>-4.2</td>
<td>lipid-transfer protein</td>
<td></td>
</tr>
<tr>
<td>At3g07450</td>
<td>12.8</td>
<td>-7.5</td>
<td>-5.6</td>
<td>-5.2</td>
<td>pectin methylesterase inhibitor</td>
<td></td>
</tr>
<tr>
<td>At3g52130</td>
<td>12.9</td>
<td>-7.4</td>
<td>-6.0</td>
<td>-3.3</td>
<td>putative cinnamoyl-CoA reductase</td>
<td></td>
</tr>
<tr>
<td>At5g07230</td>
<td>10.2</td>
<td>-8.1</td>
<td>-6.5</td>
<td>-2.7</td>
<td>pectinesterase like protein</td>
<td></td>
</tr>
<tr>
<td>At5g52160</td>
<td>10.7</td>
<td>-6.3</td>
<td>-6.2</td>
<td>-1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>pectin</td>
<td></td>
</tr>
<tr>
<td>At3g24230</td>
<td>7.6</td>
<td>-3.2</td>
<td>-3.2</td>
<td>-3.2</td>
<td>pectate lyase</td>
<td>A-P</td>
</tr>
<tr>
<td>At4g22080</td>
<td>8.5</td>
<td>-4.9</td>
<td>-4.6</td>
<td>-1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At1g75790</td>
<td>10.2</td>
<td>-5.8</td>
<td>-5.5</td>
<td>-1.9</td>
<td>pectinesterase like protein</td>
<td>R-P</td>
</tr>
<tr>
<td>At3g01270</td>
<td>6.2</td>
<td>-1.2</td>
<td>-1.2</td>
<td>-1.1</td>
<td>putative pectate lyase</td>
<td></td>
</tr>
<tr>
<td>At5g50030</td>
<td>6.5</td>
<td>-1.3</td>
<td>-1.1</td>
<td>-1.4</td>
<td>putative cinnamoyl-CoA reductase</td>
<td>A-P</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>pectin methylesterase inhibitor</td>
<td></td>
</tr>
<tr>
<td>At1g76470</td>
<td>10.3</td>
<td>-4.8</td>
<td>-5</td>
<td>-2.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
All the expression values are log2 ratio.

<table>
<thead>
<tr>
<th>AGI</th>
<th>wt</th>
<th>s/w</th>
<th>e/w</th>
<th>a/w</th>
<th>Function</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>At3g21230</td>
<td>7.7</td>
<td>-3.1</td>
<td>-3.1</td>
<td>-1.8</td>
<td>4-coumarate--CoA ligase 2</td>
<td></td>
</tr>
<tr>
<td>At3g13220</td>
<td>10</td>
<td>-5.9</td>
<td>-5.7</td>
<td>-2.2</td>
<td>WBC27 white-brown complex</td>
<td>A-S</td>
</tr>
<tr>
<td>At4g14080</td>
<td>12.5</td>
<td>-8.1</td>
<td>-7.1</td>
<td>-3.8</td>
<td>maternal effect embryo arrest 48</td>
<td>A-P</td>
</tr>
<tr>
<td>At1g02050</td>
<td>11.6</td>
<td>-4.6</td>
<td>-4.7</td>
<td>-1.9</td>
<td>LESS ADHESIVE POLLEN 6 (LAP6)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3-3. Genes related to endomembrane system affected by SPL, EMS1 and AMS
All the expression values are log2 ratio.

Table 3-4. Transcription factors in SEA-L and SE-L cluster with known or putative function in anther development

<table>
<thead>
<tr>
<th>Cluster</th>
<th>AGI</th>
<th>wt</th>
<th>s/w</th>
<th>e/w</th>
<th>a/w</th>
<th>Function</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEA-L</td>
<td>At5g58610</td>
<td>7.7</td>
<td>-1.6</td>
<td>-1</td>
<td>-1.9</td>
<td>PHD finger</td>
<td>A-P</td>
</tr>
<tr>
<td></td>
<td>At5g10140</td>
<td>8.1</td>
<td>-3.2</td>
<td>-3.1</td>
<td>-1.8</td>
<td>AGL25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At4g36590</td>
<td>5.8</td>
<td>-1.7</td>
<td>-1.4</td>
<td>-1.7</td>
<td>AGL40</td>
<td>R-P</td>
</tr>
<tr>
<td></td>
<td>At2g22800</td>
<td>8.7</td>
<td>-2.9</td>
<td>-2.0</td>
<td>-1.3</td>
<td>HAT9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At5g56110</td>
<td>7.5</td>
<td>-1.6</td>
<td>-1.5</td>
<td>-1.8</td>
<td>MYB 80/MYB103</td>
<td></td>
</tr>
<tr>
<td>SE-L</td>
<td>At1g06170</td>
<td>9.6</td>
<td>-5.9</td>
<td>-4.6</td>
<td>-0.8</td>
<td>bHLH89</td>
<td>A-P</td>
</tr>
<tr>
<td></td>
<td>At3g28470</td>
<td>8.6</td>
<td>-4.9</td>
<td>-4.5</td>
<td>0.0</td>
<td>TDF1/MYB35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At2g31210</td>
<td>7.1</td>
<td>-3.4</td>
<td>-2.9</td>
<td>-1.1</td>
<td>bHLH91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At3g57370</td>
<td>7.3</td>
<td>-2.8</td>
<td>-2.8</td>
<td>0.8</td>
<td>initiation factor IIIB</td>
<td>R-P</td>
</tr>
<tr>
<td></td>
<td>At1g77850</td>
<td>8.8</td>
<td>-2.4</td>
<td>-1.7</td>
<td>-0.1</td>
<td>auxin response factor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At2g28830</td>
<td>6.5</td>
<td>-1.1</td>
<td>-1.1</td>
<td>0.3</td>
<td>transcription activator</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At5g62320</td>
<td>8.6</td>
<td>-3.3</td>
<td>-3.4</td>
<td>-0.3</td>
<td>MYB99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At4g09460</td>
<td>8.7</td>
<td>-3.4</td>
<td>-2.3</td>
<td>0.8</td>
<td>MYB6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At4g34680</td>
<td>9.3</td>
<td>-2.6</td>
<td>-1.2</td>
<td>0.3</td>
<td>GATA 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At2g41630</td>
<td>10.6</td>
<td>-1.1</td>
<td>-1.4</td>
<td>-0.7</td>
<td>TFIIIB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>At3g10580</td>
<td>7.3</td>
<td>-2.5</td>
<td>-2.3</td>
<td>0.9</td>
<td>MYB</td>
<td></td>
</tr>
</tbody>
</table>
In this study, we identified genes whose expression were changed in spl, ems1 and ams at anther stage 4-7 and further categorized these genes according to their expression patterns. These genes might directly regulate some fundamental biological processes during anther development. In addition, both anther-specific and non-anther-specific genes are identified in anther development. Transcriptome analyses also showed AMS-dependent and -independent pathways. Careful analyses of transcriptome combined with genetic and phylogenetic information revealed an elaborate regulatory network during early anther development and expanded our understanding of the hierarchy of anther-development-related genes, especially transcription factors.
5. Materials and Methods

5.1 Plant materials

All the plants in this study were grown in soil under long day condition (16 hr light/ 8 hr dark) at constant 22°C. The wild-type in this paper refers to ecotype Landsberg erecta (Ler). The mutants of spl, emsl are of Ler background as described (Wijeratne et al., 2007b; Zhang et al., 2006b), while the ams mutant is of Columbia background. We select 21-28 day old plant to collect anther at 4-7 stage as described previously (Wijeratne et al., 2007b).

5.2 Microarray experiment

Following the Affymetrix GeneChip Expression Analysis Overview described on the website (Xu et al., 2010), cRNA was synthesized for hybridization as described (Wijeratne et al., 2007b). Hybridization, washing, staining, scanning and data collection were performed at the Genomics Core Facility, Pennsylvania State University, University Park.

5.3 Microarray analysis to identify differentially expressed genes in anther of mutants

Normalization was applied using Bioconductor package in R by RMA (Zhang et al., 2005a), and all expression values were converted to logarithms base 2. LIMMA library was then used to compare signals from mutant and wild-type anther. Only genes with more than two-fold changes were selected. To obtain more reliable result, we screened out genes with q-value (FDR)
larger than 0.05, since q-value is more stringent than p-value of T-test based on previous study (Storey and Tibshirani, 2003).

Similar data processing was performed with the microarray results from different organs. The microarray data from all organs in wild-type *Arabidopsis* were normalized together and converted to logarithms base 2 values. We defined genes as anther-specific if they met these criteria: 1) the expression in anther is significantly higher than in any other tissue with FDR<0.05; 2) gene is present in anther but absent in any other tissues. We used two alternative methods to define whether a gene is present in a tissue. One of the methods was using the Affymetrix’ MAS5 algorithm. This method uses a comparison of hybridization intensity with wild-type oligo set vs mismatched oligo set; sometimes similar levels of hybridization to both sets can actually be real expression, yet such results would lead to “absent” calls. Therefore, we also used a second method to define “presence”, by using a threshold of 50 for expression value, previously determined on basis of analysis of variation among samples of the same tissue (Wijeratne et al., 2007b; Zhang et al., 2005a).

For the anther-preferential genes, we used the criteria that the expression in anther is 1) present using both MAS and/or 50 cutoff; 2) significantly higher than in any other tissue with FDR<0.05; 3) at least 2 fold more compared with any other tissues. The reproductive-preferential genes required the expression present and significantly higher in anther than only the vegetative organs using FDR<0.05 and 2-fold changes.

Hierarchical clustering of co-expressed genes was performed by MeV 4.6 (Yang et al., 2007). We used Euclidean distance metric to conduct this analysis. For the identification of the functions of the differentially expressed genes, the annotations of genes on ATH1 microarray chip were downloaded from Affymetrix website and we used the GO categorization function on TAIR website (Ma, 2009). To verify whether one category is enriched compared with the whole
5.4 Cis-regulatory element analysis

Possible promoter sequences of all genes on the microarray chip (1kb upstream of the start codon) were obtained from TAIR website. The number of common bHLH binding site (E-box) was then counted. We then plotted the fold-changes of gene expression in *ams* against the numbers of their putative AMS binding sites using minitab (Sato et al., 1997). The identification of *cis*-regulatory binding site was conducted by perl (Jurgens et al., 2011). The binding motifs were obtained from Gene Regulation and PlantCARE (Parenicova et al., 2003b).

5.5 Real-time PCR experiments

To test the reliability of our microarray hybridizations, six genes and one reference (ACT2, At3g18780) were studied using Quantitative Real-Time PCR. RNA extraction and Real-Time experiments followed the protocols described previously (Zahn et al., 2010a). Triplicate reactions were performed for all tissues with “no reverse transcription” as a negative control. Relative transcript quantities were calculated using the ΔΔCt method (Livak and Schmittgen, 2001).
6. Acknowledgements

We greatly appreciate the help of Dr. Craig Praul in performing microarray hybridizations. We thank Dr. Naomi Altman for discussion of statistic analysis. We also thank Ms. Jiong Wang for plant care. We appreciate the suggestion from Drs. Xiaofan Zhou, Zhenhai Zhang, Zhao Su on microarray data analysis. This work was supported by a US Department of Energy grant to H.M. and funds from Department Biology and the Huck Institutes of the Life Sciences, the Pennsylvania State University, and Fudan University.
7. References


CHAPTER 4

IMPACTS OF DIFFERENT DROUGHT SEVERITIES ON
REPRODUCTIVE ORGANS IN *ARABIDOPSIS*

Zhao Su, a post-doctoral fellow in the Ma lab, initiated the study of severe drought, which provided a comparison for the analysis described in this chapter. I was involved in the microarray data analysis for the study with severe drought and am the second author of a manuscript to be submitted in the near future. Zhao and Liyana, another graduate student in the lab, contributed to morphological analyses presented in part of Figure 4-1 to Figure 4-4. I designed the study, grew the plants, collected the tissues, isolated the RNAs, and analyzed the data of the microarray experiments described in this chapter. The Penn State University Microarray Facility performed the microarray hybridizations.
1. Abstract

Abiotic stresses can impact plant growth and limit crop production. Drought is a major constraint which leads to extensive losses to agricultural yield worldwide. Previous studies discovered thousands of genes responsive to dehydration and grouped them to two main pathways, ABA-dependent and –independent pathways. However, most investigations on plant response to drought focused on vegetative development. This study describes the morphological changes of reproductive development under various drought conditions. On the basis of our observation, the transcriptomes of inflorescences under two water conditions: minimum for successful reproduction (50%, moderate) and for survival (35%, severe) were further studied at molecular level using microarrays. Our results showed more than four thousand genes with differential expression under severe drought and less than two thousand changed under moderate drought condition (with 2-fold change & q-value<0.05). We found a group of genes with increasing expression as the drought became more severe, suggesting their functions in adaptation to the dehydration. Interestingly, we also identified genes with alteration only under the moderate but not the severe drought condition, indicating the existence of distinct sets of genes responsive to different levels of water availability. Further cis-element analyses of the putative regulatory sequences provided more information about the underlying mechanisms for reproductive responses to drought, suggesting possible novel candidate genes that protect the developing flower under drought.
2. Introduction

The increasing world population (up to 7 billions by 2010) suggests a growing demand in crop production. Agricultural productivity is inevitably limited by the environmental stress, such as drought, salinity, heat and cold (Golldack et al., 2011b). Many of these abiotic factors might cause the loss of yields partially resulting from the dehydration of plant cells. Despite the abundance of water on earth, most of the water resources are not usable for irrigation due to the salinity. Thus, more and more investigations focused on the mechanisms by which plants adapt to dehydration. Dehydration resistance consists of two main categories: dehydration avoidance or by dehydration tolerance (Blum, 2005). The dehydration avoidance is defined as the plant capacity to maintain cellular hydration in spite of stress and plants could achieve it by maintaining soil moisture, limiting water use (WU), and osmotic adjustment (OA). Dehydration tolerance is defined as the relative capacity to maintain normal function even in a dehydrated state, which is also viewed as the secondary defense against desiccation. This mechanism is not commonly observed other than in seed embryo and the only main exception occurs during certain stages of grain filling under drought (Blum, 1998).

Drought, the most direct reason leading to plant dehydration, has been studied for years. It could impact plants at molecular, cellular, physiological and biochemical levels and cause multiple phenotypes in various tissues. It has been shown that drought can severely impact multiple developmental process, including seed germination (Finkelstein et al., 2002), seedling growth (van der Weele et al., 2000), root development (Xiong et al., 2006) and later leaf development (Bohnert et al., 1995; Luan, 2002; Yaish et al., 2011). However, only a few studies have studied reproductive development under drought conditions (Yaish et al., 2011). In many flowering plants, the emergence of flowers coincides with drought stress during summer. To
ensure the successful reproduction, flowering plant must possess mechanisms that protect flowers from severe dehydration.

Another challenge for scientists studying drought is controlling the water condition. It is known from both scientific studies and agricultural experience that different extents of drought stress could impact the development and yield of crop at different levels (Ali et al., 1999). A few studies tried to calculate the minimum ecological water requirement in certain regions and proposed to enhance the capacity to deal with drought using water management (Gordon et al., 2010; Jia et al., 2011). However field studies on drought might also have substantial deviation for several reasons: 1) the difficulty of controlling soil water content accurately; 2) delay of drought effects on plant due to the variation of evaporation rate and soil content; and 3) substantial deviation due to variation of nutrients in soil. Besides, it is hard to address the question of the minimum water requirement to guarantee plant survival or to maintain the yield by field studies.

Unlike large-scale field studies, studies in the lab could allow comparatively precise control of the water condition to explore the mechanisms that plants employ to survive. By reducing the water supply, scientists found many genes actively involved in a sophisticated drought response system. Both ABA-dependent and –independent pathways have been discovered as drought-responsive (Golldack et al., 2011a; Hauser et al., 2011b; Seki et al., 2007). Further study revealed key components in these pathways, including transcription factors belonging to bZIP, AP2/ERF, and MYB families (Hossain et al., 2010; Lippold et al., 2009; Yoshida et al., 2010). With the availability of transcriptomic profiling, more and more genes responsive to drought stress have been reported, especially in the model plants whose genomic information is available, such as Arabidopsis, rice and maize (Golldack et al., 2011b; Kilian et al., 2007; Sreenivasulu et al., 2010). However, most of these studies focused on vegetative tissues where more obvious phenotype could be observed, No previous studies have investigated the transcriptome of inflorescence under drought stress (Kilian et al., 2007; Rabbani et al., 2003).
A recent study in our lab has shown the impacts of severe drought on the development of reproductive organs, such as postponing flowering time, reducing inflorescences and seed numbers (Zhao Su, Xuan Ma, Huihong Guo, N. Liyana Sukiran, Hong Ma, manuscript in preparation). In this study, detailed morphological analyses showed that the development of both male and female reproductive organs was impacted, resulting in ovule abortion, failure of flowers to open, abnormal anther development and delayed elongation of the filaments and stigma papillae cells. Further examination of the inflorescence transcriptome under well-watered and severe drought conditions provided molecular evidence that floral growth can be severely affected by severe drought. To understand the impacts of different magnitude of drought stress on reproductive development, we treated the Arabidopsis plants with different drought severities right after the bolting stage (around the time of the first opening flower) and observed their morphological changes. In this study, we further compared the difference of inflorescences under various magnitudes of drought. We also collected inflorescences from the treated plants after a series of periods of drought stress and used the mRNA sample to exert microarray experiments in order to investigate the changes at molecular level. The analyses of the gene expression pattern and promoter cis-acting element searching combined with previous understanding of flower development enable us to understand more about how the reproductive tissues respond to drought stress. On the basis of recent microarray data from Arabidopsis inflorescences, we proposed that many genes related to flower development might be responsive to drought in order to assure the survival of whole plant under stress (Su, unpublished data).
3. Results and Discussion

3.1 Morphological changes in response to different drought severities in Arabidopsis

Drought stress could severely impact plant growth. To avoid variation caused by the timing of drought treatment, we sowed one seed in each pot and kept the plant well-watered (90% water content) until bolting stage (about 24 days after plant with average total height about 1cm and 8-9 rosette leaves). To investigate the impact of drought on inflorescences at different levels, we divided those plants to six groups. Except for the control group which maintained the water condition, the water supply of the other five groups were simultaneously stopped until their soil moisture reached 70%, 50%, 40%, 35%, and 30% individually. Plants can hardly survive lower soil moisture than 30% (Su, unpublished) (Figure 4-1). After 10 day drought, an obvious reduction of total plant height could be observed. The less water sources obtained, the shorter the main stem. It is also obvious that fewer branches, flower buds and siliques were produced as the drought condition became more rigorous (Figure 4-2).

To evaluate the impact of drought in detail, we counted the flower number on the main stem after the drought treatment. As seen in Figure 4-3, the plant will not compromise its reproductive development in terms of flower number, as long as the water content is above 50%. When the soil moisture is below 50%, the accumulative flower number stopped growing for a while and resumed slowly at the first few days under drought (Figure 4-3). A postponed resumption was also observed as the magnitude of drought increased. Though the impact of drought on inflorescences prolong the process of blooming, the plants under severe drought could still manage to produce flowers after revival, and reach almost the same flower number as the control group. Interestingly, the plants under most intensive drought condition (at 30% soil moisture) strived to produce even more buds than those under moderate drought (at 40% soil
moisture). To figure out whether plants could survive better under the more severe condition, we took a closer look at the seed production and found that the plant might be able to endure a certain extent of drought without obvious loss of yield on the main stem (Figure 4-4&5). Similar to the trend of flower number, there is hardly any evident difference between the three groups with soil moisture greater than 50%. However, an obvious reduction of yield could be observed when the water content reached 40%. More severe impacts on the number of siliques were seen under 30% and 35% of soil moisture condition. Different those under mild drought condition (above 50%), plants undergoing severe drought stopped producing siliques for a period. The length of the ceasing period (plants stopped producing siliques) prolonged as the severity of drought increased. As well, the seed number per silique also began to decrease as the water content went below 50%. The average seed number of each silique was reduced by 50% with 35% water supply and went down even further to only 20% of the control group with 30% water moisture. We also noticed that not all the plants with 30% soil moisture can survive till the end. Thus we used 35% as the most severe condition so that plants can survive for our experiments.

3. 2 Transcriptome analyses of inflorescences under moderate drought condition

3.2.1 Analyzing transcriptomic changes under moderate drought by microarray

On the basis of the morphological observation, we hypothesized that plant might adopt different mechanisms to deal with drought stress under mild and severe drought condition. In order to analyze the plant response to drought at the molecular level, we choose two vital conditions to conduct further analyses: the minimum soil moisture that plant can adapt without compromise yield (50%) and we compared it with previous data under the toughest water condition that plants can endure and survive (35%) (Su, unpublished data). To understand the
temporal expression pattern, samples from wild-type inflorescences in *Arabidopsis* under different water conditions (as 35%, 50% and 90% water as the total dry soil weight) were collected at a series of time points (after 0, 3, 4, 5 and 10 days drought stress). According to these criteria, total 1810 genes were differentially expressed in any of the four time points, Day 3, 4, 5 and 10.

To identify the inflorescence genes that were involved in drought response, Affymetrix GeneChips were used in this study. For each condition, we had at least two biological replicates for each time point and all the results were highly reproducible (all Pearson correlation coefficients > 0.97). To focus on the genes significantly changed under drought stress compared with well-watered condition, we only selected those whose expressions have: 1) more than two fold changes; 2) with q-values less than 0.05.

### 3.2.2 Known stress responsive pathways activated by moderate drought

By Day 3 when the soil moisture reached 50%, 421 genes were found to be up-regulated compared with Day 0 (MD3/C0) (Figure 4-6). To exclude those genes with expression shifts primarily due to developmental changes, we further compared the Day 3 to Day 0 in the control group and 181 of the 421 genes were identified as developmental related genes. We further compared the control and drought treated plants at Day 3, 154 genes were found significantly changed. 87 genes were identified in both MD3/C0 and MD3/C3 group and thus labeled as moderate drought induced genes. Take a closer look at the pathways activated at Day 3, we found that known genes involved in the plant response to water-deprivation, cold, salt and abscisic acid (ABA) stimulus were enriched in this group (Figure 4-7). This result is not surprising because all the stresses mentioned above could lead to loss of water in cell and the ABA stimulus are the most immediate signal responsive to dehydration.
We also studied those genes suppressed under drought condition within the first three days. Excluding the developmental related genes, 102 genes were identified as down-regulated genes in both comparisons (MD3/C0 and MD3/C3) (Figure 4-6). The GO analysis showed that those genes responsive to light, heat, oxidative stress and chemical stimulus were enriched (Figure 4-8).

Further analysis of the transcriptome from later time points revealed that more genes had significant expression pattern changes. Compared with the starting point, the drought treated group had an increasing number of differentially expressed genes, from 421 (at Day 3) up to 756 (at Day 4) and reaching its summit 1018 (at Day 5), then going back to 442 (at Day 10). We also compared the drought sample with more similar stage in the control group: MD 3, 4 and 5 with C3; MD10 with C10. The total number of genes with different expression is less than the ones previous identified, slowly increasing from D3 (154) to D10 (350), suggesting a continuous reaction in response to drought stress.

### 3.2.3 The Cis-element analysis shown enrichment of known binding motif

Many studies related to environmental stresses have shown that many cis-elements have essential functions in response to various signals in the transcriptional regulatory network, such as the DRE, MYBR and ABRE motifs (Golldack et al., 2011b; Huang et al., 2011; Rabbani et al., 2003; Salekdeh et al., 2005; Seki et al., 2002b; Shinozaki et al., 2003). To find the key motifs responsible for the plant response to moderate drought, we search the known motifs in the putative promoter sequence of all the differentially expressed genes (1kb upstream of start codon). We found 270 genes with ABREs (1580 with the core motif ACGT), 1172 with MYB binding site (WAACCA), as well as 236 with DRE motif (RCCGAC) in their putative promoter sequences. In addition to these known binding motifs involved in drought response, we also
search other known cis-acting regulatory elements of transcription factor family: NAC family (1327 with its core binding motif: CACG), MYC of bHLH family (1716 with canon E-box: CANNTG and 447 with core motif G-box: CATGTG) and WRKY (332 with its binding site: TGACY). Besides, several known consensuses involved in transcriptional activation were also identified in the putative sequences, such as TATA-box and CAAT-box. All the promoter sequences have at least one in their promoters. We compared those with all the genes on the chip, no obvious enrichment was found, suggesting a complex network involving multiple key regulators.

3.3 Comparison between moderate-drought and severe-drought responsive genes

3.3.1 Transcriptome analyses of inflorescences under 35% drought condition

Data from severe drought condition were also analyzed at Day 3, 4, 5 and 10. About 50% of the number of total genes experienced expression alteration under severe drought condition (35% soil moisture) (moderate drought vs. severe drought: 1810/4036). At day 3, when the water content of the soil reached 35%, 23 and 265 genes were up- and down-regulated respectively with more than two-fold change (Q<0.05). Only one day later, the number of differentially expressed genes went up to 925 (541 up regulated and 384 down-regulated). And twice as many more genes were changed on day 2. However, there were not many more genes differentially expressed during day 2 to day 7, suggesting most of the response was accomplished within 2 days.

We observed four clusters using hierarchical clustering method, and then applied K-mean method to cluster all 1810 genes. From the figure 4-8, we can see that genes in each cluster have distinct expression pattern. For the first cluster, the highest expression level was observed at C0 (Day 0 without any treatment), suggesting their putative function in early inflorescence
development. The expression of genes in the second cluster changed dramatically under the severe drought condition, but not as much under moderate condition. Interestingly, genes in the third group reached their expression peak at MD5 (Day 5 under the moderate drought condition); however, their expression levels were not very high under the severe drought condition. Genes in the last group had elevated expression level under both moderate and severe drought conditions, and the expression was induced to higher levels by the severe drought condition.

To compare putative gene functions between these clusters, we examined the GO categories for the four clusters. In all four groups, the genes responsive to stimulus were enriched, consistent with their expression changes under drought condition. In terms of molecular function, we found that no GO category was enriched in the 1st cluster, suggesting that the expression changes observed in this cluster might not be closely related to drought response. On the contrary, genes encoding transcription factors and transporters were enriched in the 2nd cluster, consistent with the increasing expression as the drought condition intensified. Genes annotated with transcription factor and catalytic activities were enriched in the third cluster, whose member shared a similar expression pattern that the expression induced by drought during initial response (before day 5 under moderate drought and before day 3 under severe drought). Genes in the catalytic activities and transporters categories were also enriched in the 4th cluster, consistent with their elevated expression under both drought conditions.

To investigate for possible relationship between the response time and the biological function, we also examined the GO annotations for genes up- and down-regulated at different time points. We found that genes encoding hydrolase and kinase were enriched in the group that quickly respond to mild drought with reduction of gene expression on day 0. And more genes with other binding activities were differentially expressed in both directions during the following day. Transcription factors were also over-represented in these groups. Interestingly, there are more transcription factors with increased but not reduced expression and this enrichment could be
observed on day 2 but not at later time points. This indicates possible functions of those genes within a limited time period. In addition, genes with other enzyme activities were over-represented in the group that had less expression from day 1 through day 7. As well, transferases were also enriched in day 7, suggesting a longer response time and possible later functions.

### 3.3.2 Expression levels of transcription factors in proportional to the increasing drought severity

As observed in our current study, the impacts of droughts on reproductive organs become more severe as the severity of drought increases. However, few studies have reported which genes might be responsible for the morphological changes. Here we compared the inflorescence transcriptome under severe and mild drought conditions and identified some key genes that might be able to explain the different impacts of the two conditions on plant growth. Most of these genes were included in the 2nd cluster. From the GO results, we learned that transcription factors and transporters might be mostly responsive to the drought intensity. We focused on those genes that have significantly increased expression under severe drought compared with mild drought and others that have preferential expression in mild drought compared with well-watered condition (q<0.05, 2 fold change). At Day 3, no genes satisfying these criteria were found but later several genes with this expression pattern were identified (14 genes at Day 4, 62 genes at Day 5 and 26 genes at Day 10). This trend is consistent with our observation that the floral development resumed at Day 10 after a short pause following the initial drought treatment (Figure 4-3).

Many genes within this group had important molecular functions, such as transcription regulation (table 4-1). For example, ATHB-7 and ATHB-12, containing a homeodomain closely linked to a leucine zipper motif belonging to the same phylogenetic cluster γ, showed preferential
expression at Day 4 which is consistent with a previous study in other vegetative tissues (root, leaf and stem) (Lee and Chun, 1998). Both transcription factors are responsive to ABA and water deficiency. Two other homeodomain factors ATHB-2 and ATHB-5 were also in this group, and the phylogenic analysis showed that they belong to a different group δ closed to the one of ATHB-7 and -12 on the basis of the similarity of their homeodomain (Henriksson et al., 2005). The four transcription factors mentioned above all belong to the same HD zip class I, most of which have been shown to be responsive to ABA and salt stress using seedling RNA for RT-PCR (Henriksson et al., 2005). The tissue specific analysis also suggested that these four genes had a wide expression pattern. Thus we speculate that they are involved in the conventional pathway dealing with drought stress.

Another transcription factor family with many members found in this group is NAC family, which consists of 105 members in Arabidopsis (Ooka et al., 2003). Three members of the NAC family were identified as drought severity associated genes, NAC19, NAC47 and NAC92. The NAC19 and NAC47 belong to the atNAC3 group while NAC92 belongs to NAM group. NAC92 has been known to function in apical meristem (SAM) formation and development, and is redundant to CUC1 (Takada et al., 2001). Functional investigation suggested that NAC92 regulates senescence in response to salt by controlling several downstream genes in a stage dependent way, similar to what we have observed in this drought study (Balazadeh et al., 2010). No study has explored the function of the other two NAC members, and our results might shed some interesting light on the understanding of stress induced senescence pathways.

Other transcription factors in this group included HSF1, PLATZ, OZF1 and OZF2. Both OZF factors are ABA-responsive and recent studies indicated that OZF2 is involved in the ABI2-mediated signaling pathway (Huang et al., 2012; Sanchez et al., 2004). Both OZFs have two CCCH motifs and they are the closely related homologs (Wang et al., 2008a). It is very possible that the two OZF function redundantly to assist the plant in response to various stresses. HSF1 is
involved in the response to a combination of drought and heat stress but more thorough experimental confirmation is needed (Rizhsky et al., 2004). No functional information of PLATZ has been obtained and further investigation is needed.

3.3.3 Plants adapt to different drought conditions by activating transportation, ABA-dependant pathway and embryogenesis.

In addition to the transcription factors, many other genes also showed expression alteration responsive to drought severity. Not surprisingly, many genes encoding transporters had elevated expression (table 4-2). We also found that genes involved in male reproduction, late embryogenesis and seed dormancy were activated. Genes encoding four late embryogenesis abundant (LEA) proteins which protect other proteins from desiccation were in this group and some of them have known function in response to drought (table 4-3) (Rizhsky et al., 2004). Interestingly, a gene called MATERNAL EFFECT EMBRYO ARREST 25 (MEE25) with UDP-glucose 4-epimerase activity was also in this group and a previous study suggested that it functions in male reproductive development (Boavida et al., 2009). Two other male reproductive genes were also found in this group, partially explaining the delayed impact of drought on flowering observed in previous morphological analysis. The identification of these functional genes suggested that both drought avoidance genes and drought tolerance genes were involved in the response of inflorescence to drought stress.

The crosstalk between multiple pathways in response to different stresses has been shown in many studies on vegetative development (Rizhsky et al., 2004). It is not surprising that many known genes involved in known stress responsive pathways were also in this group (table 4-4). The ABA signaling pathway is one of the key mechanisms that play an important role in response
to drought stress throughout plant evolution history (Fujita et al., 2011; Hauser et al., 2011a; Raghavendra et al., 2010; Seki et al., 2002a). In our study, we also found eight genes in ABA signaling pathway with increasing expression levels as the drought severity became more intensive. Besides, genes responsive to cold and salt stresses were also found in our study, suggesting crosstalk between different pathways also exists in inflorescences.

### 3.3.4 Genes activated by moderate drought but not by severe drought

As mentioned in the morphological analysis, there is no obvious reduction of yield under moderate drought condition (50%) but there is a significant loss in severe drought (35%). It would be very interesting to investigate whether there is any gene that is induced under moderate but not under severe drought condition. From our microarray data, we observed that genes in the third cluster shared a similar expression pattern that reached the highest expression levels under moderate drought. We found several interesting genes which might assist us in understanding the fundamental mechanisms. Among the 55 genes with preferential expression in MD compared with the control and SD using stringent criteria on day 5 (2 fold changes, q<0.05), four Nuclear Factor Y transcription factor subunits (NY-Fs, also known as CCAAT-bind factors) were up-regulated (Figure 4-10). In *Arabidopsis*, 36 genes for CCAAT-bind factors were identified (10 NF-YA, 13 NF-YB and 13 NF-YC) and they are thought to act as heterotrimer. After the first identification of CCAAT-binding factors in *Arabidopsis* (Edwards et al., 1998), a few studies have reported about their functions in the development process especially in flowering (Cai et al., 2007; Kumimoto et al., 2008); the adaption to various stresses, including drought (Chen et al., 2007), osmotic stress (Chen et al., 2007) and nutrient deprivation (Nilsson et al., 2010b); and the response to many plant hormones, such as ABA (Yamamoto et al., 2009). We further investigated the expression of all the genes of NF-Y family and found that seven of the ten NF-YA were up-
regulated under moderate drought condition (Figure 4-10). The hierarchical clustering on the basis of their expression levels suggested that NF-YA subgroup is more responsive to the moderate drought condition. Our results provide a possible explanation for the continuous flowering in moderate lack of water resource that NF-YA (CCAAT-binding factors) might be activated to maintain reproductive growth by slowing down the flowering process.

3.3.5 Signal pathways in response to moderate drought and severe drought

In our study, 4788 genes were identified as differentially expressed under SD condition and only 1810 genes under MD condition. Among those genes, 1467 genes were detected in both studies. Using AgriGO software, we found different categories were enriched, including transcription regulator, transporter, enzyme and catalytic activity (Figure 4-11). The enrichment of similar categories was also observed in the group only differentially expressed under SD condition compared with control, except for the GO category with binding activity (Figure 4-12). This result indicated that the mechanisms plant adopted to deal with dehydration might be very similar.

85 transcription factors were differentially expressed under both MD and SD conditions, including DREB, NAC, AP2/ERF, MYB, bZIP, PLATZ, homeodomain, WRKY, zinc finger and HSF gene family. We also analyzed those genes that are only responsive to severe drought and found 175 transcription factors in this group. In addition to gene families mentioned above, genes in the AGL and BEH families were also identified in this category. The AGL family is commonly involved in the floral developmental process, thus it is consistent with our observation that only severe drought but not moderate drought, significantly influences the essential developmental process and cause loss of yield.
To find the putative transcriptional regulatory network in response to drought, we investigated the putative promoter sequences of genes differentially expressed under drought condition. We searched for more than one hundred known binding motifs (from PLANTCARE) in the four clusters. Interestingly, the ABRE and ABRE-like binding motif were enriched in the 2\textsuperscript{nd} and 3\textsuperscript{rd} clusters (both within the first 0.5kb and 1kb). It is not surprising that ABA-independent pathway is very important in both moderate and severe drought response. Other binding motifs, such as E-box and G-box, were also enriched in these two clusters, suggesting putative transcription factors, such as those in the bHLH family, controlling some genes in the two clusters. The enrichment of cis-regulatory elements is not as significant when we searched in longer sequence (3kb upstream of genes in each cluster). Though the NF-Y family members were enriched in the 2\textsuperscript{nd} cluster, the binding site of NF-Y was not obviously enriched in any clusters using different length of putative promoter sequence. However, we still find many genes in this cluster with the CCAAT motif.

After the investigation of the promoter sequences and the expression data from microarray, we now proposed an updated regulatory network in response to drought (Figure 4-13). In addition to the known ABA-dependent and ABA-independent pathways, the CCAAT-binding factors might also play important roles in drought response. As well, our microarray data suggested, \textit{DREB1A}, which was identified as a cold responsive gene, might be also involved in the drought stress pathways.
Plants were grown under six different conditions (at 30%, 35%, 40%, 50%, 70% and 90% of soil moisture) and treated till two weeks after bolting. The Day 0 means the start time point of drought treatment (around bolting time).
Plants were treated with different drought severities (at 30%, 35%, 40%, 50%, 70% and 90% of soil moisture) for ten days. The total height and the size of inflorescences reduced obviously as the severity increased.

Figure 4-2. The whole plants and inflorescences after ten days treatment under six conditions
Figure 4-3. The accumulated flower numbers were affected by drought stress.

Plants under different drought severities (at 30%, 35%, 40%, 50%, 70% and 90% of soil moisture) showed reduction of flower number.
Figure 4-4. Drought stress affected the number of siliques

Plants under different drought severities (at 30%, 35%, 40%, 50%, 70% and 90% of soil moisture) showed reduction of yield.
Figure 4-5. Drought stress affected the number of siliques.

Plants under different drought severities (at 30%, 35%, 40%, 50%, 70% and 90% of soil moisture) showed reduction of seeds number per silique.
Figure 4-6. Venn diagram illustrating the number of genes that are induced under different condition at certain time point

(A) Genes up-regulated after drought treatment compared with those up-regulated due to developmental changes. (B) Genes down-regulated after drought treatment compared with those down-regulated due to developmental changes.

Figure 4-7. GO category of biological process for genes up-regulated due to drought stress.
Figure 4-8. GO category of biological process for genes down-regulated due to drought stress
Figure 4-9. Genes differentially expressed under mild drought condition

Heat map of 1810 genes differentially expressed under moderate drought condition. Red color represents high expression while green color represents low expression. K-mean clustering was performed on transcripts ratios of all conditions. C represents control: well-watered group; MD represents moderate drought; SD represents severe drought.
Figure 4-10. Hierarchical clustering of NF-Y gene family under drought condition

Red color represents high expression while green color represents low expression. Hierarchical clustering was performed on transcripts ratios of all conditions. C represents control: well-watered group; MD represents moderate drought; SD represents severe drought.
Figure 4-11. Percent of genes in each GO category of molecular functions for genes differentially expressed under both moderate and severe drought condition.
Figure 4-12. Percent of genes in each GO category of molecular functions for genes differentially expressed only under severe but not moderate drought condition.
Figure 4-13. Signaling pathways in response to drought stresses

Positive gene regulation is represented by arrows. *cis*-regulatory motifs are surrounded by rectangles. Genes with known function in drought response is in black and newly founded signaling pathways were labeled with color.
Table 4-1. Expression of genes known as transcription factors.

<table>
<thead>
<tr>
<th>gene ID</th>
<th>Description</th>
<th>C0</th>
<th>C3</th>
<th>C10</th>
<th>MD3</th>
<th>MD4</th>
<th>MD5</th>
<th>MD10</th>
<th>SD3</th>
<th>SD4</th>
<th>SD5</th>
<th>SD10</th>
</tr>
</thead>
<tbody>
<tr>
<td>At2g18550</td>
<td>ATHB-2/HB21</td>
<td>4.9</td>
<td>4.3</td>
<td>4.3</td>
<td>5.0</td>
<td>6.1</td>
<td>6.4</td>
<td>5.5</td>
<td>4.5</td>
<td>7.7</td>
<td>9.8</td>
<td>8.0</td>
</tr>
<tr>
<td>At2g46680</td>
<td>ATHB-7</td>
<td>5.8</td>
<td>5.7</td>
<td>5.9</td>
<td>6.9</td>
<td>7.8</td>
<td>8.9</td>
<td>6.8</td>
<td>6.7</td>
<td>9.4</td>
<td>11.5</td>
<td>10.9</td>
</tr>
<tr>
<td>At2g61890</td>
<td>ATHB-12</td>
<td>5.3</td>
<td>6.8</td>
<td>6.6</td>
<td>6.4</td>
<td>7.3</td>
<td>9.1</td>
<td>6.1</td>
<td>7.5</td>
<td>8.9</td>
<td>11.3</td>
<td>11.2</td>
</tr>
<tr>
<td>At1g52890</td>
<td>NAC19</td>
<td>7.1</td>
<td>7.8</td>
<td>8.7</td>
<td>9.3</td>
<td>9.9</td>
<td>9.5</td>
<td>9.0</td>
<td>8.4</td>
<td>10.0</td>
<td>11.4</td>
<td>9.7</td>
</tr>
<tr>
<td>At3g04070</td>
<td>NAC47</td>
<td>5.0</td>
<td>4.8</td>
<td>5.7</td>
<td>5.6</td>
<td>5.9</td>
<td>6.4</td>
<td>5.2</td>
<td>4.9</td>
<td>6.8</td>
<td>8.8</td>
<td>6.7</td>
</tr>
<tr>
<td>At5g39610</td>
<td>NAC92/CUC2</td>
<td>5.1</td>
<td>5.3</td>
<td>5.7</td>
<td>6.0</td>
<td>6.3</td>
<td>6.6</td>
<td>6.1</td>
<td>5.9</td>
<td>6.6</td>
<td>7.7</td>
<td>5.6</td>
</tr>
<tr>
<td>At2g19810</td>
<td>OZF1</td>
<td>8.4</td>
<td>8.4</td>
<td>9.0</td>
<td>8.9</td>
<td>9.3</td>
<td>9.6</td>
<td>8.6</td>
<td>8.8</td>
<td>9.7</td>
<td>10.7</td>
<td>11.4</td>
</tr>
<tr>
<td>At4g29190</td>
<td>OZF2</td>
<td>7.6</td>
<td>7.8</td>
<td>8.4</td>
<td>8.2</td>
<td>8.8</td>
<td>9.3</td>
<td>8.2</td>
<td>8.4</td>
<td>9.9</td>
<td>11.6</td>
<td>11.5</td>
</tr>
<tr>
<td>At3g24520</td>
<td>HSF1</td>
<td>5.1</td>
<td>5.3</td>
<td>5.6</td>
<td>5.9</td>
<td>6.4</td>
<td>6.6</td>
<td>6.3</td>
<td>5.9</td>
<td>7.9</td>
<td>9.5</td>
<td>10.2</td>
</tr>
<tr>
<td>At1g76590</td>
<td>PLATZ TF</td>
<td>5.5</td>
<td>5.1</td>
<td>5.4</td>
<td>6.4</td>
<td>7.3</td>
<td>6.9</td>
<td>5.7</td>
<td>5.2</td>
<td>8.3</td>
<td>10.7</td>
<td>10.7</td>
</tr>
</tbody>
</table>

All expression values are log2 ratio.

Table 4-2. Expression of genes involved in transportation.

<table>
<thead>
<tr>
<th>gene ID</th>
<th>Target Description</th>
<th>C0</th>
<th>C3</th>
<th>C10</th>
<th>MD3</th>
<th>MD4</th>
<th>MD5</th>
<th>MD10</th>
<th>SD3</th>
<th>SD4</th>
<th>SD5</th>
<th>SD10</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g02390</td>
<td>acyltransferase2</td>
<td>4.8</td>
<td>5.1</td>
<td>5.2</td>
<td>5.3</td>
<td>5.6</td>
<td>6.3</td>
<td>5.8</td>
<td>5.4</td>
<td>6.3</td>
<td>7.8</td>
<td>7.7</td>
</tr>
<tr>
<td>At5g26340</td>
<td>hexose transporter</td>
<td>6.3</td>
<td>6.5</td>
<td>7.2</td>
<td>8.1</td>
<td>8.6</td>
<td>7.5</td>
<td>7.9</td>
<td>6.2</td>
<td>7.9</td>
<td>8.8</td>
<td>8.4</td>
</tr>
<tr>
<td>At4g35190</td>
<td>decarboxylase</td>
<td>7.5</td>
<td>7.3</td>
<td>8.2</td>
<td>9.1</td>
<td>9.3</td>
<td>9.8</td>
<td>9.2</td>
<td>8.3</td>
<td>10.0</td>
<td>10.9</td>
<td>9.6</td>
</tr>
<tr>
<td>At3g43270</td>
<td>pectinesterase</td>
<td>6.7</td>
<td>7.8</td>
<td>7.5</td>
<td>8.5</td>
<td>8.8</td>
<td>8.8</td>
<td>8.7</td>
<td>7.8</td>
<td>9.0</td>
<td>9.8</td>
<td>9.2</td>
</tr>
<tr>
<td>At1g32450</td>
<td>PTR2-B</td>
<td>7.2</td>
<td>7.6</td>
<td>7.3</td>
<td>8.7</td>
<td>9.1</td>
<td>9.1</td>
<td>8.7</td>
<td>8.2</td>
<td>8.9</td>
<td>10.2</td>
<td>9.2</td>
</tr>
<tr>
<td>At5g47560</td>
<td>dicarboxylate cotransporter</td>
<td>7.2</td>
<td>8.3</td>
<td>8.1</td>
<td>8.2</td>
<td>8.7</td>
<td>9.5</td>
<td>8.8</td>
<td>8.5</td>
<td>9.2</td>
<td>10.6</td>
<td>10.4</td>
</tr>
<tr>
<td>At1g78070</td>
<td>Transducin</td>
<td>5.7</td>
<td>6.1</td>
<td>6.2</td>
<td>6.3</td>
<td>6.5</td>
<td>7.1</td>
<td>5.9</td>
<td>6.3</td>
<td>6.9</td>
<td>8.6</td>
<td>8.1</td>
</tr>
<tr>
<td>At2g41190</td>
<td>amino acid transporter</td>
<td>7.2</td>
<td>7.3</td>
<td>7.1</td>
<td>7.7</td>
<td>8.2</td>
<td>8.6</td>
<td>7.7</td>
<td>7.5</td>
<td>9.6</td>
<td>11.9</td>
<td>11.6</td>
</tr>
<tr>
<td>At5g01520</td>
<td>zinc ion binding</td>
<td>6.1</td>
<td>6.3</td>
<td>6.5</td>
<td>6.8</td>
<td>7.3</td>
<td>8.0</td>
<td>6.8</td>
<td>6.0</td>
<td>8.6</td>
<td>10.4</td>
<td>9.5</td>
</tr>
</tbody>
</table>
Table 4-3. Expression of genes involved in embryogenesis and reproductive development.

<table>
<thead>
<tr>
<th>gene ID</th>
<th>Target Description</th>
<th>C0</th>
<th>C3</th>
<th>C10</th>
<th>MD3</th>
<th>MD4</th>
<th>MD5</th>
<th>MD10</th>
<th>SD3</th>
<th>SD4</th>
<th>SD5</th>
<th>SD10</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g01470</td>
<td>LEA14</td>
<td>7.5</td>
<td>7.3</td>
<td>7.3</td>
<td>8.4</td>
<td>8.9</td>
<td>8.8</td>
<td>8.7</td>
<td>8.1</td>
<td>9.5</td>
<td>10.8</td>
<td>10.8</td>
</tr>
<tr>
<td>At2g35300</td>
<td>LEA18</td>
<td>4.4</td>
<td>5.0</td>
<td>5.5</td>
<td>6.0</td>
<td>6.5</td>
<td>7.4</td>
<td>6.7</td>
<td>5.3</td>
<td>8.6</td>
<td>10.7</td>
<td>10.1</td>
</tr>
<tr>
<td>At5g06760</td>
<td>LEA4-5</td>
<td>6.6</td>
<td>7.3</td>
<td>6.5</td>
<td>7.4</td>
<td>8.0</td>
<td>8.6</td>
<td>8.0</td>
<td>7.3</td>
<td>9.7</td>
<td>11.9</td>
<td>11.8</td>
</tr>
<tr>
<td>At1g52690</td>
<td>DUO 7</td>
<td>4.6</td>
<td>5.8</td>
<td>6.7</td>
<td>6.3</td>
<td>8.6</td>
<td>10.1</td>
<td>7.8</td>
<td>7.8</td>
<td>11.9</td>
<td>13.8</td>
<td>13.4</td>
</tr>
<tr>
<td>At1g64110</td>
<td>DUO1-activated ATPase 1</td>
<td>4.4</td>
<td>4.5</td>
<td>4.8</td>
<td>4.9</td>
<td>5.8</td>
<td>6.6</td>
<td>5.3</td>
<td>4.6</td>
<td>7.5</td>
<td>10.1</td>
<td>9.9</td>
</tr>
<tr>
<td>At2g34850</td>
<td>MEE25</td>
<td>6.1</td>
<td>6.4</td>
<td>6.0</td>
<td>7.2</td>
<td>7.4</td>
<td>7.6</td>
<td>6.7</td>
<td>6.9</td>
<td>7.3</td>
<td>8.8</td>
<td>9.3</td>
</tr>
<tr>
<td>At4g14020</td>
<td>pollen tube</td>
<td>5.1</td>
<td>5.0</td>
<td>5.1</td>
<td>6.3</td>
<td>6.4</td>
<td>6.6</td>
<td>5.5</td>
<td>6.1</td>
<td>7.5</td>
<td>8.8</td>
<td>7.9</td>
</tr>
<tr>
<td>At2g37870</td>
<td>seed storage 2S</td>
<td>6.1</td>
<td>5.8</td>
<td>5.4</td>
<td>6.2</td>
<td>7.8</td>
<td>8.7</td>
<td>6.7</td>
<td>6.2</td>
<td>9.3</td>
<td>11.0</td>
<td>11.6</td>
</tr>
</tbody>
</table>

Table 4-4. Expression of genes involved in known stress responsive genes.

<table>
<thead>
<tr>
<th>gene ID</th>
<th>Target Description</th>
<th>C0</th>
<th>C3</th>
<th>C10</th>
<th>MD3</th>
<th>MD4</th>
<th>MD5</th>
<th>MD10</th>
<th>SD3</th>
<th>SD4</th>
<th>SD5</th>
<th>SD10</th>
</tr>
</thead>
<tbody>
<tr>
<td>At5g57050</td>
<td>ABI2</td>
<td>7.1</td>
<td>6.7</td>
<td>7.0</td>
<td>7.6</td>
<td>8.2</td>
<td>8.4</td>
<td>7.7</td>
<td>7.5</td>
<td>8.8</td>
<td>10.4</td>
<td>10.3</td>
</tr>
<tr>
<td>At1g69260</td>
<td>AFP1</td>
<td>5.9</td>
<td>5.9</td>
<td>6.0</td>
<td>6.5</td>
<td>7.1</td>
<td>8.1</td>
<td>7.2</td>
<td>7.3</td>
<td>8.5</td>
<td>10.1</td>
<td>10.5</td>
</tr>
<tr>
<td>At3g29575</td>
<td>AFP3</td>
<td>8.1</td>
<td>8.0</td>
<td>8.0</td>
<td>8.6</td>
<td>9.2</td>
<td>9.4</td>
<td>9.3</td>
<td>8.7</td>
<td>9.8</td>
<td>10.8</td>
<td>10.8</td>
</tr>
<tr>
<td>At5g66400</td>
<td>ATDI8</td>
<td>6.4</td>
<td>7.0</td>
<td>7.0</td>
<td>6.9</td>
<td>8.9</td>
<td>10.6</td>
<td>7.6</td>
<td>7.8</td>
<td>12.2</td>
<td>14.1</td>
<td>14.1</td>
</tr>
<tr>
<td>At5g64260</td>
<td>Exordium like 2</td>
<td>9.4</td>
<td>9.7</td>
<td>11.1</td>
<td>9.5</td>
<td>10.2</td>
<td>10.7</td>
<td>9.8</td>
<td>9.6</td>
<td>11.4</td>
<td>12.8</td>
<td>12.5</td>
</tr>
<tr>
<td>At5g59220</td>
<td>PP2C GENE 1</td>
<td>8.0</td>
<td>8.6</td>
<td>8.7</td>
<td>8.7</td>
<td>9.4</td>
<td>9.8</td>
<td>9.3</td>
<td>9.0</td>
<td>10.6</td>
<td>12.1</td>
<td>12.2</td>
</tr>
<tr>
<td>At1g07430</td>
<td>PP2C GENE 2</td>
<td>6.1</td>
<td>5.9</td>
<td>5.8</td>
<td>8.9</td>
<td>9.4</td>
<td>9.1</td>
<td>8.3</td>
<td>7.6</td>
<td>9.0</td>
<td>10.9</td>
<td>10.8</td>
</tr>
<tr>
<td>At3g11410</td>
<td>PP2CA</td>
<td>8.7</td>
<td>9.1</td>
<td>9.9</td>
<td>9.8</td>
<td>10.2</td>
<td>10.4</td>
<td>10.1</td>
<td>9.7</td>
<td>11.0</td>
<td>12.2</td>
<td>12.0</td>
</tr>
<tr>
<td>At3g50970</td>
<td>LTI30-cold</td>
<td>4.7</td>
<td>4.7</td>
<td>6.7</td>
<td>9.3</td>
<td>10.3</td>
<td>9.9</td>
<td>6.8</td>
<td>5.9</td>
<td>9.5</td>
<td>11.5</td>
<td>11.3</td>
</tr>
<tr>
<td>At4g30960</td>
<td>CIPK6-salt</td>
<td>8.8</td>
<td>9.6</td>
<td>9.7</td>
<td>10.4</td>
<td>10.5</td>
<td>10.8</td>
<td>10.3</td>
<td>9.7</td>
<td>10.8</td>
<td>11.9</td>
<td>12.0</td>
</tr>
<tr>
<td>At5g02020</td>
<td>salt induced serine rich</td>
<td>6.6</td>
<td>7.2</td>
<td>6.9</td>
<td>6.9</td>
<td>7.9</td>
<td>9.1</td>
<td>7.9</td>
<td>7.0</td>
<td>10.4</td>
<td>12.3</td>
<td>11.9</td>
</tr>
</tbody>
</table>
4. Conclusions

In this study, we analyzed the morphological changes in reproductive development and found two key thresholds of water content (30% and 50%). We then identified genes whose expression levels were changed under severe and moderate drought using microarray chips. Further analyses of the microarray data showed some key regulators with putative functions in plant response to drought via ABA-dependent and –independent pathways. Moreover, our results explained the phenomenon that flower development was impacted under severe but not moderate drought conditions at the molecular level. A complicated regulatory network in response to drought stress was then elaborated, proposing hypothesis to be tested by future experiments.

5. Materials and Methods

5.1 Plant material and stress treatments

In this study, morphological analyses under different water conditions were performed on Col-0, whose genome has been completely sequenced. Seeds were directly planted into pots containing 100g soil consisting of soil and greens grades by volume ratio 3:2. After two day’s vernalization in dark at 4°C all the plants were grown in growth chamber at 22 °Ccondition 16h/8h until the flowering stage when plants were subjected to different types of drought treatment when the main stem is about 1cm high (as described in Su’s paper).

The Moderate drought (MD) and severe drought (SD) treatments started by withholding water. The relative soil moisture content was reduced to the expected degree (MD: 50% and SD: 30%) three days after the starting point (D0). We maintained the soil water condition (30%, 50 % and 90% in control group) until almost all the siliques were matured and ready to be harvested.
about 50 days after plant). Unopened flower samples were collected, from both drought treated groups and control groups, at day 3, 4, 5 and 10 after drought. Two biological replicates from the inflorescences were collected at each time point from each group.

5.2 RNA isolation and hybridization to the ATH1 GeneChip

Total RNA was isolated from unopened flower buds using RNeasy columns following instructions from the manufacturer (Qiagen, Valencia, CA, USA). All the Hybridization, washing, staining, scanning and data collection were performed at the Genomics Core Facility, Pennsylvania State University, University Park.

5.3 Microarray Analysis

Two biological replicates were isolated from each condition. We used Bioconductor package in R by RMA to normalize all the data (Zhang et al., 2005a), and converted all expression values to logarithms base 2. We then used to compare signals from control and well-watered inflorescences using LIMMA library. Only genes with more than two-fold changes were selected in addition to the statistical criterion: q-value (FDR) less than 0.05 (Storey and Tibshirani, 2003).

Hierarchical clustering of co-expressed genes was performed by MeV 4.6 (Yang et al., 2007). We first normalized the expression data by row (divided by average expression of the gene under all condition) and then used Euclidean distance metric to conduct this analysis. For the identification of the functions of the differentially expressed genes, the annotations of genes on ATH1 microarray chip were downloaded from Affymetrix website and we used the GO categorization function on TAIR website (Ma, 2009). To verify whether one category is enriched
compared with the whole genome, we applied hypergeometric test and only the categories with p-value less than 0.05 were called statistically enriched (Zhou and Su, 2007).

5.4 Cis-regulatory element analysis and GO analysis

Possible promoter sequences of all genes on the microarray chip (1kb upstream of the start codon) were obtained from the TAIR website. The numbers of binding sites of different transcriptional regulators were then counted. The identification of cis-regulatory binding site was conducted by perl (Jurgens et al., 2011). The binding motifs were obtained from Gene Regulation and PlantCARE (Parenicova et al., 2003b). The Gene Ontology (GO) analysis was done by the agriGO software (Du et al., 2010).

6. Acknowledgements

I greatly appreciate the help of Dr. Craig Praul in performing microarray hybridizations. We also thank Ms. Yi Hu for plant care and lab management. I appreciate the suggestions from Drs. Xiaofan Zhou, Xinwei Han and Yazhou Sun on microarray data analysis. This work was supported by a US Department of Energy grant to H.M. and funds from Department Biology and the Huck Institutes of the Life Sciences, the Pennsylvania State University, and Fudan University.
7. References


binding factor that enhances abiotic stress signaling in rice. Plant Molecular Biology 72, 557-566.


A comparative proteomics approach to identifying salt and drought tolerance genes in plants. Mol Cell Proteomics 4, S258-S258.
Curriculum Vitae

Xuan Ma

Education

2007-present  The Pennsylvania State University, University Park, PA
               Ph.D. in interdisciplinary program of Cell and Developmental Biology,
               Minor in Computational Science, Cumulative GPA: 3.9/4.0

2003-2007    China Agricultural University, Beijing, China
               Bachelor of Science in Biological Sciences  Cumulative GPA: 3.50

Publications

Ma X., Feng B., and Ma H. (2012) AMS-dependent and independent regulation of anther
transcriptome and comparison with those affected by other *Arabidopsis* anther genes. BMC, Plant
biology, 12:23.

Zahn L.M.*, Ma X.*, Altman N.S., Zhang Q., Wall P.K., Tian D., Gibas C.J., Gharaibeh R.,
organs of the basal eudicot *Eschscholzia californica* as reference for floral evolutionary
developmental studies. Genome Biology. (*: co-first author) 11:R101.

Yang K., Xia C., Liu X.L., Dou X.Y., Wang W., Chen L.Q., Zhang X.Q., Xie L.F., He L., Ma X.,
DnaJ and PDI domains, leads to thermosensitive gametophytic male sterility in *Arabidopsis*. Plant
J. 57:870.

Ma X., Su Z., Sukiran L. and Ma H. Impacts of different drought severities on reproductive
organs in *Arabidopsis*. In preparation.

Su Z., Ma X., Guo H.H., Sukiran L., and Ma H. Developmental and transcriptomic adaptations to

Feng B.M., Lu D.H., Ma X. et al. Regulation of the male reproductive transcriptome in
*Arabidopsis thaliana* by DYT1 via multiple mechanisms. In preparation.

Teaching Experience:

2009-2011 Spring: Teaching Assistant for the Biology course, Function and Development of
2012 Spring: Teaching Assistant for the Biology course, Function and Development of
Organisms (Honor section) (BIOL240H). Responsibility: lab instruction.

Honors and awards

2009 and 2010  Travel grant from Huck Institute and Biology Department
2008-present  Graduate Research Assistantship, Penn State University
2007-2008    Cell and Developmental Biology Program Fellowship