

The Pennsylvania State University

The Graduate School

Department of Biology

**TRANSCRIPTIONAL AND POST TRANSCRIPTIONAL
REGULATION IN ARABIDOPSIS REPRODUCTIVE
DEVELOPMENT**

A Dissertation in Biology

by

Baomin Feng

©2008 Baomin Feng

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2008

The dissertation of Baomin Feng was reviewed and approved* by the following:

Hong Ma

Professor of Biology

Dissertation Advisor

Chair of Committee

Teh-hui Kao

Professor of Biochemistry and Molecular Biology

Richard Cyr

Professor of Biology

Paula McSteen

Assistant Professor of Biology

Douglas Cavener

Professor and Head of Biology

*Signatures are on file in the Graduate School

ABSTRACT

Sexual reproduction is important for propagation and requires a series of cell division and differentiation events. In flowering plants, it is incompletely understood how these events are genetically controlled. Because transcriptional and post transcriptional regulations are suggested to be main molecular mechanisms in these events, it is important to examine the changes of all transcripts (transcriptome) and proteins (proteome) during plant reproductive development. However, only limited studies have been done to analyze the transcriptome and proteome during plant reproductive development. Part of my thesis work is an analysis of the transcriptome in the tapetum, which is an important sporophytic tissue that contributes to microsporogenesis and pollen formation. It is known that DYSFUNCTIONAL TAPETUM1 (DYT1) and ABORTED MICROSPORES (AMS), two bHLH transcription factors, are required for the development and functions of the tapetum. However, it is poorly understood which genes are regulated by DYT1 and/or AMS. In Chapter 2, a transcriptome analysis of the *dyl1-1* and *ams* mutants by using microarray is presented. It was found that DYT1 positively and negatively regulates approximately 600 and 300 genes, respectively, and that AMS positively and negatively regulates ~600 and ~900 genes, respectively. DYT1 is required to control a broad spectrum of tapetum functions in part by activating at least 32 transcription factors including AMS. However, AMS regulates not only common but also different sets of genes compared to DYT1, which suggests AMS might have distinct functions. In order to differentiate the direct and indirect regulation by DYT1, I searched the binding sites of DYT1 by *in vitro* (Selex and gel-shift) and *in vivo* (ChIP) protein-DNA interaction experiments, which is described in Chapter 3. The results indicate that DYT1 recognizes the E-box (CANNTG) *in vitro* and binds to its own promoter and the promoters of downstream genes *in vivo*. These findings not only reveal the spectrum of functions regulated by *DYT1* but also suggest some mechanisms of the transcriptional controls in the tapetum. In Chapter 4, I present a systematic exploration of *Arabidopsis* floral proteome where I identify more than 2400 floral proteins and provide pieces of evidence for protein modifications (methylation, acetylation, and glycosylation etc.) of some proteins by using 2-Dimensional Gel Electrophoresis/mass spectrometry and Multidimensional Protein Identification Technique (MudPIT). These studies have provided new insights into transcriptional and post transcriptional regulation in plant reproductive development.

TABLE OF CONTENTS

List of figures.....	viii
----------------------	------

List of tables	ix
----------------------	----

Acknowledgements	x
------------------------	---

Chapter 1 Introduction	1
------------------------------	---

1.1 Significance and Objectives of Studying <i>Arabidopsis</i> Reproductive Development	2
1.2 An Overview of <i>Arabidopsis</i> Reproductive Development and Genetic Controls.....	3
1.3 Rationales to Study Anther Development	6
1.4 The Structure and Functions of <i>Arabidopsis</i> Anther.....	6
1.5 The Cell Division and Differentiation in Early Anther Development are regulated by Transcriptional Cascades and Cell-cell Signaling	8
1.6 Genetic Controls of the Tapetum Differentiation	12
1.7 The Tapetum Cells are Highly Active in the Biosynthesis and Secretion of Pollen Wall Components	18
1.8 Current Understanding about the Endothecium and the Middle layer: Structure, Functions and Genetic controls.....	22
1.9 Reference.....	26

Chapter 2 DYT1 and AMS regulate common and distinct sets of genes in anther...	37
--	----

2.1 Abstract.....	38
2.2 Introduction.....	39
2.3 Materials and Methods.....	40
2.3.1 Plant materials, growth conditions, and anther collection.....	40
2.3.2 Microarray analysis.....	41
2.3.3 RT-PCR and Real time PCR.....	41
2.4 Results.....	41

2.4.1 The transcriptomes of the <i>dyt1-1</i> and <i>ams</i> anther were uncovered.....	41
2.4.2 <i>DYT1</i> positively regulates at least 32 transcription factors, among which 14 are also positively regulated by <i>AMS</i>	42
2.4.3 The genes involved in lipid metabolism, cell wall and secondary metabolism are down regulated in <i>dyt1-1</i> , whereas heat shock protein genes are up regulated.....	43
2.4.4 <i>DYT1</i> and <i>AMS</i> have common and distinct downstream genes	45
2.5 Discussion.....	45
2.5.1 A transcriptional cascade regulated by <i>DYT1</i> involves at least 32 transcription factors in the tapetum	45
2.5.2 <i>DYT1</i> regulates important functions in the tapetum including lipid metabolism, cell wall modification, and secondary metabolism and transporter activities.	48
2.5.3 Transcriptional regulation by <i>DYT1</i> might be partially mediated by complexes of DYT1 and AMS, whereas AMS might have distinct functions.....	49
2.6 References.....	67

Chapter 3 Identification of the direct targets of DYT1 by Selex and ChIP.....74

3.1 Abstract.....	75
3.2 Introduction.....	75
3.3 Materials and Methods.....	77
3.3.1 Transgenic plant construction.....	77
3.3.2 Selex.....	78
3.3.3 Gel-shift assay.....	78
3.3.4 ChIP.....	79
3.3.5 Yeast two hybrid assay.....	79
3.4 Results.....	79
3.4.1 DYT1 binds to the E-boxes (CANNTG)	77
3.4.2 DYT1 binds to its own promoter <i>in vivo</i>	80

3.4.3 DYT1 binds to the promoters of <i>MYB35</i> , <i>MS</i> and other transcription factor genes except <i>AMS</i>	80
3.4.4 DYT1 binds to the promoter of chalcone synthase gene, a key gene in flavonol biosynthesis pathway.	81
3.4.5 DYT1, AMS and three anther preferential bHLHs interact in yeast two hybrid assays.....	81
3.5 Discussion.....	82
3.5.1 DYT1 might bind to E-boxes as dimmer.....	82
3.5.2 The establishment of the of <i>DYT1</i> expression requires self regulation and the functions of <i>SPL</i> and <i>EMS</i>	82
3.5.3 <i>DYT1</i> directly regulate downstream transcription factor genes and enzyme genes.....	84
3.6 References.....	91

Chapter 4 Analysis of the *Arabidopsis* floral proteome: detection of over 2000 proteins and evidence for posttranslational modifications.....97

4.1 ABSTRACT.....	98
4.2 INTRODUCTION.....	99
4.3 Materials and methods.....	102
4.3.1 Plant materials and growth conditions.....	102
4.3.2 Protein isolation and separation by 2-D electrophoresis	102
4.3.3 Identification of proteins by MALDI-TOF MS and/or Q-TOF MS.....	102
4.3.4 Multi-dimensional protein identification (MudPIT)	103
4.3.5 Go annotation and pathway analysis	104
4.3.6 Phylogenetic Analysis.....	104
4.4 RESULTS.....	104

4.4.1 Identification of floral proteins by 2-DGE/MS and MudPIT.....	106
4.4.2 Multiple sampling and bioinformatic analyses increase protein detection.....	106
4.4.3 Comparison between protein and transcript levels.....	107
4.4.4 Protein modifications were suggested by 2-D image and MudPIT.....	108
4.4.5 Putative functions of the detected floral proteins	108
4.4.6 Phylogenetic analysis of SAM synthesis pathway components.....	110
4.5 DISCUSSION.....	112
4.5.1 A combination of proteomics methods and analyses allows increased detection of floral proteins.	112
4.5.2 Evidence for post-translational modifications in the <i>Arabidopsis</i> floral proteome.....	114
4.6 References.....	125

List of figures

Figure 1.1 The stages and genes of early anther development.....	25
Figure 2.1 Scatter plots to evaluate the reproducibility of microarray replicates.....	51
Figure 2.2 the overlap between the group of genes that are up or down regulated in <i>dyt1</i> or <i>ams</i>	52
Figure 2.3 RT-PCR and Real time PCR analysis of the genes down regulated in <i>dyt1-1</i> .	53
Figure 2.4: Cellular component analysis of the genes down regulated in <i>dyt1-1</i>	54
Figure 2.5: Cellular component analysis of the genes down regulated in <i>ams</i>	55
Figure 2.6 Molecular function analysis of the genes down regulated in <i>dyt1-1</i>	56
Figure 2.7 Molecular function analysis of the genes down regulated in <i>ams</i>	57
Figure 2.8: A MAPMAN visualization of the genes down regulated in <i>dyt1-1</i>	58
Figure 2.9: A MAPMAN visualization of the genes down regulated in <i>ams</i>	59
Figure 3.1: Expression and purification of the 6×His-DYT1 recombinant protein	85
Figure 3.2 Identification of DYT1 binding sites by Selex procedure.	86
Figure 3.3: DYT1 binds to the E-box of the <i>MSI</i> promoter.	87
Figure 3.4: SPL binds to a 6bp motif specifically.....	88
Figure 3.5: DYT1 binds to the promoters of downstream genes as revealed by ChIP.....	89
Figure 3.6: DYT1 interacts with 3bHLHs and itself in yeast two hybrid assay.....	91
Figure 4.1 A 2-D gel image for identification of 216 proteins from <i>Arabidopsis</i> floral tissues.....	119
Figure 4.2 Overlap between the lists of protein identifications by 2-D/MS and MudPITs and the comparison between two MudPIT experiments.....	120
Figure 4.3 The distribution of proteins identified by MudPIT at different RNA ranges.	121
Figure 4.4 Comparison between the RNA and protein abundances.	122
Figure 4.5 GO categories of the proteins identified by 2-D MS and MudPIT.....	123

Figure 4.6 Phylogenetic trees of SAM synthetase	124
---	-----

List of tables

Table 2.1 The list of transcription factors down regulated in the <i>dyl1-1</i> and <i>ams</i> mutant anther revealed by ATH1 microarray analysis.....	60
Table 2.2 The list of 18 transcription factors down regulated in the <i>dyl1-1</i> mutant anther	61
Table 2.3 The list of transcription factors down regulated in the <i>ams</i> mutant anther	62
Table 2.4 The known genes that are down regulated in <i>dyl1-1</i>	63
Table 2.5 The acyl lipid metabolism genes that are down regulated in <i>dyl1-1</i>	64
Table 2.6 The genes for cell wall modification/degradation down regulated in <i>dyl1-1</i>	65
Table 2.7 The genes involved in secondary metabolism are down regulated in <i>dyl1-1</i>	66
Table 2.8 The genes encoding pollen wall or coat proteins are down regulated in <i>dyl1-1</i>	67
Table 2.9 The transporter genes that are down regulated in <i>dyl1-1</i>	68
Table 4.1 Proteins with multiple spots on 2-D and modifications detected by MudPIT.....	116
Table 4.2 Number of peptide modifications detected by MudPIT with 99% confidence.....	117
Table 4.3 Important categories of proteins identified by MudPIT.....	118

ACKNOWLEDGEMENTS

This thesis project was carried out under the guidance of Dr. Hong Ma, distinguished professor in Biology. I would like to express my deepest gratitude to him for his support and advice. His intelligence and knowledge stimulated me to think deeper and broader, and his encouragement instilled in me the strength to conquer difficulties. I will never forget the time and effort he spent to improve my writing. He even asked his talented son, Jason Ma at Stanford University, to write scripts for my proteomics analysis. He also educated us what an excellent scientist and a great person should be and I have benefited much from it.

I would like also thank my thesis committee members, Drs. Teh-hui Kao, Paula McSteen and Richard J. Cyr for their suggestions on my projects and critical reading of my thesis. Their expertise and knowledge are invaluable to my research. They also advised on my future career and supported me in application for a post doctoral position. It is a great honor to have these scientists on my committee.

The members from the Ma lab are very helpful colleagues in research and good friends in life. I would like to thank Drs. Weimin Ni and Xiaohong Zhang for their guidance when we worked on the same project during the first two years of my study. Dr. Wei Hu accommodated me when I first arrived in U.S.A. and helped me with computer scripts and experiments. Dr. Yiben Peng from Peking University spent months with me to collect anthers for microarray analysis. I also received many constructive suggestions on the DYT1 project from Drs. Wei Zhang and Yujin Sun. I would like also thank Yi Hu for taking care of lab business, Jiong Wang for plant care and Carey Hendrix, John Diller, Xiaofan Zhou, Dihong Lu, Xuan Ma for manuscript revision and data analysis. Other lab members are also helpful in different aspects. I also would like to acknowledge Drs. Andrea Skirpan and Zhixin Zhao for their valuable suggestions.

Dr. Lianchao Li at the Proteomics Facility in Penn State trained me for the proteomic experimental skills. Dr. Bruce Stanley at the Proteomic Facility at Hershey aided in mass spectrometry analysis and provided lot of valuable comments. Drs. Craig Paul, Dr. Haikun Zhang and Dr. Naomi Altman helped with microarray hybridization and statistical analysis. Dr. Greg Ning contributed to the TEM analysis. I really appreciate their expertise and willingness to contribute.

I would like to acknowledge all the graduate staff assistants in Biology Department, especially Kathryn McClintock, for their help during my study. At the Penn State, I enjoyed very much the inspiring talks and communications on the seminar series and work shops etc. I would like to thank the organizers of these activities.

Finally, I would like to sincerely express my gratitude to my family. Their endless love is the stimulus for me to pursue excellence and beauty in life.

Chapter 1

Introduction

1.1 Significance and objectives of studying *Arabidopsis* reproductive development

Flowering plants are important plant species that have been subjected to extensive studies. They constitute one major source of energy and carbohydrate for the ecosystem on earth because they have dominating populations and the function of photosynthesis. They also benefit human as the source of food, energy, and decorating materials for landscape etc. One main reason for flowering plants to flourish on earth is that they developed flower organs and fruits during evolution. These reproductive organs can increase the success rate of sexual reproduction *via* protecting reproductive tissues, facilitating pollination, and improving seed dispersal. Sexual reproduction is an important mechanism of propagation because through this process not only the genetic materials are distributed to the next generation but also some individual plants which are more adaptive to the environment are generated through recombination. Because reproductive development is essential for the success of sexual reproduction, plant biologists have been studying the mechanisms of reproductive development of flowering plants for many decades. The knowledge obtained from the studies has substantially contributed to the genetic engineering of important economic crops.

In the previous decades, *Arabidopsis* was recognized as a model system for the study of sexual reproduction in flowering plants. It belongs to the Brassicaceae family and has many favorable features for genetic studies such as small stature, short life cycle, small genome size, and complete genome sequence information etc. More importantly, many genetic mechanisms were suggested to be conserved between *Arabidopsis* and other flowering plant species. Therefore the discoveries made from *Arabidopsis* are also very valuable for the studies of similar processes in other plant species. Recently, genomic and proteomic approaches became more powerful than before and further aided to discover new molecular mechanisms underlying sexual reproduction of *Arabidopsis*. Because of these advantages, I choose *Arabidopsis* as a model system in my study. Previous studies suggested that transcriptional and post transcriptional controls of cell differentiation are important genetic mechanisms in *Arabidopsis* reproductive development. Therefore, I applied multiple approaches including genetics, expression profiling, and proteomics in

order to make new insights into transcriptional and post transcriptional regulation of plant reproductive development.

1.2 An overview of *Arabidopsis* reproductive development and genetic controls

Reproductive development in *Arabidopsis* includes several stages. It starts from the initiation of inflorescence meristem and floral meristem, through floral organ formation, sporogenesis, gametogenesis, and finally ends up with fertilization and embryo development. The studies of these developmental processes have suggested that both transcriptional and post transcriptional controls are essential.

The initiation of *Arabidopsis* reproductive development is controlled by both environmental and endogenous factors. During vegetative development, the shoot apical meristem (SAM) gives rise to axillary meristems, which will differentiate into branches with leaves. The initiation of the inflorescence meristem (IM) is inhibited by the function of genes such as *TERMINAL FLOWER1 (TFL1)* and *EMBRYONIC FLOWER1 (EMF1)* (Ahn *et al.*, 2006; Aubert *et al.*, 2001; Bradley *et al.*, 1997; Calonje *et al.*, 2008; Chen *et al.*, 1997; Kobayashi *et al.*, 1999; Liljegren *et al.*, 1999; Moon *et al.*, 2003; Shannon *et al.*, 1991; Shannon *et al.*, 1993; Simon *et al.*, 1996; Yang *et al.*, 1995). Under proper conditions including long day photoperiod, a phase of cold treatment (vernalization), and appropriate temperature, the inhibition by genes such as *TFL1* and *EMF1* is lifted and the SAM acquires the competence to flower and changes to the IM. Extensive genetic studies have been performed to understand the control of this transition. Photoperiod, vernalization, and GA pathways are three major pathways that promote the floral transition (Komeda, 2004). These pathways are integrated by genes such as *FLOWERING LOCUS C (FLC)*, *FLOWERING LOCUS T (FT)*, and *SUPPRESSOR OF OVEREXPRESSION OF CO1(SOC1)* (Bastow *et al.*, 2004; He *et al.*, 2003; Helliwell *et al.*, 2006; Hepworth *et al.*, 2002; Levy *et al.*, 2002; Michaels *et al.*, 1999). Recently, the FT protein was suggested to be a mobile signal that promotes floral initiation (Corbesier *et al.*, 2007; Helliwell *et al.*, 2006; Hepworth *et al.*, 2002; Moon *et al.*, 2003; Samach *et al.*, 2000; Searle *et al.*, 2006; Tamaki *et al.*, 2007). FT is expressed in the leaf and transported to the SAM to promote the IM identity, suggesting that it is a candidate for

the hypothesized “florigen” (Corbesier *et al.*, 2007). *Arabidopsis* also has an autonomous pathway that promotes flowering only after a period of time (Komeda, 2004). These studies suggest that the initiation of reproductive development requires the responses of gene functions to environmental factors.

The cells within the inflorescence meristem have two roles: the replenishment of the cells at the central zone to keep the size of the meristem and the depletion of the cells at the peripheral zone *via* differentiating into floral meristems (FM). *WUSCHEL* (*WUS*) and *CLAVATA1/2/3* (*CLV1/2/3*) promote and inhibit the meristem identity, respectively. The balance between these two antagonizing functions regulates the stem cell population and maintains the indeterminate growth of the meristem (Brand *et al.*, 2000). In the peripheral region of the IM, *LEAFY* (*LFY*) acts to establish the FM identity because the *lfy* mutants have only leaf-like structures at the presumed position of floral organs (Kobayashi *et al.*, 1999; Liljegren *et al.*, 1999; Shannon *et al.*, 1993; Simon *et al.*, 1996). *APETALA1* (*API*) is a gene also involved in FM formation. *API* is activated at the transcriptional level by *LFY* (Liljegren *et al.*, 1999). In turn, *API* also positively regulates the *LFY* expression. Such a positive regulatory loop could reinforce the commitment to form the FM.

In different regions of FM, the identities of different floral organs are established. In *Arabidopsis*, there are four whorls of floral organs: sepals, petals, stamens and the pistil. Forward genetic screens have discovered many floral homeotic genes that regulate the identities of floral organs. Based on current understanding about the floral homeotic genes, the ABC model proposes that floral organ identities are determined by the combination of three functions that are mainly carried out by MADS box genes. The A function genes are *API* and *APETALA2/AP2* (Drews *et al.*, 1991; Gustafson-Brown *et al.*, 1994; Mizukami *et al.*, 1992); The B function genes are *APETALA3/AP3* and *PISTILLATA/PI* (Goto *et al.*, 1994; Jack *et al.*, 1994; Kramer *et al.*, 1999; Lamb *et al.*, 2002; McGonigle *et al.*, 1996; Ng *et al.*, 2001); The C function gene is *AGAMOUS/AG* (Drews *et al.*, 1991; Mizukami *et al.*, 1992). In the ABC model, the A function specifies the sepal, the A+B functions determine the petal identity, the B+C functions control stamen formation, and the C function alone regulates the pistil identity. Among these

functions, the A function and the C function play antagonistic roles (Drews *et al.*, 1991). One of the A function genes, *AP2*, encodes an AP2 domain transcription factor which is expressed in all four whorls (Drews *et al.*, 1991). It is required to repress the *AG* function in the outer two whorls. The balance between *AP2* and *AG* is regulated by microRNA172 (Aukerman *et al.*, 2003; Chen, 2004). The ABC functions are activated by floral meristem identity genes such as *LFY* and *API* (Lamb *et al.*, 2002; Ng *et al.*, 2001; Simon *et al.*, 1996).

Besides the ABC functions, the E function is also required to establish floral organ identities. The E function genes, *SEPALLATA 1/2/3* (*SEP1/2/3*, also MADS box genes), are needed to establish a ground floral state, which is suggested by the findings that the *sep1/2/3* triple mutant has leaf-like structures in the position of floral organs (Pelaz *et al.*, 2000). This function is named as the E function. The Quartet model proposes that a tetramer complex of MADS box proteins of the ABC and E functions can specify one floral organ identity. This is supported by the findings that ectopic expression of the *API/AP3/PI/SEP3* genes or the *AP3/PI/AG/SEP3* genes changes leaves into petal-like or stamen-like structures, respectively (Honma *et al.*, 2001). These studies suggest that transcriptional control is a major mechanism in the regulation of floral organ identities. More studies are needed to identify the downstream genes of the floral organ identity genes in order to understand how the tissues are specified within each floral organ.

The tissues within the floral organs are formed after several rounds of asymmetric cell divisions. For the anther and the pistil that produce pollen grains and egg cells respectively, both the sporophytic tissues and the gametophytic tissues need to be properly formed. The gametophytic tissues undergo sporogenesis and gametogenesis, which is supported by the surrounding sporophytic tissues. Genetic studies have uncovered tens of genes that are required for the formation and functions of the tissues in reproductive organs. To exemplify how reproductive tissues are formed under genetic controls, I will introduce our current understanding about anther development and the underlying genetic network in *Arabidopsis*.

1.3 Rationales to study anther development

Many studies have been carried out to understand anther development because the anther is the male reproductive organ, and its normal development is essential for the success of male reproduction in flowering plants that requires pollen formation. To produce pollen, not only the pollen mother cells (PMCs) but also the sporophytic cells surrounding PMCs should be properly formed. The sporophytic cells in the endothecium, the middle layer and the tapetum have essential roles supporting pollen formation (Ma, 2005). However, the regulation of the cell division and differentiation events that give rise to these cells has not been well understood.

Furthermore, the anther is an excellent genetic system to study mechanisms of plant development (Ma, 2005). First of all, the anther has a simple structure. It consists of only five layers of different tissues in one lobe, which makes it accessible to many approaches such as cytological imaging and histochemical staining, etc. This feature and the advantages of using *Arabidopsis* as a model system make the anther a powerful genetic system. Furthermore, anther development exhibits complexity of cell division and differentiation events so that it provides plenty of chances to explore the genetic mechanisms. For example, premature or delayed programmed cell death of the tapetum causes the collapse of meiocytes and the failure to develop into viable pollen grains (Kapoor *et al.*, 2002). This suggests that cell differentiation events in the anther are coordinated and synchronized so that the anther could be a good system for studying cell-cell communication. Because of the simplicity of the structure and the complexity of developmental mechanisms, I chose the anther as a major subject of my thesis research.

1.4 The structure and functions of *Arabidopsis* anther

The anther is part of the stamen. A filament connects the anther to the base of the flower and transports nutrients to the anther. A typical *Arabidopsis* anther has four lobes connected by the vascular tissue. For each lobe, there are five cell layers: the epidermis, the endothecium, the middle layer, the tapetum and the meiocytes (Goldberg *et al.*, 1993;

Ma, 2005). These cells have distinct functions and collaborate to support the formation of pollen grains.

The meiocytes are located inside each locule. At stage 5 of anther development, the large meiocytes initiate meiosis. After DNA duplication, the homologues pair, recombine, and segregate in meiosis I. Right after the separation of homologues in meiosis I, meiosis II is initiated without cell plate formed to separate the two clusters of chromosomes, and the sister chromatids are separated to form four clusters of chromosomes. Then, the cell plates are formed to separate the four chromosomal clusters and encapsulate them in a callose and pectin enriched cell wall. This structure is called tetrad. Finally, the microspores are released from the tetrad and develop into pollen grains by dividing into two sperms and one vegetative cell. Pollen coat is formed around the pollen grain, which is mainly contributed by the tapetum. However, the microspore also plays a role in the formation of the pollen wall. For example, in rice, UDP-glucose pyrophosphorylase1 is expressed in pollen and is required for callose deposition (Chen *et al.*, 2007). Besides this, meiocytes also provide signals to regulate tapetum functions. It is thought that cell-cell signaling is critical to coordinate the tapetum activities with events in meiocytes. However, besides TPD1, which might be a microspore-generated peptide signal that directs tapetum formation, little is known about if other signals also exist.

The four layers of sporophytic cells are not directly involved in the formation of microspores; however, their functions are not dispensable. The epidermis protects all ground tissues from the environment. The endothecium forms a relatively thin layer which generates the mechanical force required for anther dehiscence. After pollen matures in the locule, the cell wall of the endothecium becomes lignified in preparation for pollen release. The middle layer is even thinner than the endothecium, and its function is not clear. The tapetal cells directly contact meiocytes, and provide nutrients and signaling molecules to meiocytes. They actively synthesize lipid, protein, and sets of secondary metabolites, which might explain the presence of abundant cytoplasm in the tapetal cells as revealed on a thin section or TEM image. The tapetal cells undergo endoreduplication and endomitosis so that the genome reaches $4n$ or $6n$ to fulfill the need

for active transcription. At stage 8 of anther development, the tapetal cells have degenerated *via* programmed cell death and released all contents into the locule. This event is synchronized with meiocyte development and occurs just when microspores are ready to be released from the tetrads and begin to form mature pollen grains. The materials from the tapetal cells are the source of pollen wall components. Mutants with defective tapetal cells are usually male sterile.

1.5 The cell division and differentiation in early anther development are regulated by transcriptional cascades and cell-cell signaling

The early anther development starts with the establishment of the anther primordium identity in the third whorl of the flower, which is controlled by the combinatorial function of the B function genes (*AP3* and *PI*) and the C function gene (*AG*). It is largely unknown how these MADS box genes control the anther identity. A key transcription factor gene, *SPOROCTELESS* (*SPL*), was identified as the direct target of *AG* and as being required for early cell differentiation in the anther (Balasubramanian *et al.*, 2000; Ito *et al.*, 2004; Sieber *et al.*, 2004; Yang *et al.*, 1999). This suggests that a transcriptional cascade involving the floral organ identity genes and downstream transcription factor genes regulates early anther cell differentiation. Further identification of the genes downstream of the B+C functions will elucidate how the early anther differentiation is regulated at the transcription level.

At stage 1, the anther primordia have three layers of undifferentiated cells. The outmost layer of cells (L1) differentiate into the epidermis; the innermost layer (L3) consists of the founder cells for vasculatures. The L2 cells are the progenitor cells of meiocytes, the tapetum, the middle layer and the endothecium. At stage 2, the L2 cells enlarge and form archesporal cells after asymmetric cell divisions, and as a result the anther exhibits an oval shape. At stage 3, the archesporal cells undergo periclinal cell divisions to form the primary parietal cells close to the epidermis and the primary sporogenous cells inside. Forward genetic studies have revealed several loci required for the cell differentiation during these early stages. The *SPL* gene is a key factor in the differentiation of

archesporal cells (Ito *et al.*, 2004; Yang *et al.*, 1999). Multiple alleles of *spl* mutants are both male and female sterile. The anthers from the *spl* mutants are shrunk and devoid of pollen. In the *spl* mutant, anther development is blocked at the stage 2, and the cells inside of the epidermis are undifferentiated. SPL is a novel transcription factor as suggested by its sequence and its localization in the nucleus. Its mRNA can be detected in stage 1 anthers and is highly expressed in the tapetum and meiocytes at stage 5. It is possible that *SPL* also has important functions in the tapetum and meiocyte at later stages. However, an early blockage in the anther development of the *spl* mutant makes such a potential function difficult to study. It was shown that the floral homeotic protein AG can bind to the CArG box in the 3' un-translated region (3'UTR) of *SPL* and promote the *SPL* expression in the tapetum and meiocytes (Ito *et al.*, 2004; Yang *et al.*, 1999). Therefore, regulation of anther cell differentiation by AG is in part mediated by transcription factors such as SPL. A transcriptional profiling analysis of the *spl* mutants during anther stages 4-7 suggests that *SPL* both activates and represses a set of genes, including the genes preferentially expressed in the tapetum and meiocytes (Wijeratne *et al.*, 2007). However, it is not clear what the transcriptional cascade is downstream of *SPL* during the early anther stages.

Cell-cell signaling also plays an important role in defining the cell types in early anther. Through signaling, a cluster of cells acquire their cell fate according to the cues from the neighboring cells. At the same time, they also release signals to affect the identities of neighboring cells. Recently, BARELY ANY MERISTEM 1/2 (BAM1/2), a pair of CLAVATA1-related leucine rich repeat receptor protein kinases (LRR-RPKs), were found to be important not only for the patterning of the shoot apical meristem but also for early anther cell differentiation (Hord *et al.*, 2006). In the *bam1/2* double mutant, the anther lacks the endothecium, the middle layer and the tapetum. Instead, the cells inside of the epidermis are not properly differentiated and acquire a partial identity of pollen mother cells (PMCs). Because the sporophytic tissues are missing, these PMC-like cells degenerate early. LRR-RPKs belong to a super family of receptor like kinases, which are localized on the plasma membrane and required for the transduction of different signals into the cell. The involvement of LRR-RPKs in early anther development suggests that a

signaling process is required for the specification of the cells in the early anther. A model based on expression and phenotypic analyses proposes that *BAM1/2* are activated by *SPL*, but in turn they repress the *SPL* expression. This negative feedback loop is analog to that of *CLV1/2/3* and *WUS* in the shoot apical meristem. However, the mechanisms of this signaling pathway, especially the putative ligand for the BAM1/2 receptors, remain to be elucidated.

At stage 4, the primary parietal cells differentiate into two layers of secondary parietal cells. The primary sporogenous cells initiate the S-phase DNA synthesis in preparation for meiosis. At stage 5, the outer secondary parietal cells divide and differentiate into the endothecium and the middle layer, whereas the inner secondary parietal cells differentiate into the tapetum. At this stage, all types of cells in the anther are formed and patterned as concentric layers.

The microsporocytes initiate meiosis at stage 5. As for how their identity is established, very limited understanding is obtained, and a mutant that lacks only the sporogenous cells has not been found in *Arabidopsis*. It seems that the neighboring tapetal cells have less effects on the microsporocyte identity because the microsporocytes can form in the absence of the tapetum in the *ems1* mutants (Yang *et al.*, 2005; Yang *et al.*, 2003; Zhao *et al.*, 2002; Zhao *et al.*, 2008). *SPL* may have positive roles in the formation of microsporocytes. However, it is unclear whether the formation of microsporocytes directly requires *SPL* because the anther development is blocked at very early stage in the *spl* mutant. Also, little is known about the molecular switch of the entry into meiosis. In the maize *ameiotic 1* (*am1*) mutants, meiosis is replaced with mitosis in both micro- and mega- sporocytes (Golubovskaya *et al.*, 1997; Golubovskaya *et al.*, 1993). The *am1* gene might be important for the entry of meiosis in maize, but it is still unclear whether there are similar genes in *Arabidopsis*. However, one gene seems to play a role in the switch from mitosis to meiosis in the megasporocyte in *Arabidopsis*. In the mutant defective in the *SWITCH1* (*SWI1*) gene, additional rounds of mitosis proceed meiosis (Mercier *et al.*, 2003; Mercier *et al.*, 2001; Ravi *et al.*, 2008). This suggests that *SWI1* may repress

mitosis before meiosis, and synchronize the entry into meiosis in megasporocytes. The direct positive factors for the entry into meiosis are still missing.

In *Arabidopsis*, meiosis includes two rounds of cell divisions: meiosis I and II. Extensive literature is available about the genetic controls of meiosis in *Arabidopsis*. Here I briefly summarize the events during and after meiosis. Meiosis I has a long prophase, which can be further divided into five phases: leptotene, zygotene, pachytene, diatene, and diakinesis (Hamant *et al.*, 2006; Ma, 2005). During meiosis I, the homolog pair, synapsis, recombine and separate. Little is known about how homologs recognize each other and pair. Evidence suggests that telomere cluster, centromere and double strand break repair may be required for homolog pairing. In maize, the *phs1* locus prevents non-homolog pairing (Pawlowski *et al.*, 2004). However, the mechanism is not clear. The central event in meiosis is recombination. Double Strand Break Repair (DSBR) has been recognized as the most important recombination pathway (Ma, 2005). In the DSBR model, SPO11, a topoisomerase IV, and its homologs generate double strand breaks. The AtDMC1 and AtRAD51 recombinases (and homologous recombinases) promote the recombination between homologous chromosomes (Bleuyard *et al.*, 2004; De Muyt *et al.*, 2007; Grelon *et al.*, 2001; Hartung *et al.*, 2007; Klimyuk *et al.*, 1997; Li *et al.*, 2004; Li *et al.*, 2005; Sanchez-Moran *et al.*, 2007; Stacey *et al.*, 2006). During recombination, the mismatch repair proteins (MSH4/5) are required to stabilize the recombination intermediates, Holiday Junctions (HJs), and DNA helicases, such as ROCK-N-ROLLERS (RCK), might promote the branch immigration (Hamant *et al.*, 2006; Higgins *et al.*, 2004)(Chen *et al.*, 2005). Finally, the HJs are resolved by presumed resolvases. In plant, such enzyme activity has not been found. However, PARTING DANCERS (PTD), a homolog of yeast XRCC1, might be involved in this process (Wijeratne *et al.*, 2006). It was found that the proper recombination between homologs is also required for normal alignment of homologs at the equator plate at metaphase I. Then upon the action of spindle and motor proteins such as kinesins, homologs are separated to the opposite poles of the meiocytes (Chen *et al.*, 2002; Marcus *et al.*, 2002; Quan *et al.*, 2008). Without the formation of the cell plate to separate the two clusters of chromosomes after anaphase I, sister chromatids start to separate in meiosis II. Finally, four clusters of chromosomes are formed, and they

are encapsulated by a cell wall rich in callose and pectin. This structure is called tetrad. Then the tetrad cell wall is dissolved to release four microspores (Ma, 2005). Each microspore undergoes two round of mitosis to form a 3-cell male gamete: two sperm cells and one vegetative cell. These three cells are packed into a chemical inert wall to form a mature pollen grain, which is released to fertilize the female gamete in the ovary after anther dehiscence (Ma, 2005; McCormick, 1991).

1.6 Genetic controls of the tapetum differentiation

The tapetum is differentiated from the innermost secondary parietal cells during stages 4-5. During these stages, the tapetal cells develop a set of specific functions to support the development of meiocytes and pollen grains. Several genetic loci have been found important to tapetum differentiation and functions. The major molecular events underlying tapetum development and functions are: (1) the cell-cell signaling mediated by TPD1/EMS1 and transcriptional regulation that are required to establish tapetum identity; (2) expression of early transcription factors including DYT1 and MYB33/65, and regulation of intermediate transcription factors including AMS and MS1, etc.; (3) regulation of structural and enzyme proteins.

The fate of tapetal cells is regulated by a signaling process mediated by receptor like protein kinases. Two genes, *EXCESS MICROSPOROCYTES 1 (EMS1/EXS)* and *TAPETUM DETERMINANT 1 (TPD1)*, encode a LRR-RPK and a short peptide respectively (Yang *et al.*, 2005; Yang *et al.*, 2003; Zhao *et al.*, 2002; Zhao *et al.*, 2008). Mutation in either gene causes the missing of the tapetal cell layer and the formation of the excess meiocytes in the presumed position of the tapetum. In the *ems1* mutant, meiosis can proceed without detectable defects; however, the cytokinesis at meiosis II is not completed. This suggests that the sporophytic cells play an important role in supporting the cytokinesis after meiosis. At stages 3-4, *EMS1* and *TPD1* show overlapping expression patterns in the precursor cells of the meiocytes and the tapetum (the secondary parietal cells). Then at stage 5, they establish complimentary expression domains: *EMS1* is highly expressed in the tapetum and *TPD1* is mainly expressed in the

meiocytes. The anther phenotype of the *ems1/tpd1* double mutant resembles that of each single mutant, which suggests they function in the same pathway or in the same complex.

Recent evidence suggests that TPD1 is a peptide that is produced and released from PMCs and directs tapetum differentiation by binding to EMS1 as a ligand (Yang *et al.*, 2005; Yang *et al.*, 2003; Zhao *et al.*, 2002; Zhao *et al.*, 2008). This ligand-LRR RPK signaling complex might be analogous to the complexes involved in the control of the shoot apical meristem size (CLV3/CLV1), root epidermis development (SCRAMBLED [SCM]), flower dehiscence (INFLORESCENCE DEFICIENT IN ABSCISSION1 [IDA1]/HAESA) and stomata patterning (EPIDERMAL PATTERNING FACTOR 1 [EPF1]/TOO MANY MOUTHS[TMM]) (Fiers *et al.*, 2005; Hara *et al.*, 2007; Ito *et al.*, 2006; Jinn *et al.*, 2000; Kwak *et al.*, 2007; Kwak *et al.*, 2005; Nadeau *et al.*, 2002; Ogawa *et al.*, 2008; Shpak *et al.*, 2005). It is interesting to know what is required to establish the complementary expression patterns for a ligand and its receptor, and how EMS1/TPD1 controls tapetum development.

Recently, it was suggested that the EMS1 receptor interacts with other receptor like kinases. *SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1/2* (*SERK1* and *SERK2*) encode two redundant receptor-like kinases, which might be the co-receptors of EMS1 because the *serk1/2* double mutant exhibits similar phenotype as that of *ems1* (Albrecht *et al.*, 2005; Colcombet *et al.*, 2005; Shah *et al.*, 2002). The EMS1 receptor complex might be analogous to the brassinosteroid signaling complex BRASSINOSTEROID INSENSITIVE 1/BRI1-ASSOCIATED RECEPTOR KINASE1 (BRI/BAK1) (He *et al.*, 2007; Heese *et al.*, 2007; Li *et al.*, 2002; Nam *et al.*, 2002). Recently, a transcriptome analysis of the *ems1* anther revealed that many transcription factor genes exhibit differential expressions in *ems1*. This suggests that one important function of EMS1/TPD1 is to regulate transcriptions in the tapetum. However, it is not clear what molecular events are that links the signal perception by EMS1 on the cell surface to the transcriptome changes in the nucleus.

At stage 5, the tapetal cells become specialized in structure and functions. Several key genes have been found to be important for the specification of tapetum functions. Most of these genes were discovered in genetic screens for male sterile mutants. Among them, there are several genes encoding transcription factors: *DYSFUNCTIONAL TAPETUM1* (*DYT1*), *MYB33/65*, *ABORTED MICROSPORES* (*AMS*), and *MALE STERILE 1* (*MS1*). Another category of genes encode enzymes and structural proteins that are required for microspore and pollen maturation. This suggests that a transcriptional cascade exists and regulates the expression of the genes required for tapetum functions.

DYT1 is a gene required for normal tapetum functions (Zhang *et al.*, 2006). In our lab, one Ds line that can't produce any pollen was analyzed. The results of the cross pollinations between wild type plants and the *dyt1-1* mutant suggest that the *dyt1-1* mutant is only male sterile. Dramatic phenotypes were found on the thin sections of the *dyt1-1* anther: thin cytoplasm and vacuolation of the tapetum cells and early degeneration of the meiocyte. Although meiosis can be initiated and completed, the cytokinesis at meiosis II is defective, suggesting a role for the tapetum in meiosis. By map-based cloning, the *DYT1* gene was isolated, which encodes a bHLH transcription factor. In the *dyt1-1* mutant, a transposon is inserted at the proximal region of the *DYT1* promoter so that the expression of *DYT1* is reduced 3-4 folds. Two independent experiments were performed to further support the finding that the *dyt1-1* phenotypes are caused by a mutation in *DYT1*: a genomic fragment of *DYT1* can rescue the mutant; a transgene harboring the *DYT1* RNAi construct driven by the *DYT1* native promoter can cause the *dyt1-1* phenotypes. The DYT1 protein has 207 amino acid residues, and the bHLH domain is predicted to span residues 29-78. An expression analysis suggests that the *DYT1* transcripts are detected at low level in the stage 4 anthers, at a dramatically high level in the tapetum from stage 5 to stage 6, and then at a reduced level after stage 7. Its expression is reduced in the *spl* and *ems1* mutants (Wijeratne *et al.*, 2007; Zhang *et al.*, 2006). Therefore, both *SPL* and *EMS1* are upstream of *DYT1*. Some genes downstream of *DYT1* are also found, which include *AMS* and *MS1*. These results suggest that *DYT1* is activated by early genes such as *SPL* and *EMS1* and then regulate a set of downstream

genes in the tapetum. I believe that a careful study of the DYT1 function will lead us to the discoveries of novel genes and functions of the tapetum.

MYB33 and *MYB65* are a pair of GA-MYB-like genes that have functions at similar stages to that of *DYT1*. Because the *MYB33/65* expression is reduced in the *spl* and *ems1* mutants but not in the *dyl1-1* mutant, it is proposed that *MYB33/65* are downstream of *EMS1* and *SPL* and function at a parallel level with *DYT1* (Millar *et al.*, 2005). Although the single mutant of each gene does not exhibit obvious phenotypes, the *myb33/65* double mutant is male sterile. At stage 5, the tapetal cells from the double mutant are vacuolated which is similar to that of *dyl1-1*. However, the tapetal cells of the *myb33/65* double mutant are squeezed by neighboring tapetal cells so that they become narrow and protrude toward meiocyte, whereas in the *dyl1-1* anther, the tapetal cells are enlarged and swollen. Furthermore, the *myb33/myb65* double mutant can restore its fertility at low temperature such as 16°C, whereas the *dyl1-1* mutant can not restore its fertility (unpublished data) (Millar *et al.*, 2005). This suggests that at least two genetic pathways (the DYT1 and MYB33/65 pathways) regulate the tapetum functions from stage 5, both of which are downstream of *SPL* and *EMS1*. Interestingly, this pair of *MYB* genes are regulated by microRNAs which target their coding sequences (Millar *et al.*, 2005). However, our knowledge about the *MYB33/65* downstream genes is very limited, so is our understanding of tapetum functions regulated by these two *MYB* genes.

Other known transcription factors that are important for late tapetum functions are downstream of *DYT1*. These include AMS, MS1, and MYB103. In a screen of T-DNA lines for mutants with defects in silique elongation, a recessive sporophytic mutant was isolated and named as *aborted microspores (ams)* (Sorensen *et al.*, 2003; Thorstensen *et al.*, 2008). In the anther of the *ams* mutant, no mature pollen grains are formed. The microspores are degenerated after they are released from the tetrad. This might be a consequence of the premature degeneration of the tapetum. The *AMS* gene encodes a bHLH transcription factor. It is preferentially expressed in the tapetum and is positively regulated by *DYT1*. Recently, the ortholog of *AMS* in rice, *TAPETUM DEGENERATION RETARDATION (TDR)*, was studied (Li *et al.*, 2006). The *tdr* mutant shows delayed

program cell death (PCD) in the tapetum as suggested by TUNEL assays and the occurrence of premature microspore degeneration. *TDR* directly regulates two genes involved in PCD, *OsCP1* and *Osc6*, which encode a cysteine protease and a protease inhibitor respectively. This suggests that *TDR* might regulate PCD by directly regulating the expression of PCD related genes in rice (Li *et al.*, 2006). *AMS* might regulate similar processes in *Arabidopsis*. However, a genome wide search for the *TDR* or *AMS* downstream genes has not been done. Therefore, the importance of the pathways regulated by *AMS* or *TDR* remains to be recognized. Recently, it was found that *AMS* interacts with *ASHR3*, a protein with a SET domain and a PHD domain (Thorstensen *et al.*, 2008). The SET domain has been suggested to be important to gene repression. Therefore, it is possible that *AMS* can repress gene expression by forming a transcriptional complex with other proteins such as *ASHR3*. *ASHR3* is expressed in the anther and its over-expression can cause the degeneration of the anther, which suggests that the *AMS*-*ASHR3* interaction might have important *in vivo* functions (Thorstensen *et al.*, 2008).

MS1 is another transcription factor that functions at a slightly later stage than that of *AMS* (Ito *et al.*, 2007; Ito *et al.*, 2002; Vizcay-Barrena *et al.*, 2006; Wilson *et al.*, 2001; Yang *et al.*, 2007). It is expressed in the tapetum during late tetrad stages and microspore release. Phenotypic analyses of different *ms1* mutants revealed that it controls tapetum development and the pollen exine formation. Abnormal PCD was also observed in the *ms1* tapetum cells (Vizcay-Barrena *et al.*, 2006). Due to the pollen wall defects, the pollen grains from the *ms1* anther have abnormal shape and are not viable. *MS1* encodes a PHD domain transcription factor. It is suggested that *MS1* functions as a transcriptional activator because a *MS1* transgene fused with *SRDX*, a transcriptional repressor domain derived from *SUPERMAN* (*SUP*) and *CUP SHAPEDS COTYLEDEN1* (*CUC1*), can cause the *ms1* phenotypes (Ito *et al.*, 2007; Yang *et al.*, 2007). By using DEX inducible system and microarray, two labs analyzed the genes regulated by *MS1* and discovered more than 200 genes that are downstream of *MS1*. These genes encode different transcription factors (among which *MYB99* might be one direct target), proteases, and proteins related to pollen wall formation (Ito *et al.*, 2007; Yang *et al.*, 2007). The finding

that proteases are the *MSI* downstream targets is coincident with the observation that PCD happens in the *msl* anther.

Although both *AMS* and *MSI* are important genes downstream of *DYT1*, it is still not clear whether *DYT1* can directly regulate these two genes. Recently, a R2R3 type MYB gene, *TAPETUM DEVELOPMENT and FUNCTIONS1 (TDF1)*, was discovered to be important for tapetum functions. It is downstream of *DYT1* and upstream of both *AMS* and *MYB103* (Zhu *et al.*, 2008). This suggests that *DYT1* might regulate *AMS* indirectly. However, because of the potential dimerization between the bHLH and R2R3 MYB proteins, I think *DYT1* might also directly regulate these downstream genes by forming a complex with *TDF1*. The transcription factors which function further downstream are closer to the end of the transcriptional cascade and may directly activate structural and enzyme proteins. *MYB103* was found downstream of *AMS* and is required for callose dissolution and exine formation of pollen (Zhang *et al.*, 2007). Its close homolog, *MYB99*, is downstream of *MS1*, which may have similar functions (Alves-Ferreira *et al.*, 2007; Ito *et al.*, 2007; Yang *et al.*, 2007). *MYB32* regulates the phenylpropanoid biosynthesis pathways and is also activated by *DYT1* (Preston *et al.*, 2004).

The discoveries of aforementioned genes provide us a solid base to study the genetic network of the tapetum development. So far, it is still a challenging problem to uncover all genetic components in tapetum development and to define the relationships among those components in a network. By searching for the downstream genes of the transcription factors that have been characterized, we will be able to gradually reveal the overall theme for the regulation of tapetum development and functions. To resolve the structure of the transcriptional network in the tapetum, more studies of the protein-DNA interactions are also needed. Chromatin Immuno-precipitation-sequencing (ChIP-seq) will provide valuable information (Lee *et al.*, 2007). Meanwhile, isolation and analyses of new mutants are also needed.

1.7 The tapetum cells are highly active in biosynthesis and secretion of pollen wall components

The tapetum cells are very active which is suggested by the observations that the tapetum cells are enlarged with abundant cytoplasm, and that the genome is 4-6N. The activities of the tapetum cells are required to support microsporogenesis. Downstream of the transcription factors discussed above, there are the genes encoding structural proteins and enzymes that are transcribed in the tapetum (Goldberg *et al.*, 1993). These proteins include the precursors of structural proteins on the pollen coat, the enzymes involved in the biosynthesis and secretion of essential metabolites required for pollen maturation, and the structural proteins of the organelles in the tapetum.

To support these biosynthesis and secretion activities, the tapetal cells become one major sink for nitrogen, sugar and other organic compounds. It was found that several transporters have important functions in the tapetum. For example, AtLHT2 is a proline/asparagine high affinity transporter (Lee *et al.*, 2004). It is needed to transport amino acids to the tapetum. In rice, an anther specific monosaccharide transporter (OsMST8) is repressed by a chilling treatment (Mamun *et al.*, 2006). Such a treatment can cause vacuolation and hypertrophy of the tapetum and accumulation of starch in the endothecium, which might be explained by the defects in the saccharide transporter system involving OsMST8.

Energy is also required for the biosynthesis and secretion activities of the tapetum so that energy metabolisms that involve the functions of mitochondria are active in the tapetum. This is supported by the fact that Cytoplasm Male Sterility (CMS) is usually caused by mitochondria dysfunctions in the tapetum. For example, defects in the Fe-S subunit of the mitochondria complex II, SDH2-1 and SDH2-1, can cause male sterility (Elorza *et al.*, 2004). CMS can also be caused by inhibition of important enzymes in glycolysis such as pyruvate dehydrogenase. It is expected that other components of mitochondria or other enzymes in energy metabolisms are also required for the tapetum functions.

One main function of the tapetal cells is to synthesize different types of compounds with the materials transported from vascular tissues and the energy generated by mitochondria; these compounds are needed for microsporogenesis at different stages. During meiosis, the microsporocytes exhibit dramatic cell wall changes. A cell wall mainly composed of callose and pectin is formed surrounding the microsporocyte during meiosis (Ma, 2005). After meiosis II, this type of cell wall encapsulates four microspores to form the tetrad. It is suggested that the majority of the components of this cell wall are synthesized in the tapetum. The mutants that are defective in the biosynthesis of these cell wall components in the tapetum are usually male sterile. For example, mutations in the callose synthase genes including *GSL5* can cause defective in the tapetum and result in male sterility.

When the microspore matures, the cell wall of the tetrad is dissolved so that four microspores are released. The dissolution of the tetrad wall is also dependent on the enzymes secreted from the tapetum. For example, a β -1,3-glucanase from the tapetum is required to degrade the callose around the tetrad (Bucciaglia *et al.*, 1994; Worrall *et al.*, 1992). In *Arabidopsis* and *Brassica*, an anther specific gene, *A6*, also encodes this type of enzyme and is required for the dissolution of the callose around the tetrad (Hird *et al.*, 1993). A similar anther preferential enzyme, *Tag1*, was found in tobacco (Bucciaglia *et al.*, 1994). Pectin modification and dissolution is also required for the release of microspores. In the *quartet1* mutant, the tetrad wall is not dissolved although each microspore is viable, which is caused by a mutation in the *QUARTET1* (*QRT1*) gene encoding a pectin methylesterase (Francis *et al.*, 2006; Rhee *et al.*, 1998). This suggests changes in pectin ester modification might be required for the dissolution of the pectin structure of the tetrad.

After microspores are released from tetrads, the tapetum cells initiate PCD and degenerate so that the compounds synthesized in the tapetum are released and deposited onto the maturing pollen grains. However, before the degeneration of the tapetum, the majority of compounds for pollen coat formation are not released. An important question is how the compounds are stored and then released.

In most plants, the lipid rich organelles in the tapetum are important compartments for storing the pollen coat materials. These organelles also function as carriers to release and deposit the pollen coat materials. The tapetosomes are tapetum specific organelles that have the functions of both storage and delivery of pollen coat materials. It is assembled on the rough ER as oleosin-coated oil droplets (mainly triacylglycerols, TAGs), and then detached from ER to form isolated organelle (Hernandez-Pinzon *et al.*, 1999; Hsieh *et al.*, 2005; Ting *et al.*, 1998; Wu *et al.*, 1999; Yui *et al.*, 2003). Oleosin is one major protein components on the pollen wall which is originally accumulated in the tapetosome. Upon degradation of the tapetal cells, the oleosin is deposited onto the pollen wall (Hernandez-Pinzon *et al.*, 1999; Hsieh *et al.*, 2005; Roberts *et al.*, 1993; Ross *et al.*, 1996; Ruiter *et al.*, 1997; Ting *et al.*, 1998; Wang *et al.*, 1997; Wu *et al.*, 1999). Oleosin was first discovered from olive as a major pollen allergen and then found from the pollen grains of *Brassica*, *Arabidopsis*, maize and rice. It was suggested that the oleosin functions as an emulsifier for lipid and facilitates the deposition of lipid onto pollen grains. Calreticulin is a calcium binding protein that is also abundant in the tapetosome, however, it is unclear what role it plays (Hsieh *et al.*, 2005; Nardi *et al.*, 2006; Nelson *et al.*, 1997). The tapetosome also functions as a storage compartment for other compounds such as flavonoids (Kapoor *et al.*, 2002; Napoli *et al.*, 1999; Pollak *et al.*, 1993; Preston *et al.*, 2004; van Eldik *et al.*, 1997; Ylstra *et al.*, 1992). Recently, a fluorescence staining of the flavonoids in the tapetum revealed that flavonoids are synthesized in the ER of the tapetum, stored in the tapetosome, and finally deposited onto pollen. (Hsieh *et al.*, 2007). The flavonoids likely protect male gametes from UV light, which is suggested by the fact that the pollen grains from the flavonoids synthesis mutants show reduced survival rate under UV light (Hsieh *et al.*, 2007). Coincident with this, inhibition of the key gene in flavonoids biosynthesis, the chalcone synthase gene (*CHS*), can cause male sterility (Atanassov *et al.*, 1998; Pollak *et al.*, 1993; van der Meer *et al.*, 1992; Yang *et al.*, 2007).

In cereals, Ubisch body is a hallmark organelle for the secretory tapetum. Its biogenesis also occurs on the ER and may have similar functions to that of tapetosomes. RAFTIN is a protein accumulated in Ubisch bodies and can be detected on pollen suggesting that the Ubisch body also has both storage and delivery functions (Wang *et al.*, 2003). The C-

terminal domain of RAFTIN shows low homology with the BURP domain found in *Arabidopsis*. Inhibition of the RAFTIN expression by antisense RNA causes the collapse of pollen which suggests that the Ubisch body is important for tapetum functions and pollen formation. Elaioplast is another type of lipid rich organelle in the tapetum (Hernandez-Pinzon *et al.*, 1999; Hsieh *et al.*, 2007; Ting *et al.*, 1998; Wu *et al.*, 1999; Yui *et al.*, 2003). Different from the tapetosome, the elaioplast is rich in degradable neutral esters and proteins, which are the components of pollen wall.

Although the organelles like tapetosomes are important, their functions have not been fully appreciated. It is necessary to study their biogenesis, chemical composition and regulations. Recently, it was reported that a type V p-type ATPase encoded by *MALE GAMETOGENESIS IMPAIRED (MIA)* is required to transport the secretion vesicles to the tapetum membrane (Jakobsen *et al.*, 2005). Mutations in the *MIA* gene will disrupt the cation homeostasis in the tapetum and affect the expression of secretion related genes. It is not clear this type V p-type ATPase is also involved in the regulation of the lipid rich organelles.

The effects of the tapetum on pollen exist even after pollen maturation. The tapetum functions are also required for pollen-pistil recognition, pollen rehydration, and pollen tube penetration. For example, the S-locus cysteine rich protein (SCR) is a pollen determinant for self-incompatibility in *Brassica*. This protein is synthesized in the tapetum and can be recognized by S-locus glycoprotein (SLG) and the S-locus receptor kinase (SRK) on the pistil so that the specificity of pollen-pistil recognition is established (Boyes *et al.*, 1995; Cui *et al.*, 2000; Pastuglia *et al.*, 1997; Schopfer *et al.*, 1999; Takayama *et al.*, 2000; Toriyama *et al.*, 1991; Yu *et al.*, 1996). In rice, a xylenase synthesized in the tapetum, ZmXYN1, is a prominent pollen coat component and important for the penetration of pollen tubes into the silk (Suen *et al.*, 2007). Arabinogalactan-rich glycoproteins (AGPs) are also tapetum derived materials that are important to guide the pollen tube growth in the process of fertilization (Coimbra *et al.*, 2007). These facts suggest that we should fully appreciate the important roles for the tapetum in order to better understand male gamete formation.

1.8 Current understanding about the endothecium and the middle layer: structure, functions and genetic controls.

The differentiation of the middle layer and the endothecium is less understood compared to that of the tapetum and microsporocytes. The genes important for the early anther development such as *SPL* and *BAM1/2* might play a role in the establishment of the identities of the endothecium and the middle layer. Recently, a receptor like kinase, *RPK2*, was found to play a role in the specification of the middle layer and the thickening of the endothecium cell wall (Mizuno *et al.*, 2007). However, the *RPK2* mRNA is abundant in the tapetum, and might be required to activate genes involved in cell wall metabolism and lignin biosynthesis. For this mutant, it is not clear if the developmental defects in the middle layer and endothecium are directly caused by the *rpk1* mutation or indirectly caused by tapetum dysfunctions. More genetic loci need to be discovered in order to get a better understanding about the formation of the middle layer and the endothecium.

The endothecium cell may have the functions of photosynthesis, cell wall lignification and dehiscence. In young anther, endothecium cells have chloroplasts. Under chilling treatment, starch will be accumulated in the endothecium. Evidence suggests that endothecium cells are either photosynthetic, or able to transport sugar. At late stages, the endothecium cell wall is thickened by lignification in preparation for anther dehiscence. The transcription factor gene *MYB26/MS35* regulates the genes involved in wall thickening: *Irregular Xylem 1 (IRX1)*, 3, 8, and 12 (Steiner-Lange *et al.*, 2003; Yang *et al.*, 2007). Mutants of *MYB26/MS35* exhibit defects in lignification and dehiscence (Steiner-Lange *et al.*, 2003; Yang *et al.*, 2007). Two NAC transcription factors, NAC SECONDARY WALL-PROMOTING FACTOR1 (NST1) and NST2, which also regulate the *IRX* genes, are downstream of *MYB26/MS35*. It was thought that *MYB26/MS35* regulates the *IRX* genes through the functions of *NST1/2* (Mitsuda *et al.*, 2005). Plant hormones also regulate the wall thickening of the endothecium. In *Arabidopsis*, the *AHP4* gene encodes a histidine-containing phosphotransfer factor 4 protein, which can negatively regulate the secondary cell wall thickening by repressing

the *IRX* gene expression in the endothecium cells (Shin *et al.*, 2008). It is activated by cytokinin, therefore, cytokinin may be also involved in the secondary wall thickening. In the investigations of anther indehiscence defects in cotton induced by treatment of glyphosate, a group found that the changes in microtubule and auxin level caused by glyphosate inhibited anther dehiscence (Yasuor *et al.*, 2006). This suggests that Auxin is a player in anther dehiscence. In barley, the *HvGAMYB* gene is expressed in the epidermis and the endothecium before dehiscence. Over expression of *HvGAMYB* causes the defects in dehiscence (Murray *et al.*, 2003). This suggests GA is involved in wall thickening of the endothecium. Recently, it was found that a MAP kinase kinase kinase gene from *Solanum chacoense*, *FERTILIZATION-RELATED KINASE 2* (*ScFRK2*), shows a similar expression pattern to that of *HvGAMYB* (O'Brien *et al.*, 2007). Initially, it is expressed in the tapetum and meiocytes in the young anthers, whereas at the dehiscence stage, it is only expressed in the epidermis and endothecium. The over expression of *ScFRK2* can disrupt dehiscence and pollen development suggesting the involvement of MAPK signaling in dehiscence. Transcriptional control is also needed. In maize, a MADS transcriptional factor, *ZmMADS2*, is accumulated in the degenerating nuclei of the endothecium and required for anther dehiscence (Ma *et al.*, 2008; Schreiber *et al.*, 2004). PCD in the endothecium is required for dehiscence. Interestingly, the endothecium PCD proceeds progressively without killing the cells until dehiscence. One type of molecules that are usually associated with PCD are cystein proteases. In *brinjal*, the *SmCP* gene encoding a cystein protease is expressed in the endothecium and other cells such as the xylem and nucellar cells that also undergo PCD (Xu *et al.*, 1999). The protease inhibitor might also be involved in PCD, which is suggested by that evidence that *OsPI8-1*, a protease inhibitor, is expressed in the middle layer and endothecium and has a role in PCD in rice, (Wang *et al.*, 2007).

The function of the middle layer is not well understood. It is a thin cell layer between the tapetum and the endothecium. Obviously, the aforementioned gene, *SPL*, *BAM1/2* and *RPK2*, play a role in the differentiation of the middle layer at very early stage. In maize, a genetic locus, *mac1*, is required for the middle layer identity because both the tapetum and the middle layer are missing in the *mac1* mutant (Ma *et al.*, 2007; Sheridan *et al.*,

1999). It is different from the *ems1* mutant which only lacks the tapetum. Recently, some genes that are preferentially expressed in the middle layer were isolated. For example, the *END1* gene from pea is a gene preferentially expressed in the epidermis, the endothecium and the middle layer, and the *END1* promoter has been used to regulate the male reproduction (Gomez *et al.*, 2004; Roque *et al.*, 2007). The middle layer also undergoes PCD before anther dehiscence. So far, we only have very limited information about the middle layer. Transcriptional analysis of the middle layer preferential genes may help to reveal its role in anther development and its relationship to other cells in the anther.

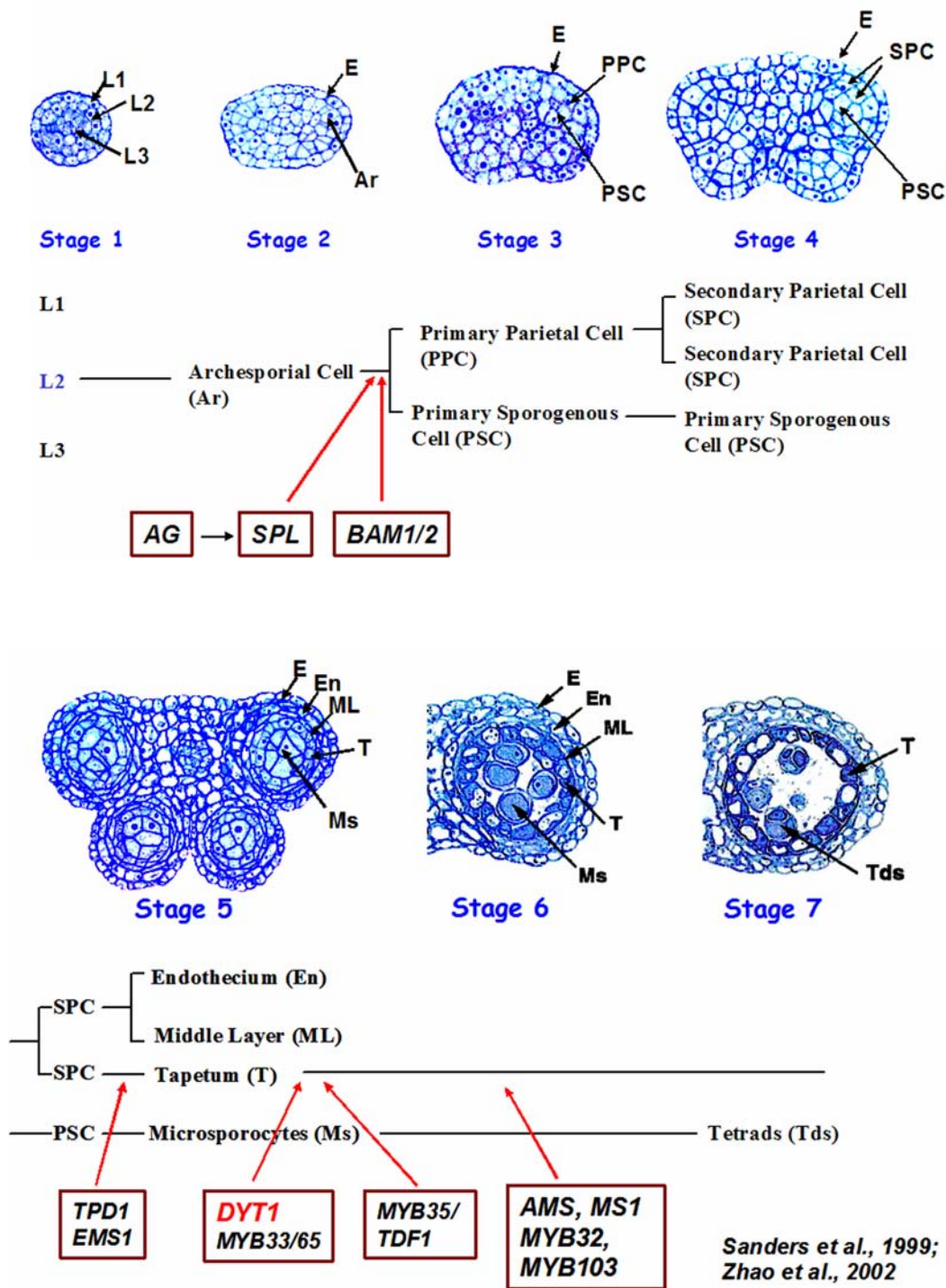


Figure 1.1 The stages and genes of early anther development.

1.9 Reference

- Ahn, J.H., Miller, D., Winter, V.J., Banfield, M.J., Lee, J.H., Yoo, S.Y., Henz, S.R., Brady, R.L. and Weigel, D. (2006) A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. *Embo J*, **25**, 605-614.
- Albrecht, C., Russinova, E., Hecht, V., Baaijens, E. and de Vries, S. (2005) The Arabidopsis thaliana SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1 and 2 control male sporogenesis. *Plant Cell*, **17**, 3337-3349.
- Alves-Ferreira, M., Wellmer, F., Banhara, A., Kumar, V., Riechmann, J.L. and Meyerowitz, E.M. (2007) Global expression profiling applied to the analysis of Arabidopsis stamen development. *Plant Physiol*, **145**, 747-762.
- Atanassov, I., Russinova, E., Antonov, L. and Atanassov, A. (1998) Expression of an anther-specific chalcone synthase-like gene is correlated with uninucleate microspore development in Nicotiana sylvestris. *Plant Mol Biol*, **38**, 1169-1178.
- Aubert, D., Chen, L., Moon, Y.H., Martin, D., Castle, L.A., Yang, C.H. and Sung, Z.R. (2001) EMF1, a novel protein involved in the control of shoot architecture and flowering in Arabidopsis. *Plant Cell*, **13**, 1865-1875.
- Aukerman, M.J. and Sakai, H. (2003) Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell*, **15**, 2730-2741.
- Balasubramanian, S. and Schneitz, K. (2000) NOZZLE regulates proximal-distal pattern formation, cell proliferation and early sporogenesis during ovule development in Arabidopsis thaliana. *Development*, **127**, 4227-4238.
- Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A. and Dean, C. (2004) Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature*, **427**, 164-167.
- Bleuyard, J.Y. and White, C.I. (2004) The Arabidopsis homologue of Xrcc3 plays an essential role in meiosis. *Embo J*, **23**, 439-449.
- Boyes, D.C. and Nasrallah, J.B. (1995) An anther-specific gene encoded by an S locus haplotype of Brassica produces complementary and differentially regulated transcripts. *Plant Cell*, **7**, 1283-1294.
- Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R. and Coen, E. (1997) Inflorescence commitment and architecture in Arabidopsis. *Science*, **275**, 80-83.
- Brand, U., Fletcher, J.C., Hobe, M., Meyerowitz, E.M. and Simon, R. (2000) Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity. *Science*, **289**, 617-619.
- Bucciaglia, P.A. and Smith, A.G. (1994) Cloning and characterization of Tag 1, a tobacco anther beta-1,3-glucanase expressed during tetrad dissolution. *Plant Mol Biol*, **24**, 903-914.
- Calonje, M., Sanchez, R., Chen, L. and Sung, Z.R. (2008) EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in Arabidopsis. *Plant Cell*, **20**, 277-291.
- Chen, C., Marcus, A., Li, W., Hu, Y., Calzada, J.P., Grossniklaus, U., Cyr, R.J. and Ma, H. (2002) The Arabidopsis ATK1 gene is required for spindle morphogenesis in male meiosis. *Development*, **129**, 2401-2409.

- Chen, C., Zhang, W., Timofejeva, L., Gerardin, Y. and Ma, H.** (2005) The Arabidopsis ROCK-N-ROLLERS gene encodes a homolog of the yeast ATP-dependent DNA helicase MER3 and is required for normal meiotic crossover formation. *Plant J*, **43**, 321-334.
- Chen, L., Cheng, J.C., Castle, L. and Sung, Z.R.** (1997) EMF genes regulate Arabidopsis inflorescence development. *Plant Cell*, **9**, 2011-2024.
- Chen, R., Zhao, X., Shao, Z., Wei, Z., Wang, Y., Zhu, L., Zhao, J., Sun, M., He, R. and He, G.** (2007) Rice UDP-glucose pyrophosphorylase1 is essential for pollen callose deposition and its cosuppression results in a new type of thermosensitive genic male sterility. *Plant Cell*, **19**, 847-861.
- Chen, X.** (2004) A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science*, **303**, 2022-2025.
- Coimbra, S., Almeida, J., Junqueira, V., Costa, M.L. and Pereira, L.G.** (2007) Arabinogalactan proteins as molecular markers in Arabidopsis thaliana sexual reproduction. *J Exp Bot*, **58**, 4027-4035.
- Colcombet, J., Boisson-Dernier, A., Ros-Palau, R., Vera, C.E. and Schroeder, J.I.** (2005) Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASES1 and 2 are essential for tapetum development and microspore maturation. *Plant Cell*, **17**, 3350-3361.
- Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C. and Coupland, G.** (2007) FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science*, **316**, 1030-1033.
- Cui, Y., Bi, Y.M., Brugiere, N., Arnoldo, M. and Rothstein, S.J.** (2000) The S locus glycoprotein and the S receptor kinase are sufficient for self-pollen rejection in Brassica. *Proc Natl Acad Sci U S A*, **97**, 3713-3717.
- De Muyt, A., Vezon, D., Gendrot, G., Gallois, J.L., Stevens, R. and Grelon, M.** (2007) AtPRD1 is required for meiotic double strand break formation in Arabidopsis thaliana. *Embo J*, **26**, 4126-4137.
- Drews, G.N., Bowman, J.L. and Meyerowitz, E.M.** (1991) Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. *Cell*, **65**, 991-1002.
- Elorza, A., Leon, G., Gomez, I., Mouras, A., Holuigue, L., Araya, A. and Jordana, X.** (2004) Nuclear SDH2-1 and SDH2-2 genes, encoding the iron-sulfur subunit of mitochondrial complex II in Arabidopsis, have distinct cell-specific expression patterns and promoter activities. *Plant Physiol*, **136**, 4072-4087.
- Fiers, M., Golemiec, E., Xu, J., van der Geest, L., Heidstra, R., Stiekema, W. and Liu, C.M.** (2005) The 14-amino acid CLV3, CLE19, and CLE40 peptides trigger consumption of the root meristem in Arabidopsis through a CLAVATA2-dependent pathway. *Plant Cell*, **17**, 2542-2553.
- Francis, K.E., Lam, S.Y. and Copenhaver, G.P.** (2006) Separation of Arabidopsis pollen tetrads is regulated by QUARTET1, a pectin methylesterase gene. *Plant Physiol*, **142**, 1004-1013.
- Goldberg, R.B., Beals, T.P. and Sanders, P.M.** (1993) Anther development: basic principles and practical applications. *Plant Cell*, **5**, 1217-1229.

- Golubovskaya, I., Avalkina, N. and Sheridan, W.F.** (1997) New insights into the role of the maize *ameiotic1* locus. *Genetics*, **147**, 1339-1350.
- Golubovskaya, I., Grebennikova, Z.K., Avalkina, N.A. and Sheridan, W.F.** (1993) The role of the *ameiotic1* gene in the initiation of meiosis and in subsequent meiotic events in maize. *Genetics*, **135**, 1151-1166.
- Gomez, M.D., Beltran, J.P. and Canas, L.A.** (2004) The pea *END1* promoter drives anther-specific gene expression in different plant species. *Planta*, **219**, 967-981.
- Goto, K. and Meyerowitz, E.M.** (1994) Function and regulation of the Arabidopsis floral homeotic gene *PISTILLATA*. *Genes Dev*, **8**, 1548-1560.
- Grelon, M., Vezon, D., Gendrot, G. and Pelletier, G.** (2001) *AtSPO11-1* is necessary for efficient meiotic recombination in plants. *Embo J*, **20**, 589-600.
- Gustafson-Brown, C., Savidge, B. and Yanofsky, M.F.** (1994) Regulation of the arabidopsis floral homeotic gene *APETALA1*. *Cell*, **76**, 131-143.
- Hamant, O., Ma, H. and Cande, W.Z.** (2006) Genetics of meiotic prophase I in plants. *Annu Rev Plant Biol*, **57**, 267-302.
- Hara, K., Kajita, R., Torii, K.U., Bergmann, D.C. and Kakimoto, T.** (2007) The secretory peptide gene *EPF1* enforces the stomatal one-cell-spacing rule. *Genes Dev*, **21**, 1720-1725.
- Hartung, F., Wurzel-Wildersinn, R., Fuchs, J., Schubert, I., Suer, S. and Puchta, H.** (2007) The catalytically active tyrosine residues of both *SPO11-1* and *SPO11-2* are required for meiotic double-strand break induction in Arabidopsis. *Plant Cell*, **19**, 3090-3099.
- He, K., Gou, X., Yuan, T., Lin, H., Asami, T., Yoshida, S., Russell, S.D. and Li, J.** (2007) *BAK1* and *BKK1* regulate brassinosteroid-dependent growth and brassinosteroid-independent cell-death pathways. *Curr Biol*, **17**, 1109-1115.
- He, Y., Michaels, S.D. and Amasino, R.M.** (2003) Regulation of flowering time by histone acetylation in Arabidopsis. *Science*, **302**, 1751-1754.
- Heese, A., Hann, D.R., Gimenez-Ibanez, S., Jones, A.M., He, K., Li, J., Schroeder, J.I., Peck, S.C. and Rathjen, J.P.** (2007) The receptor-like kinase *SERK3/BAK1* is a central regulator of innate immunity in plants. *Proc Natl Acad Sci U S A*, **104**, 12217-12222.
- Helliwell, C.A., Wood, C.C., Robertson, M., James Peacock, W. and Dennis, E.S.** (2006) The Arabidopsis *FLC* protein interacts directly in vivo with *SOC1* and *FT* chromatin and is part of a high-molecular-weight protein complex. *Plant J*, **46**, 183-192.
- Hepworth, S.R., Valverde, F., Ravenscroft, D., Mouradov, A. and Coupland, G.** (2002) Antagonistic regulation of flowering-time gene *SOC1* by *CONSTANS* and *FLC* via separate promoter motifs. *Embo J*, **21**, 4327-4337.
- Hernandez-Pinzon, I., Ross, J.H., Barnes, K.A., Damant, A.P. and Murphy, D.J.** (1999) Composition and role of tapetal lipid bodies in the biogenesis of the pollen coat of *Brassica napus*. *Planta*, **208**, 588-598.
- Higgins, J.D., Armstrong, S.J., Franklin, F.C. and Jones, G.H.** (2004) The Arabidopsis *MutS* homolog *AtMSH4* functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. *Genes Dev*, **18**, 2557-2570.

- Hird, D.L., Worrall, D., Hodge, R., Smartt, S., Paul, W. and Scott, R.** (1993) The anther-specific protein encoded by the *Brassica napus* and *Arabidopsis thaliana* A6 gene displays similarity to beta-1,3-glucanases. *Plant J*, **4**, 1023-1033.
- Honma, T. and Goto, K.** (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, **409**, 525-529.
- Hord, C.L., Chen, C., Deyoung, B.J., Clark, S.E. and Ma, H.** (2006) The BAM1/BAM2 receptor-like kinases are important regulators of *Arabidopsis* early anther development. *Plant Cell*, **18**, 1667-1680.
- Hsieh, K. and Huang, A.H.** (2005) Lipid-rich tapetosomes in *Brassica* tapetum are composed of oleosin-coated oil droplets and vesicles, both assembled in and then detached from the endoplasmic reticulum. *Plant J*, **43**, 889-899.
- Hsieh, K. and Huang, A.H.** (2007) Tapetosomes in *Brassica* tapetum accumulate endoplasmic reticulum-derived flavonoids and alkanes for delivery to the pollen surface. *Plant Cell*, **19**, 582-596.
- Ito, T., Nagata, N., Yoshiba, Y., Ohme-Takagi, M., Ma, H. and Shinozaki, K.** (2007) *Arabidopsis* MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development. *Plant Cell*, **19**, 3549-3562.
- Ito, T. and Shinozaki, K.** (2002) The MALE STERILITY1 gene of *Arabidopsis*, encoding a nuclear protein with a PHD-finger motif, is expressed in tapetal cells and is required for pollen maturation. *Plant Cell Physiol*, **43**, 1285-1292.
- Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J.L. and Meyerowitz, E.M.** (2004) The homeotic protein AGAMOUS controls microsporogenesis by regulation of SPOROCTELESS. *Nature*, **430**, 356-360.
- Ito, Y., Nakanomyo, I., Motose, H., Iwamoto, K., Sawa, S., Dohmae, N. and Fukuda, H.** (2006) Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science*, **313**, 842-845.
- Jack, T., Fox, G.L. and Meyerowitz, E.M.** (1994) *Arabidopsis* homeotic gene APETALA3 ectopic expression: transcriptional and posttranscriptional regulation determine floral organ identity. *Cell*, **76**, 703-716.
- Jakobsen, M.K., Poulsen, L.R., Schulz, A., Fleurat-Lessard, P., Moller, A., Husted, S., Schiott, M., Amtmann, A. and Palmgren, M.G.** (2005) Pollen development and fertilization in *Arabidopsis* is dependent on the MALE GAMETOGENESIS IMPAIRED ANTHERS gene encoding a type V P-type ATPase. *Genes Dev*, **19**, 2757-2769.
- Jinn, T.L., Stone, J.M. and Walker, J.C.** (2000) HAESA, an *Arabidopsis* leucine-rich repeat receptor kinase, controls floral organ abscission. *Genes Dev*, **14**, 108-117.
- Kapoor, S., Kobayashi, A. and Takatsuji, H.** (2002) Silencing of the tapetum-specific zinc finger gene TAZ1 causes premature degeneration of tapetum and pollen abortion in *petunia*. *Plant Cell*, **14**, 2353-2367.
- Klimyuk, V.I. and Jones, J.D.** (1997) AtDMC1, the *Arabidopsis* homologue of the yeast DMC1 gene: characterization, transposon-induced allelic variation and meiosis-associated expression. *Plant J*, **11**, 1-14.
- Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. and Araki, T.** (1999) A pair of related genes with antagonistic roles in mediating flowering signals. *Science*, **286**, 1960-1962.

- Komeda, Y.** (2004) Genetic regulation of time to flower in *Arabidopsis thaliana*. *Annu Rev Plant Biol*, **55**, 521-535.
- Kramer, E.M. and Irish, V.F.** (1999) Evolution of genetic mechanisms controlling petal development. *Nature*, **399**, 144-148.
- Kwak, S.H. and Schiefelbein, J.** (2007) The role of the SCRAMBLED receptor-like kinase in patterning the *Arabidopsis* root epidermis. *Dev Biol*, **302**, 118-131.
- Kwak, S.H., Shen, R. and Schiefelbein, J.** (2005) Positional signaling mediated by a receptor-like kinase in *Arabidopsis*. *Science*, **307**, 1111-1113.
- Lamb, R.S., Hill, T.A., Tan, Q.K. and Irish, V.F.** (2002) Regulation of APETALA3 floral homeotic gene expression by meristem identity genes. *Development*, **129**, 2079-2086.
- Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W., Zhao, H., Lee, I. and Deng, X.W.** (2007) Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell*, **19**, 731-749.
- Lee, Y.H. and Tegeder, M.** (2004) Selective expression of a novel high-affinity transport system for acidic and neutral amino acids in the tapetum cells of *Arabidopsis* flowers. *Plant J*, **40**, 60-74.
- Levy, Y.Y., Mesnage, S., Mylne, J.S., Gendall, A.R. and Dean, C.** (2002) Multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control. *Science*, **297**, 243-246.
- Li, J., Wen, J., Lease, K.A., Doke, J.T., Tax, F.E. and Walker, J.C.** (2002) BAK1, an *Arabidopsis* LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell*, **110**, 213-222.
- Li, N., Zhang, D.S., Liu, H.S., Yin, C.S., Li, X.X., Liang, W.Q., Yuan, Z., Xu, B., Chu, H.W., Wang, J., Wen, T.Q., Huang, H., Luo, D., Ma, H. and Zhang, D.B.** (2006) The rice tapetum degeneration retardation gene is required for tapetum degradation and anther development. *Plant Cell*, **18**, 2999-3014.
- Li, W., Chen, C., Markmann-Mulisch, U., Timofejeva, L., Schmelzer, E., Ma, H. and Reiss, B.** (2004) The *Arabidopsis* AtRAD51 gene is dispensable for vegetative development but required for meiosis. *Proc Natl Acad Sci U S A*, **101**, 10596-10601.
- Li, W., Yang, X., Lin, Z., Timofejeva, L., Xiao, R., Makaroff, C.A. and Ma, H.** (2005) The AtRAD51C gene is required for normal meiotic chromosome synapsis and double-stranded break repair in *Arabidopsis*. *Plant Physiol*, **138**, 965-976.
- Liljgren, S.J., Gustafson-Brown, C., Pinyopich, A., Ditta, G.S. and Yanofsky, M.F.** (1999) Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *Plant Cell*, **11**, 1007-1018.
- Ma, H.** (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annu Rev Plant Biol*, **56**, 393-434.
- Ma, J., Duncan, D., Morrow, D.J., Fernandes, J. and Walbot, V.** (2007) Transcriptome profiling of maize anthers using genetic ablation to analyze pre-meiotic and tapetal cell types. *Plant J*, **50**, 637-648.
- Ma, J., Yan, B., Qu, Y., Qin, F., Yang, Y., Hao, X., Yu, J., Zhao, Q., Zhu, D. and Ao, G.** (2008) Zm401, a short-open reading-frame mRNA or noncoding RNA, is

- essential for tapetum and microspore development and can regulate the floret formation in maize. *J Cell Biochem*.
- Mamun, E.A., Alfred, S., Cantrill, L.C., Overall, R.L. and Sutton, B.G.** (2006) Effects of chilling on male gametophyte development in rice. *Cell Biol Int*, **30**, 583-591.
- Marcus, A.I., Ambrose, J.C., Blickley, L., Hancock, W.O. and Cyr, R.J.** (2002) Arabidopsis thaliana protein, ATK1, is a minus-end directed kinesin that exhibits non-processive movement. *Cell Motil Cytoskeleton*, **52**, 144-150.
- McCormick, S.** (1991) Molecular analysis of male gametogenesis in plants. *Trends Genet*, **7**, 298-303.
- McGonigle, B., Bouhidel, K. and Irish, V.F.** (1996) Nuclear localization of the Arabidopsis APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression. *Genes Dev*, **10**, 1812-1821.
- Mercier, R., Armstrong, S.J., Horlow, C., Jackson, N.P., Makaroff, C.A., Vezon, D., Pelletier, G., Jones, G.H. and Franklin, F.C.** (2003) The meiotic protein SWI1 is required for axial element formation and recombination initiation in Arabidopsis. *Development*, **130**, 3309-3318.
- Mercier, R., Vezon, D., Bullier, E., Motamayor, J.C., Sellier, A., Lefevre, F., Pelletier, G. and Horlow, C.** (2001) SWITCH1 (SWI1): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis. *Genes Dev*, **15**, 1859-1871.
- Michaels, S.D. and Amasino, R.M.** (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell*, **11**, 949-956.
- Millar, A.A. and Gubler, F.** (2005) The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell*, **17**, 705-721.
- Mitsuda, N., Seki, M., Shinozaki, K. and Ohme-Takagi, M.** (2005) The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell*, **17**, 2993-3006.
- Mizukami, Y. and Ma, H.** (1992) Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity. *Cell*, **71**, 119-131.
- Mizuno, S., Osakabe, Y., Maruyama, K., Ito, T., Osakabe, K., Sato, T., Shinozaki, K. and Yamaguchi-Shinozaki, K.** (2007) Receptor-like protein kinase 2 (RPK 2) is a novel factor controlling anther development in Arabidopsis thaliana. *Plant J*, **50**, 751-766.
- Moon, J., Suh, S.S., Lee, H., Choi, K.R., Hong, C.B., Paek, N.C., Kim, S.G. and Lee, I.** (2003) The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. *Plant J*, **35**, 613-623.
- Moon, Y.H., Chen, L., Pan, R.L., Chang, H.S., Zhu, T., Maffeo, D.M. and Sung, Z.R.** (2003) EMF genes maintain vegetative development by repressing the flower program in Arabidopsis. *Plant Cell*, **15**, 681-693.
- Murray, F., Kalla, R., Jacobsen, J. and Gubler, F.** (2003) A role for HvGAMYB in anther development. *Plant J*, **33**, 481-491.

- Nadeau, J.A. and Sack, F.D.** (2002) Control of stomatal distribution on the Arabidopsis leaf surface. *Science*, **296**, 1697-1700.
- Nam, K.H. and Li, J.** (2002) BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell*, **110**, 203-212.
- Napoli, C.A., Fahy, D., Wang, H.Y. and Taylor, L.P.** (1999) white anther: A petunia mutant that abolishes pollen flavonol accumulation, induces male sterility, and is complemented by a chalcone synthase transgene. *Plant Physiol*, **120**, 615-622.
- Nardi, M.C., Feron, R., Navazio, L., Mariani, P., Pierson, E., Wolters-Arts, M., Knuiman, B., Mariani, C. and Derksen, J.** (2006) Expression and localization of calreticulin in tobacco anthers and pollen tubes. *Planta*, **223**, 1263-1271.
- Nelson, D.E., Glaunsinger, B. and Bohnert, H.J.** (1997) Abundant accumulation of the calcium-binding molecular chaperone calreticulin in specific floral tissues of Arabidopsis thaliana. *Plant Physiol*, **114**, 29-37.
- Ng, M. and Yanofsky, M.F.** (2001) Activation of the Arabidopsis B class homeotic genes by APETALA1. *Plant Cell*, **13**, 739-753.
- O'Brien, M., Gray-Mitsumune, M., Kapfer, C., Bertrand, C. and Matton, D.P.** (2007) The ScFRK2 MAP kinase kinase kinase from Solanum chacoense affects pollen development and viability. *Planta*, **225**, 1221-1231.
- Ogawa, M., Shinohara, H., Sakagami, Y. and Matsubayashi, Y.** (2008) Arabidopsis CLV3 peptide directly binds CLV1 ectodomain. *Science*, **319**, 294.
- Pastuglia, M., Ruffio-Chable, V., Delorme, V., Gaude, T., Dumas, C. and Cock, J.M.** (1997) A functional S locus anther gene is not required for the self-incompatibility response in Brassica oleracea. *Plant Cell*, **9**, 2065-2076.
- Pawlowski, W.P., Golubovskaya, I.N., Timofejeva, L., Meeley, R.B., Sheridan, W.F. and Cande, W.Z.** (2004) Coordination of meiotic recombination, pairing, and synapsis by PHS1. *Science*, **303**, 89-92.
- Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M.F.** (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, **405**, 200-203.
- Pollak, P.E., Vogt, T., Mo, Y. and Taylor, L.P.** (1993) Chalcone Synthase and Flavonol Accumulation in Stigmas and Anthers of Petunia hybrida. *Plant Physiol*, **102**, 925-932.
- Preston, J., Wheeler, J., Heazlewood, J., Li, S.F. and Parish, R.W.** (2004) AtMYB32 is required for normal pollen development in Arabidopsis thaliana. *Plant J*, **40**, 979-995.
- Quan, L., Xiao, R., Li, W., Oh, S.A., Kong, H., Ambrose, J.C., Malcos, J.L., Cyr, R., Twell, D. and Ma, H.** (2008) Functional divergence of the duplicated AtKIN14a and AtKIN14b genes: critical roles in Arabidopsis meiosis and gametophyte development. *Plant J*, **53**, 1013-1026.
- Ravi, M., Marimuthu, M.P. and Siddiqi, I.** (2008) Gamete formation without meiosis in Arabidopsis. *Nature*, **451**, 1121-1124.
- Rhee, S.Y. and Somerville, C.R.** (1998) Tetrad pollen formation in quartet mutants of Arabidopsis thaliana is associated with persistence of pectic polysaccharides of the pollen mother cell wall. *Plant J*, **15**, 79-88.

- Roberts, M.R., Hodge, R., Ross, J.H., Sorensen, A., Murphy, D.J., Draper, J. and Scott, R.** (1993) Characterization of a new class of oleosins suggests a male gametophyte-specific lipid storage pathway. *Plant J*, **3**, 629-636.
- Roque, E., Gomez, M.D., Ellul, P., Wallbraun, M., Madueno, F., Beltran, J.P. and Canas, L.A.** (2007) The PsEND1 promoter: a novel tool to produce genetically engineered male-sterile plants by early anther ablation. *Plant Cell Rep*, **26**, 313-325.
- Ross, J.H. and Murphy, D.J.** (1996) Characterization of anther-expressed genes encoding a major class of extracellular oleosin-like proteins in the pollen coat of Brassicaceae. *Plant J*, **9**, 625-637.
- Ruiter, R.K., Van Eldik, G.J., Van Herpen, R.M., Schrauwen, J.A. and Wullems, G.J.** (1997) Characterization of oleosins in the pollen coat of Brassica oleracea. *Plant Cell*, **9**, 1621-1631.
- Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F. and Coupland, G.** (2000) Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science*, **288**, 1613-1616.
- Sanchez-Moran, E., Santos, J.L., Jones, G.H. and Franklin, F.C.** (2007) ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in Arabidopsis. *Genes Dev*, **21**, 2220-2233.
- Schopfer, C.R., Nasrallah, M.E. and Nasrallah, J.B.** (1999) The male determinant of self-incompatibility in Brassica. *Science*, **286**, 1697-1700.
- Schreiber, D.N., Bantin, J. and Dresselhaus, T.** (2004) The MADS box transcription factor ZmMADS2 is required for anther and pollen maturation in maize and accumulates in apoptotic bodies during anther dehiscence. *Plant Physiol*, **134**, 1069-1079.
- Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasino, R.A. and Coupland, G.** (2006) The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. *Genes Dev*, **20**, 898-912.
- Shah, K., Russinova, E., Gadella, T.W., Jr., Willemse, J. and De Vries, S.C.** (2002) The Arabidopsis kinase-associated protein phosphatase controls internalization of the somatic embryogenesis receptor kinase 1. *Genes Dev*, **16**, 1707-1720.
- Shannon, S. and Meeks-Wagner, D.R.** (1991) A Mutation in the Arabidopsis TFL1 Gene Affects Inflorescence Meristem Development. *Plant Cell*, **3**, 877-892.
- Shannon, S. and Meeks-Wagner, D.R.** (1993) Genetic Interactions That Regulate Inflorescence Development in Arabidopsis. *Plant Cell*, **5**, 639-655.
- Sheridan, W.F., Golubeva, E.A., Abrhamova, L.I. and Golubovskaya, I.N.** (1999) The mac1 mutation alters the developmental fate of the hypodermal cells and their cellular progeny in the maize anther. *Genetics*, **153**, 933-941.
- Shin, J.S., Jung, K.W., Oh, S.I., Kim, Y.Y., Yoo, K.S. and Cui, M.H.** (2008) Arabidopsis Histidine-containing Phosphotransfer Factor 4 (AHP4) Negatively Regulates Secondary Wall Thickening of the Anther Endothecium during Flowering. *Mol Cells*, **25**.
- Shpak, E.D., McAbee, J.M., Pillitteri, L.J. and Torii, K.U.** (2005) Stomatal patterning and differentiation by synergistic interactions of receptor kinases. *Science*, **309**, 290-293.

- Sieber, P., Petrascheck, M., Barberis, A. and Schneitz, K.** (2004) Organ polarity in Arabidopsis. NOZZLE physically interacts with members of the YABBY family. *Plant Physiol*, **135**, 2172-2185.
- Simon, R., Igeno, M.I. and Coupland, G.** (1996) Activation of floral meristem identity genes in Arabidopsis. *Nature*, **384**, 59-62.
- Sorensen, A.M., Krober, S., Unte, U.S., Huijser, P., Dekker, K. and Saedler, H.** (2003) The Arabidopsis ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. *Plant J*, **33**, 413-423.
- Stacey, N.J., Kuromori, T., Azumi, Y., Roberts, G., Breuer, C., Wada, T., Maxwell, A., Roberts, K. and Sugimoto-Shirasu, K.** (2006) Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. *Plant J*, **48**, 206-216.
- Steiner-Lange, S., Unte, U.S., Eckstein, L., Yang, C., Wilson, Z.A., Schmelzer, E., Dekker, K. and Saedler, H.** (2003) Disruption of Arabidopsis thaliana MYB26 results in male sterility due to non-dehiscent anthers. *Plant J*, **34**, 519-528.
- Suen, D.F. and Huang, A.H.** (2007) Maize pollen coat xylanase facilitates pollen tube penetration into silk during sexual reproduction. *J Biol Chem*, **282**, 625-636.
- Takayama, S., Shiba, H., Iwano, M., Shimosato, H., Che, F.S., Kai, N., Watanabe, M., Suzuki, G., Hinata, K. and Isogai, A.** (2000) The pollen determinant of self-incompatibility in Brassica campestris. *Proc Natl Acad Sci U S A*, **97**, 1920-1925.
- Tamaki, S., Matsuo, S., Wong, H.L., Yokoi, S. and Shimamoto, K.** (2007) Hd3a protein is a mobile flowering signal in rice. *Science*, **316**, 1033-1036.
- Thorstensen, T., Grini, P.E., Mercy, I.S., Alm, V., Erdal, S., Aasland, R. and Aalen, R.B.** (2008) The Arabidopsis SET-domain protein ASHR3 is involved in stamen development and interacts with the bHLH transcription factor ABORTED MICROSPORES (AMS). *Plant Mol Biol*, **66**, 47-59.
- Ting, J.T., Wu, S.S., Ratnayake, C. and Huang, A.H.** (1998) Constituents of the tapetosomes and elaioplasts in Brassica campestris tapetum and their degradation and retention during microsporogenesis. *Plant J*, **16**, 541-551.
- Toriyama, K., Thorsness, M.K., Nasrallah, J.B. and Nasrallah, M.E.** (1991) A Brassica S locus gene promoter directs sporophytic expression in the anther tapetum of transgenic Arabidopsis. *Dev Biol*, **143**, 427-431.
- van der Meer, I.M., Stam, M.E., van Tunen, A.J., Mol, J.N. and Stuitje, A.R.** (1992) Antisense inhibition of flavonoid biosynthesis in petunia anthers results in male sterility. *Plant Cell*, **4**, 253-262.
- van Eldik, G.J., Reijnen, W.H., Ruiter, R.K., van Herpen, M.M., Schrauwen, J.A. and Wullems, G.J.** (1997) Regulation of flavonol biosynthesis during anther and pistil development, and during pollen tube growth in Solanum tuberosum. *Plant J*, **11**, 105-113.
- Vizcay-Barrena, G. and Wilson, Z.A.** (2006) Altered tapetal PCD and pollen wall development in the Arabidopsis ms1 mutant. *J Exp Bot*, **57**, 2709-2717.
- Wang, A., Xia, Q., Xie, W., Datla, R. and Selvaraj, G.** (2003) The classical Ubisch bodies carry a sporophytically produced structural protein (RAFTIN) that is essential for pollen development. *Proc Natl Acad Sci U S A*, **100**, 14487-14492.
- Wang, J., Shi, Z.Y., Wan, X.S., Shen, G.Z. and Zhang, J.L.** (2007) The expression pattern of a rice proteinase inhibitor gene OsPI8-1 implies its role in plant development. *J Plant Physiol*.

- Wang, T.W., Balsamo, R.A., Ratnayake, C., Platt, K.A., Ting, J.T. and Huang, A.H.** (1997) Identification, subcellular localization, and developmental studies of oleosins in the anther of *Brassica napus*. *Plant J*, **11**, 475-487.
- Wijeratne, A.J., Chen, C., Zhang, W., Timofejeva, L. and Ma, H.** (2006) The *Arabidopsis thaliana* PARTING DANCERS gene encoding a novel protein is required for normal meiotic homologous recombination. *Mol Biol Cell*, **17**, 1331-1343.
- Wijeratne, A.J., Zhang, W., Sun, Y., Liu, W., Albert, R., Zheng, Z., Oppenheimer, D.G., Zhao, D. and Ma, H.** (2007) Differential gene expression in *Arabidopsis* wild-type and mutant anthers: insights into anther cell differentiation and regulatory networks. *Plant J*, **52**, 14-29.
- Wilson, Z.A., Morroll, S.M., Dawson, J., Swarup, R. and Tighe, P.J.** (2001) The *Arabidopsis* MALE STERILITY1 (MS1) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. *Plant J*, **28**, 27-39.
- Worrall, D., Hird, D.L., Hodge, R., Paul, W., Draper, J. and Scott, R.** (1992) Premature dissolution of the microsporocyte callose wall causes male sterility in transgenic tobacco. *Plant Cell*, **4**, 759-771.
- Wu, S.S., Moreau, R.A., Whitaker, B.D. and Huang, A.H.** (1999) Steryl esters in the elaioplasts of the tapetum in developing *Brassica* anthers and their recovery on the pollen surface. *Lipids*, **34**, 517-523.
- Xu, F.X. and Chye, M.L.** (1999) Expression of cysteine proteinase during developmental events associated with programmed cell death in brinjal. *Plant J*, **17**, 321-327.
- Yang, C., Vizcay-Barrena, G., Conner, K. and Wilson, Z.A.** (2007) MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. *Plant Cell*, **19**, 3530-3548.
- Yang, C., Xu, Z., Song, J., Conner, K., Vizcay Barrena, G. and Wilson, Z.A.** (2007) *Arabidopsis* MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. *Plant Cell*, **19**, 534-548.
- Yang, C.H., Chen, L.J. and Sung, Z.R.** (1995) Genetic regulation of shoot development in *Arabidopsis*: role of the EMF genes. *Dev Biol*, **169**, 421-435.
- Yang, S.L., Jiang, L., Puah, C.S., Xie, L.F., Zhang, X.Q., Chen, L.Q., Yang, W.C. and Ye, D.** (2005) Overexpression of TAPETUM DETERMINANT1 alters the cell fates in the *Arabidopsis* carpel and tapetum via genetic interaction with excess microsporocytes1/extra sporogenous cells. *Plant Physiol*, **139**, 186-191.
- Yang, S.L., Xie, L.F., Mao, H.Z., Puah, C.S., Yang, W.C., Jiang, L., Sundaresan, V. and Ye, D.** (2003) Tapetum determinant1 is required for cell specialization in the *Arabidopsis* anther. *Plant Cell*, **15**, 2792-2804.
- Yang, W.C., Ye, D., Xu, J. and Sundaresan, V.** (1999) The SPOROCTELESS gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev*, **13**, 2108-2117.
- Yang, X.Y., Li, J.G., Pei, M., Gu, H., Chen, Z.L. and Qu, L.J.** (2007) Over-expression of a flower-specific transcription factor gene AtMYB24 causes aberrant anther development. *Plant Cell Rep*, **26**, 219-228.

- Yasuor, H., Abu-Abied, M., Belausov, E., Madmony, A., Sadot, E., Riov, J. and Rubin, B.** (2006) Glyphosate-induced anther indehiscence in cotton is partially temperature dependent and involves cytoskeleton and secondary wall modifications and auxin accumulation. *Plant Physiol*, **141**, 1306-1315.
- Ylstra, B., Touraev, A., Moreno, R.M., Stoger, E., van Tunen, A.J., Vicente, O., Mol, J.N. and Heberle-Bors, E.** (1992) Flavonols Stimulate Development, Germination, and Tube Growth of Tobacco Pollen. *Plant Physiol*, **100**, 902-907.
- Yu, K., Schafer, U., Glavin, T.L., Goring, D.R. and Rothstein, S.J.** (1996) Molecular characterization of the S locus in two self-incompatible Brassica napus lines. *Plant Cell*, **8**, 2369-2380.
- Yui, R., Iketani, S., Mikami, T. and Kubo, T.** (2003) Antisense inhibition of mitochondrial pyruvate dehydrogenase E1alpha subunit in anther tapetum causes male sterility. *Plant J*, **34**, 57-66.
- Zhang, W., Sun, Y., Timofejeva, L., Chen, C., Grossniklaus, U. and Ma, H.** (2006) Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development*, **133**, 3085-3095.
- Zhang, Z.B., Zhu, J., Gao, J.F., Wang, C., Li, H., Zhang, H.Q., Zhang, S., Wang, D.M., Wang, Q.X., Huang, H., Xia, H.J. and Yang, Z.N.** (2007) Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis. *Plant J*, **52**, 528-538.
- Zhao, D.Z., Wang, G.F., Speal, B. and Ma, H.** (2002) The excess microsporocytes1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the Arabidopsis anther. *Genes Dev*, **16**, 2021-2031.
- Zhao, X., de Palma, J., Oane, R., Gamuyao, R., Luo, M., Chaudhury, A., Herve, P., Xue, Q. and Bennett, J.** (2008) OsTDL1A binds to the LRR domain of rice receptor kinase MSP1, and is required to limit sporocyte numbers. *Plant J*, **54**, 375-387.
- Zhu, J., Chen, H., Li, H., Gao, J.F., Jiang, H., Wang, C., Guan, Y.F. and Yang, Z.N.** (2008) Defective in Tapetal Development and Function1 (TDF1) is Essential for Anther Development and Tapetal Function for Microspore Maturation in Arabidopsis. *Plant J*.

Chapter 2

***DYT1* and *AMS* regulate common and distinct sets of genes in the anther**

Dr. Yiben Peng, a visiting scholar from Peking University helped with anther sample collection. Dr. Craig Paul and Dr. Haikun Zhang in the Microarray facility at Penn State did the hybridization. Dr. Naomi Altman did the normalization of microarray data. Xuan Ma contributed to the data analysis.

2.1 ABSTRACT

The tapetum is an important sporophytic tissue in the anther. It surrounds microsporocytes and supports microsporogenesis by providing nutrient and signal molecules (Ma, 2005). Previous genetic screens for male sterile mutants have led to the discoveries of two bHLH transcription factor genes that are required for tapetum development: *DYT1* and *AMS* (Sorensen *et al.*, 2003; Zhang *et al.*, 2006). *DYT1* is highly expressed at the stage of meiosis, whereas *AMS* is highly expressed after meiosis and before the release of microspores from the tetrad. Mutations in either gene cause defects in the tapetum and the degeneration of meiocytes or microspores. *DYT1* might positively regulate *AMS* because the expression of *AMS* is reduced in the *dyt1-1* mutant. An important question is how *DYT1* and *AMS* regulate anther development. Previous studies revealed only several known genes that act downstream of *DYT1* (Zhang *et al.*, 2006). Little is known about the downstream genes of *AMS*. A genome-wide analysis of the *DYT1* and *AMS* downstream genes is needed in order to understand the roles for *DYT1* and *AMS* in anther development.

In this study, I analyzed the gene expression patterns in the *dyt1-1* and *ams* mutants by performing microarray analyses, and uncovered the differentially expressed genes in each mutant. The results suggest *DYT1* regulate a broad spectrum of tapetum functions including lipid metabolism, cell wall changes, and secondary metabolism in part by activating at least 32 transcription factor genes. It was revealed that the *dyt1-1* and *ams* transcriptomes partially overlap, which suggests that part of the transcriptional regulation by *DYT1* is mediated by *AMS*. However, *DYT1* and *AMS* also have different downstream genes. For example, the expressions of heat shock protein genes were increased in the *dyt1-1* mutant anther but not in the *ams* mutant anther, whereas a group of plastid genes show reduced expression in *ams* but not *dyt1-1*. These findings provide new insights into the transcriptional regulation of tapetum functions and useful information to generate hypotheses about novel genes and functions in the tapetum.

2.2 INTRODUCTION

In *Arabidopsis*, the proper differentiation and function of tapetum cells are required for pollen formation and thus the success of male reproduction (Ma, 2005). Although the pollen grains are formed from pollen mother cells, the neighboring tapetal cells are not dispensable. They support the maturation of pollen mother cells and pollen by active biosynthesis, secretion and deposition of different types of molecules: lipid, polysaccharide, protein and other metabolites. Defects in tapetum usually cause failures in pollen formation. Given the important roles of the tapetum, it is critical to know how the tapetum cells are formed.

Transcriptional regulation is a major mechanism for the regulation of plant development as well as tapetum differentiation; this was revealed from the discoveries of several genes. *SPL* is required for the differentiation of early anther cells including the tapetum progenitor cells because in the *spl* mutants, all cells inside the epidermis fail to differentiate (Balasubramanian *et al.*, 2000; Ito *et al.*, 2004; Yang *et al.*, 1999). *SPL* encodes a novel transcription factor and is a direct target of *AG* (Ito *et al.*, 2004). This suggests that floral homeotic genes including *AG* and perhaps *AP3* and *PI* as well direct anther cell differentiation through a transcriptional cascade in which *SPL* is an early factor. TPD1/EMS1 kinase signaling is also critical for the establishment of the tapetum identity since mutations in either gene will cause “lack of tapetum” phenotype (Yang *et al.*, 2005; Yang *et al.*, 2003; Zhao *et al.*, 2002; Zhao *et al.*, 2008). This suggests that tapetum differentiation is regulated by developmental cues such as TPD1 emitted from pollen mother cells. Transcription profiling of the *spl* and *ems1* mutants has revealed that a large number of tapetum preferential genes including transcription factor genes are downstream of *SPL* and *EMS1*. This suggests that transcriptional regulation is a key mechanism of tapetum differentiation and an important output of signaling processes.

Two anther preferential bHLH transcription factors, DYT1 and AMS, act downstream of *SPL* and *EMS1* and are important to tapetum development (Sorensen *et al.*, 2003; Wijeratne *et al.*, 2007; Zhang *et al.*, 2006). *DYT1* has a peak expression in the tapetum at stages 5 and 6. In the *dyl1-1* mutant, the tapetum exhibits dramatic phenotypes: cell

enlargement, vacuolation, and degeneration. Previous studies suggested that some tapetum preferential genes are downstream of *DYT1* including the genes encoding transcription factors such as *AMS* and *MSI*, and some metabolic genes. However, it is not clear what are other genes regulated by *DYT1*. *AMS* is expressed later than *DYT1* and might be involved in the tapetum function during pollen wall formation. *TDR*, a possible ortholog of *AMS* in rice, regulates the program cell death in the tapetum by directly regulating the expression of cysteine protease and protease inhibitor genes (Li *et al.*, 2006). Whether *AMS* has similar functions is not known, and very little is known about the downstream genes of *AMS*.

To understand the roles of *DYT1* and *AMS* in the tapetum, I collected large numbers of small anthers (stage 4-7) under microscope and carried out microarray analyses on the *dyt1* and *ams* anthers. I found that *DYT1* regulates a broad spectrum of tapetum functions in part by activating at least 32 transcription factor genes, and that *AMS* have both common and distinct downstream genes. Based on the results, I provided new insights into tapetum functions and transcriptional regulation.

2.3 MATERIALS and METHODS

2.3.1 Plant materials, growth conditions, and anther collection

The wild type plants used were *Arabidopsis thaliana*, Landsberg. The *dyt1-1* mutants were described by Dr. Wei Zhang *et al.* and confirmed by genotyping PCR (Zhang *et al.*, 2006). A SALK T-DNA line of *AMS*, Salk-0152147, was identified by PCR and found to be defective in pollen formation. All plants were grown at 22C° under long day photoperiod (16 hours in light and 8 hours of darkness). The anthers at stages 4-7 were dissected and collected from the young floral buds of the plants of 3-4 weeks. Three biological replicates from each genetic background were prepared. Total RNA was extracted by using Plant RNease Mini kit according to the manufacture's manual (Qiagen).

2.3.2 Microarray analysis

For microarray analysis, the *Arabidopsis* ATH1 chips from Affymetrix were used. The procedure was the same that was used in the previous experiments done by Dr. Xiaohong Zhang and Dr. Asela Wijerantne (Wijeratne *et al.*, 2007; Zhang *et al.*, 2005). Probe synthesis, hybridization and signal acquisition were done at the microarray facility at Penn State. The signals from different chips were then normalized, and a scatter plot analysis was done to evaluate the reproducibility of chips. A two-fold cut off was used to select the differentially expressed genes. To analyze the pathways or Gene Ontology (GO) categories that the differentially expressed genes belong to, a MAPMAN analysis was done and the GO annotations were downloaded from the TAIR8 website (Thimm *et al.*, 2004; Usadel *et al.*, 2005).

2.3.3 RT-PCR and Real time PCR

To further validate the microarray results, a subset of differentially expressed genes were selected for RT-PCR and Real time PCR analysis. The cDNAs were synthesized by using the Superscript II reverse transcriptase from Invitrogen according to the manufacture's manual. For Real time PCR, the primers and amplicon were designed for the CyberGreen reaction from STRATAGENE. The actin gene was used as control. Three replicates were used for each gene and the normalized data was used for comparison.

2.4 RESULTS

2.4.1 The transcriptomes of the *dyl1-1* and *ams* anther were uncovered

In order to get reliable transcriptional profiles from the two anther mutants, three biological replicates were used. The reproducibility of the arrays was evaluated by scatter plots with the R value shown. The reproducibility is $R^2=0.9844$, $R^2=0.9667$, and $R^2=0.9783$ for *dyl1-1*, *ams*, and *ler* respectively (figure 2.1). The average of each gene expression from all replicates of the same genetic background was then used to evaluate the expression pattern and the log2 value was used to represent the expression level. With two-fold cut off, I found about 600 genes that were down regulated and 300 genes that

were up regulated in the *dyl1-1* mutant. In the *ams* mutant, about 600 genes were down regulated and about 900 genes were up regulated (figure 2.2).

To further validate the microarray results, I carried out RT-PCR experiments. The cDNA samples were normalized by using the *UBIQUITIN 1 (UBQ1)* gene as an internal control. By using the same RNA samples for microarray analysis, I detected reduced *DYT1* mRNA level in *dyl1-1*, which further confirmed the identity of the anther sample (the *DYT1* probes were not present on the ATH1 chip). For the other 11 selected genes, the expression of each gene was reduced in the *dyl1-1* mutant; this is consistent with the microarray result. These genes included the genes encoding transcription factors (AMS, MYB32, MYB103, AGL40, FLC, two unknown AP2, and homeodomain proteins) and the genes encoding enzymes: pectate lyase (At3g24230), cellulase (At3g26140), pectin esterase (At1g75790), and respiratory burst oxidase (At1g19230) (figure 2.3).

Real time PCR results also support the microarray results mentioned above. The actin gene was used to normalize the data and 3 replicates were used in order to reduce experimental errors. The expression of 9 transcription factor genes is similar to the expression shown by microarray (*AMS*, *MS1*, *MYB103*, *MYB35*, *AP2 gene/At1g44830*, *HD-LeuZipper gene/At4g37790*, *PHD gene/At5g58610*). The following 7 genes encoding enzymes also show similar results: chalcone synthase (At1g02050), caffeoyl-CoA 3-O-methyltransferase (At1g67990), cytochrome P450 (At1g74540), cinnamoyl-CoA reductase family (At1g62940), VANGUARD (At2g47040, pectinesterase), pectate lyase (At3g24230), and 2OG-Fe(II) oxygenase (At3g60260). The Real time PCR result suggests that my microarray results may reflect real expression patterns (figure 2.4).

2.4.2 *DYT1* positively regulates at least 32 transcription factors, among which 14 are also positively regulated by AMS.

Among the genes that were down regulated in the *dyl1-1* anther, at least 32 genes encode different types of transcription factors: 6 MYB, 3 AP2, 3 homeodomain, 2 MADS, 2 TAZ, 2 HD-ZIP, 2 NAC, 2 ZIP, 1 bHLH, 1 Trihelix, 1 ARF, 1 LIM, 1 B-box, 1 Leu Zipper, and 3 unknown (table 2.1 and table 2.2). Some transcription factors are known to

be important for tapetum functions: AMS, MYB103, MYB99, and MYB32. *MYB35* was recently identified as a gene upstream of *AMS* and *MYB103*. ANTHOCYANINLESS2 (ANL2, a homolog of a homeodomain protein, ANL1) regulates anthocyanin biosynthesis. Roles for other genes in anther development are not well characterized. *FLC* has been characterized as a repressor in flowering time control. In this study, it is shown to be downstream of *DYT1* in anther development. Further studies of these transcription factors might lead to the discoveries of new tapetum functions and transcriptional regulation.

I compared the transcription factor genes that were down regulated in *dyt1-1* and *ams*. I found that 14 transcription factor genes including *MYB103*, *AGL40*, 2 *TAZ* and other uncharacterized genes were down regulated in both mutants (table 2.2). This suggests that *DYT1* and *AMS* have overlapping functions in tapetum transcriptome. However, *DYT1* and *AMS* also regulate different transcription factors. 18 transcription factors are positively regulated only by *DYT1* and 11 are positively regulated only by *AMS*. This might reflect the divergent functions of *DYT1* and *AMS* (table 2.1 and table 2.3).

2.4.3 The genes involved in lipid metabolism, cell wall and secondary metabolism are down regulated in *dyt1-1*, whereas heat shock protein genes are up regulated

To understand the biological pathways that are affected in the *dyt1-1* mutant, I categorized the genes down regulated in *dyt1-1* by GO annotations and by biological pathways annotated in MAPMAN. Among the genes that show reduced expression in *dyt1-1*, metabolic genes involved in lipid metabolism, cell wall changes, and secondary metabolism are frequently seen.

The 29 genes involved in lipid metabolisms show reduced expression in *dyt1-1*, and they encode 8 lipid transfer proteins, 15 acyl lipid metabolic enzymes, and 6 lipases (table 2.5). A number of genes in this category have been shown to be anther specific and/or important for male fertility. For example, *MALE STERILE 2/MS2* encodes a fatty acid

reductase and its expression shows dramatic reduction in *dyl1-1* ($\text{Log}_2 \text{dyl1-1/wt} = -4.0$) (table 2.4). Mutations in *MS2* cause male sterility.

In the *dyl1-1* mutant, 45 genes encoding the enzymes involved in cell wall changes show reduced expression (table 2.6). During anther development, the cell wall of meiocytes shows dynamic changes, and these changes are contributed by functions of the tapetum. Among the 45 genes, 19 genes encode cell wall hydrolases that break down the covalent bonds of glycoside structures. They might function in the dissolution of cell wall components such as cellulose and xyloglucan, etc. One β -1,3-glucanase gene exhibited great reduction in expression level ($\text{Log}_2 \text{dyl1-1/wt} = -4.1$) which suggests that the enzyme for callose dissolution might be regulated by *DYL1*. Another group of cell wall enzymes that showed reduced gene expressions in *dyl1-1* are related to pectin modification and degradation, which include 8 polygalacturonases, 4 pectate lyases, 6 pectinesterases and 2 invertases/pectin methylesterase inhibitors. The genes that encode pollen coat proteins were also affected in *dyl1-1*. These proteins include 8 glycine-rich proteins (2 oleosins), 2 arabinogalactan-proteins (AGP), and 3 anther specific proline-rich proteins, which might be synthesized in the tapetum and deposited onto pollen (table 2.8).

Secondary metabolism is regulated by *DYL1* in the tapetum as suggested by microarray analysis. It was found that the expression of 14 genes involved in the biosynthesis of lignin and flavonoids were reduced in *dyl1-1*. These genes encode following enzymes: phenylalanine ammonia lyase (PAL1), chalcone synthase (CHS), caffeoyl-CoA 3-O-methyltransferase (COMT), cinnamoyl-CoA reductase, cinnamyl alcohol dehydrogenase, 4-coumarate--CoA ligase (4-CL), didyhydroflavonol-4-reductase (DFR) (table 2.7). A number of transporter genes also show lower expression in *dyl1-1* (37 genes). These transporters might be important for the transportation of both organic (sugar, peptide, amino acid etc.) and inorganic (Fe, K, Cu etc.) materials in order to support the active metabolisms in the tapetum (table 2.9).

2.4.4 *DYT1* and *AMS* have common and distinct downstream genes

A Venn diagram was made to compare the genes that were differentially regulated in *dyl1-1* and *ams*. I found that *DYT1* and *AMS* positively regulate the expression of about 600 genes respectively, among which 230 genes were positively regulated by both genes. More than 300 genes were only down regulated in *dyl1-1* and more than 300 genes were only down regulated in *ams*. As for the up regulated genes, more than 900 genes were up regulated in *ams* whereas only 300 genes were up regulated in *dyl1-1* (figure 2.2). In *dyl1-1*, a group of heat shock protein genes are up regulated, whereas such a change was not observed in *ams*. A MAPMAN analysis revealed that metabolic pathways were differentially affected in *dyl1-1* and *ams*. It is noticeable that fewer genes involved in secondary metabolisms were affected in *ams* compared to those in *dyl1-1*. A set of plastid genes were up regulated in *ams*; these genes were not up regulated in *dyl1-1* (figure 2.4-2.7). This supports a point that *DYT1* and *AMS* have distinct roles in tapetum development.

2.5 DISCUSSIONS

2.5.1 A transcriptional cascade regulated by *DYT1* involve at least 32 transcription factor genes in the tapetum

Among the genes that are positively regulated by *DYT1*, at least 32 genes encode different types of transcription factors. A plausible scenario is that *DYT1* regulates these transcription factor genes, and each transcription factor then regulates a different set of genes. Clearly, there is a transcriptional cascade regulated by *DYT1*. My microarray results and available expression analyses suggest that there are at least two hierarchical pathways in this cascade: *DYT1-MYB35-AMS-MYB103* and *DYT1-X-MS1-MYB99*. This is supported by following evidence: *TDF1/MYB35* is downstream of *DYT1* and upstream of *AMS* and *MYB103* (Zhu *et al.*, 2008); *AMS* is upstream of *MYB103*; *DYT1* is upstream of *MS1*, and *MS1* might directly regulates *MYB99* (Ito *et al.*, 2007; Yang *et al.*, 2007); I speculate that such a hierarchical transcriptional network might be required for: (1) high

expression level of genes in the tapetum; (2) regulation of broad spectrum of tapetum functions by *DYT1*; (3) timing and coordination of tapetum functions.

It is found that the average expression level of the genes positively regulated by *DYT1* is $2^{7.9}$. This is about 2 fold higher than that of whole genome. High expression level of genes might be required for the active metabolism in the tapetum. I think multiple layers of transcriptional regulation might be important to reach such a high expression level. If each molecule of transcription factor can transcribe multiple copies of mRNA, multiple levels of transcriptional regulation might be able to amplify target transcripts with more fold changes than those by a single level of transcriptional regulation.

Through such a cascade, *DYT1* also regulates a broad spectrum of functions which are carried out by other transcription factors. Because different transcription factors recognize different target genes, the combinations of transcription factors downstream of *DYT1* could potentially regulate more genes than that could be activated only by *DYT1*. For example, one of the downstream genes, *MSI*, encodes a PHD domain transcription factor which is required for pollen wall formation (Ito *et al.*, 2007; Ito *et al.*, 2002; Wilson *et al.*, 2001; Yang *et al.*, 2007). By comparing the microarray results of *dyt1-1* and *ms1*, I found that a set of genes that are down regulated in *dyt1-1* are also down regulated in *ms1*. These genes mainly encode pollen wall and pollen coat proteins including At5g07550 (Gly-rich protein PUTG1), At1g67990 (caffeoyl-CoA 3-O-methyltransferase), At1g68875 (expressed protein), At5g07560 (oleosin-like), and At1g75910 (anther-specific Pro-rich-like protein) (Ito *et al.*, 2007; Yang *et al.*, 2007). This suggests that *DYT1* regulates pollen wall formation in part by regulating *MSI*.

This is also supported by the findings about other known transcription factors such as *AMS*, *MYB32*, *MYB103* and *MYB99*. *AMS* regulates the exine formation of pollen and the programmed cell death in the tapetum; *MYB103* is involved in exine formation (Sorensen *et al.*, 2003); *MYB32* regulates the phenylpropanoid biosynthesis pathway in the tapetum (Preston *et al.*, 2004); The *TAZI* gene might control the programmed cell death in the tapetum as suggested by a study of its homolog in petunia (Kapoor *et al.*,

2002). I speculate that some of the defects in the *dyl1-1* tapetum and microsporocytes might be contributed by the reduction of expression of aforementioned genes. It is possible that in *dyl1-1* there are multiple defects caused by the reduction of these genes so that the meiocytes degenerate right after meiosis II. However, it makes late defects in *dyl1-1* difficult to study. Therefore, conditional knockout of *DYT1* might be needed to further confirm the broad spectrum of the DYT1 functions. This transcriptome analysis also revealed several transcription factors down stream of DYT1, which have not been characterized. They might have novel functions in the tapetum. Further functional studies on these factors will lead to the understandings about the novel tapetum functions regulated by *DYT1*.

A hierarchical structure might be required to coordinate different tapetum functions. A essential function of tapetum is to support the formation of the pollen wall. Because the composition of the pollen wall is stable for a given species, the genes involved in the biosynthesis of pollen wall components should be coordinated. A hierarchical structure has the potential to control the ratio of the speeds in different pathway so that the end products of each pathway are in proportion. Similarly, the coordination might also be important for the assembly of tapetosomes that consists of lipids, oleosins and other proteins at a given ratio (Hernandez-Pinzon *et al.*, 1999; Hsieh *et al.*, 2005; Hsieh *et al.*, 2007; Ting *et al.*, 1998).

Accurate timing of tapetum functions is another important factor for proper microsporogenesis. Previous studies suggest that the timing of tapetum functions is in pace with the development of meiocytes and microspores. In a transcriptional cascade, the transcription factors at low positions in the cascade might function late because their activation requires the time for upstream transcription factors to accumulate. This might be a mechanism for a function to be activated at accurate time. To better understand the coordination and timing in the transcriptional network downstream of *DYT1*, it is necessary to search for the direct targets of each transcription factor. One initiative step is to search for the direct targets of *DYT1*, which is described in Chapter 3.

2.5.2 *DYT1* regulates important functions in the tapetum including lipid metabolism, cell wall modification, secondary metabolism and transporter activities.

The genes involved in lipid metabolism, cell wall modification/degradation, secondary metabolism and transporter activities show reduced expression in the *dyl1-1* mutant. These activities might have important roles in the tapetum to support microsporogenesis and pollen formation.

During microsporogenesis, the meiocytes experience dramatic cell wall changes. A cell wall rich in callose and pectin is formed to enclose the meiocytes and subsequently separate four microspores in the tetrad. After meiosis, this cell wall enclosing the tetrad is dissolved to release the microspores. The tapetum plays an important role in these processes. In the *dyl1-1* mutant, the genes responsible for callose dissolution, pectin modification/degradation, and cell wall hydrolysis show reduced expression. This suggests that one important function of *DYT1* is to activate the genes required for the cell wall modification/degradation during microsporogenesis.

DYT1 also regulates the biosynthesis of pollen coat materials including lipid, flavonoids, oleosin and AGPs (Coimbra *et al.*, 2007; Hernandez-Pinzon *et al.*, 1999; Hsieh *et al.*, 2005; Hsieh *et al.*, 2007; Roberts *et al.*, 1993; Ross *et al.*, 1996; Ruiter *et al.*, 1997; Ting *et al.*, 1998; Wang *et al.*, 1997). PUTG1 is a known glycine-rich protein that is abundant on the pollen coat, and the expression of its gene is regulated by *DYT1* (Ito *et al.*, 2007; Yang *et al.*, 2007). Large number of genes involved in acyl lipid metabolism showed reduced expression in *dyl1-1*. One recent study suggested that flavonoids are important pollen coat components which are stored in tapetosomes before tapetum degeneration (Napoli *et al.*, 1999; Pollak *et al.*, 1993; Preston *et al.*, 2004; van Eldik *et al.*, 1997; Ylstra *et al.*, 1992). As one end product of phenylpropanoid biosynthesis pathway, the flavonoids is secreted to form a screen to protect pollen from UV irradiation (Hsieh *et al.*, 2007; Kapoor *et al.*, 2002). Interestingly, several genes in the phenylpropanoid pathway for flavonoid synthesis are affected in *dyl1-1*, which suggests that *DYT1* also regulates the flavonoids production in the tapetum. Previous studies suggest these materials are

stored in lipid rich organelles such as tapetosomes (Hsieh *et al.*, 2005; Hsieh *et al.*, 2007; Ting *et al.*, 1998; Wu *et al.*, 1999). The regulation of these related pathways by a common transcription factor, *DYT1*, might suggest that coordination among these pathways is needed to assembly normal lipid rich organelles.

Another group of genes affected in *dyl1-1* encode transporters for both organic and inorganic compounds. Their functions might be required to support the biosynthesis and secretion activities in the tapetum. Functional studies on these transporters will allow us to determine their roles in the tapetum. For example, we could identify which transporters are important for uploading materials to the tapetum and which transporters are required for secretion into locule.

2.5.3 Transcriptional regulation by *DYT1* might be partially mediated by complexes of *DYT1* and *AMS*, whereas *AMS* might have distinct functions

The result that the *dyl1-1* and *ams* transcriptomes partially overlap suggests that *DYT1* and *AMS* could regulate common genes. It is possible that the *AMS* expression is reduced one tenth in *dyl1-1* so that some genes that are downstream of *AMS* also show reduced expression in *dyl1-1*. Because bHLH proteins usually form complexes, I tested the interactions between DYT1 and other anther preferential bHLHs including AMS by yeast two hybrid assays (Chapter 3). I found that DYT1 can interact with AMS. It suggests that DYT1 and AMS might regulate common genes by being incorporated into the same complex. To further confirm this possibility, it is necessary to test if DYT1 and AMS could bind to the same promoter *in vivo*. Because *DYT1* is expressed earlier than *AMS*, the activation of AMS at late stages might be able to reinforce the DYT1 functions by regulating common genes. It is also possible that the DYT1-AMS hetero-dimers have different DNA binding specificities from that of homodimers so that DYT1-AMS interaction is required to regulate a specific set of genes. Similar mechanisms might be conserved in monocot. UDT1 and TDR are the rice orthologs of DYT1 and AMS respectively and have similar functions. It will be interesting to learn whether UDT1 and

TDR also have common and different downstream genes (Jung *et al.*, 2005; Li *et al.*, 2006).

However, there are some differences between the transcriptomes of *dyl1-1* and that of *ams*. In *dyl1-1*, a large number of heat shock protein genes are upregulated, which are not up regulated in *ams*. Furthermore, about 300 genes are only positively regulated by *DYT1* but not *AMS*. Because the *ams* mutant is a null allele, these changes should reflect the functions of *DYT1* that are not shared by *AMS*. In another word, there is at least one *AMS*-independent pathway downstream of *DYT1*. *DYT1* might regulate such pathway either by interacting with other proteins that *AMS* does not interact with or by regulating other transcription factors such as *MS1* which is not regulated by *AMS*.

On the other hand, *AMS* also regulate some genes that are not regulated by *DYT1*. A group of plastid genes are down regulated in *ams* whereas they are not in *dyl1-1*. These genes might be related to the biogenesis of elaioplast, a lipid rich organelle which is a storage compartment for lipid and proteins in the tapetum. It is possible that these genes are also regulated by *DYT1* if the residue activity of *AMS* in *dyl1-1* is still enough to activate a set of downstream genes. However, if evidence suggests that the residue expression of *AMS* in *dyl1-1* is not sufficient, I think *AMS* has distinct functions. Recently, it was shown that *AMS* interacts with a SET domain protein which is a negative regulator of transcription. This finding might explain the different functions of *DYT1* and *AMS*. It is also possible that a set of genes are positively regulated by *AMS* but negatively regulated by *DYT1*. In *dyl1-1*, although the positive regulation by *AMS* is reduced, the negative regulation by *DYT1* is also reduced; therefore the expression of these genes will be close to normal. However, in *ams*, *DYT1* still represses these genes so that the expressions will be reduced. Some genes may need to be repressed by *DYT1* at early stages and to be activated only after *AMS* is expressed. Under regulation by both *DYT1* and *AMS*, some genes can only be transcribed at a high level after the *AMS* activity can overcome the repression by *DYT1*. To understand the common and distinct regulation by *DYT1* and *AMS* *in vivo*, it is necessary to search the direct targets of both transcription factors by performing chromatin immuno precipitation (ChIP) experiments.

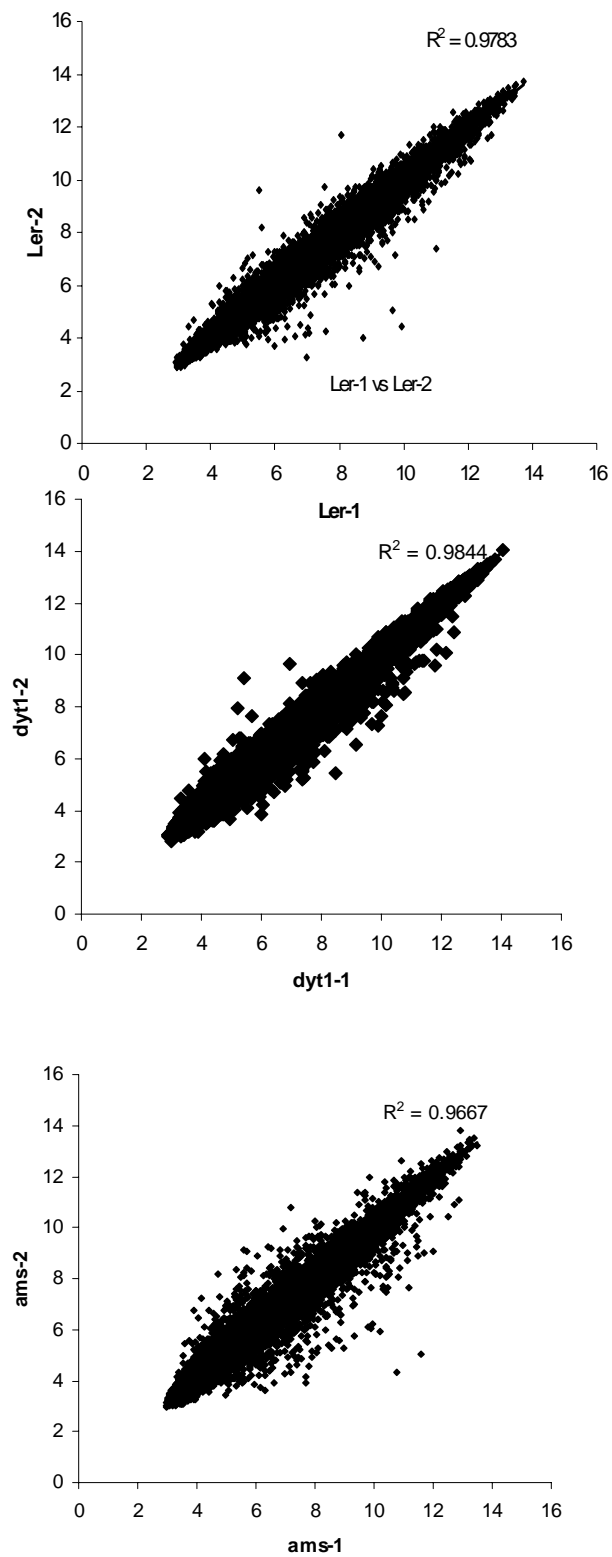


Figure 2.1 Scatter plots to evaluate the reproducibility of microarray replica.

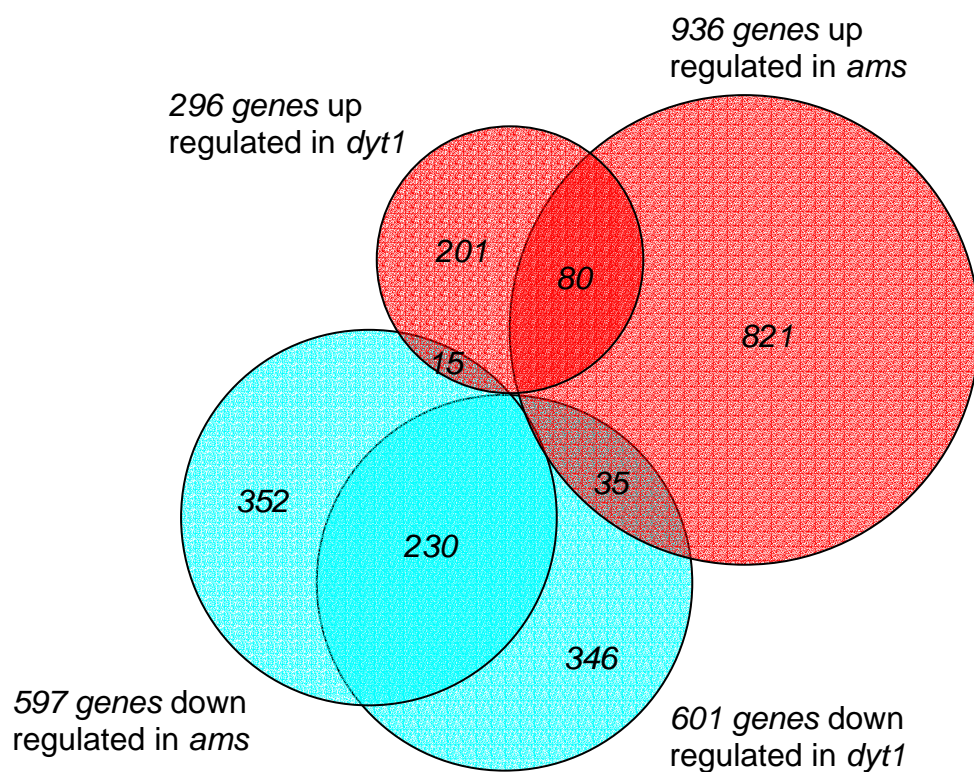


Figure 2.2 The overlap between the group of genes that are up or down-regulated in *dyt1* or *ams*. A two-fold change was used as the cut off to select differentially expressed genes.

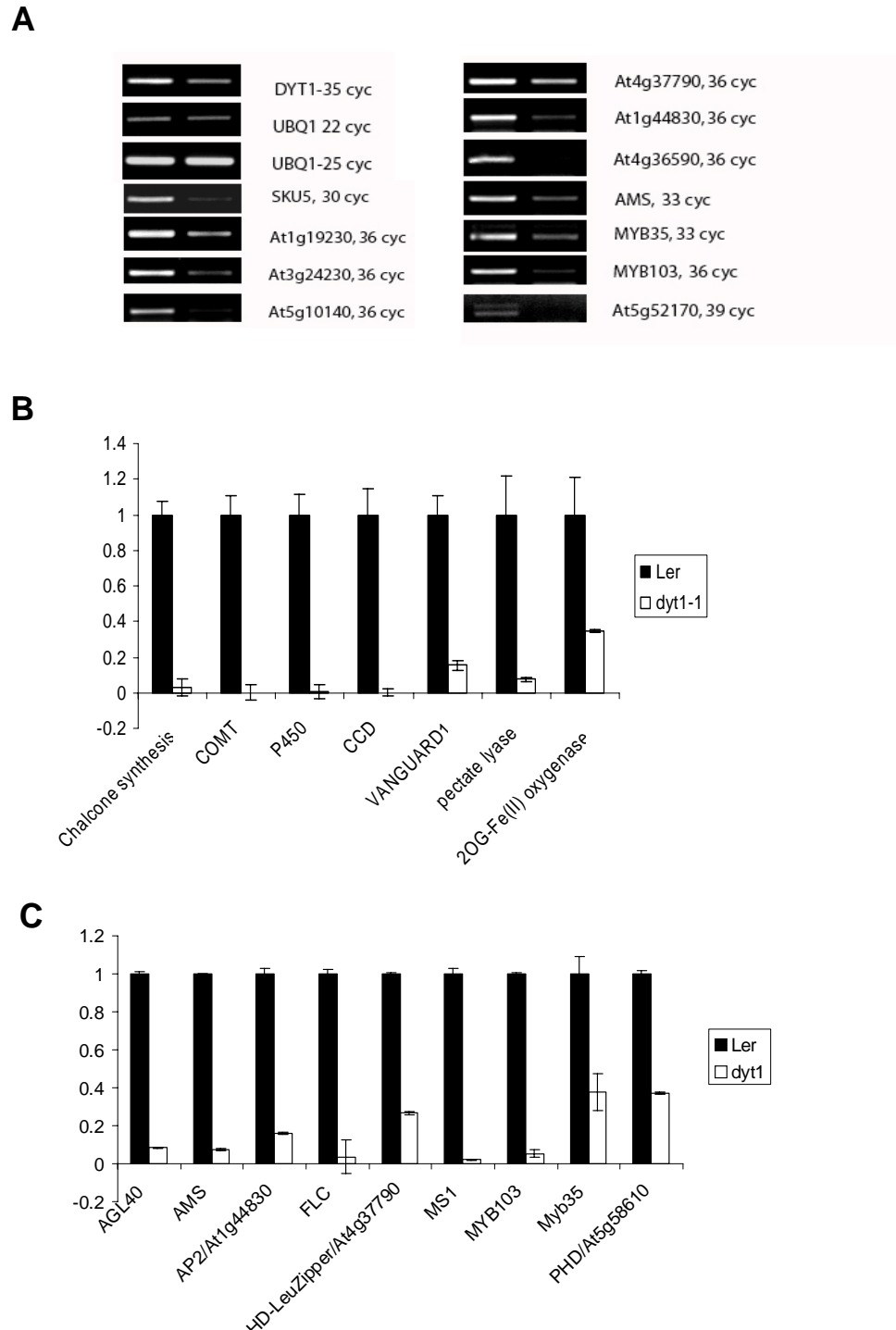


Figure 2.3: RT-PCR and Real time PCR analysis of the genes down-regulated in *dtl1-1*: A. RT-PCR analysis: UBQ1 was used to normalize the cDNA quantity; B. Real time PCR analysis of 7 enzyme genes down-regulated in *dtl1-1*; Three replica were done and the fold change between the expression in *ler* and *dtl1-1* is shown for each genes. C. Real time PCR analysis of 9 transcription factor genes down regulated in *dtl1-1*.

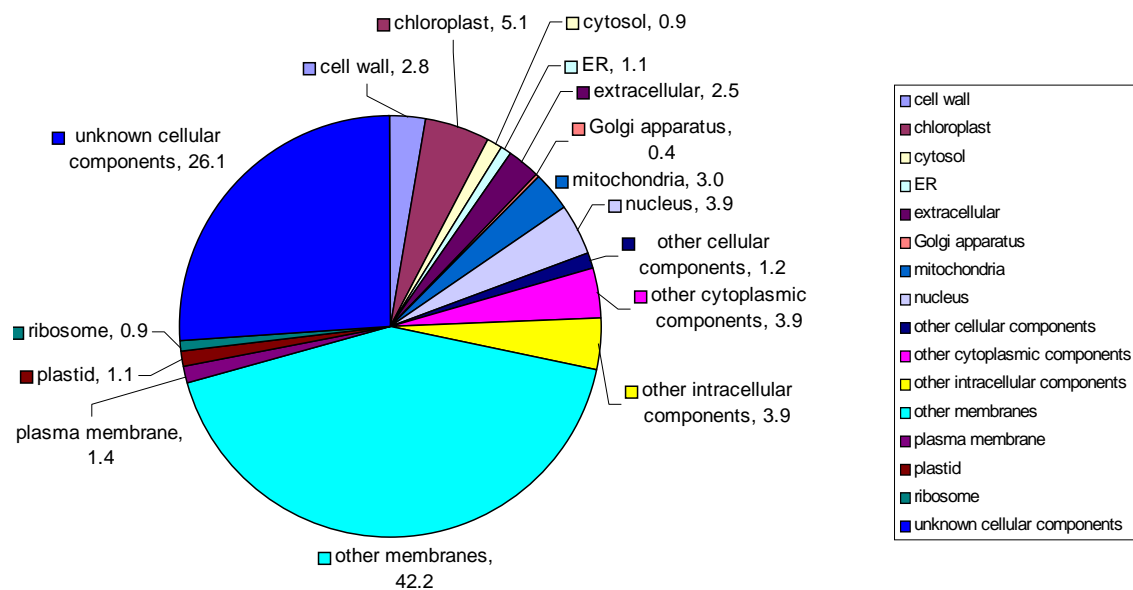


Figure 2.4: Cellular component analysis of the genes down-regulated in *dyt1-1*. The GO annotation was analyzed on TAIR8 website. Percentage of each category is shown.

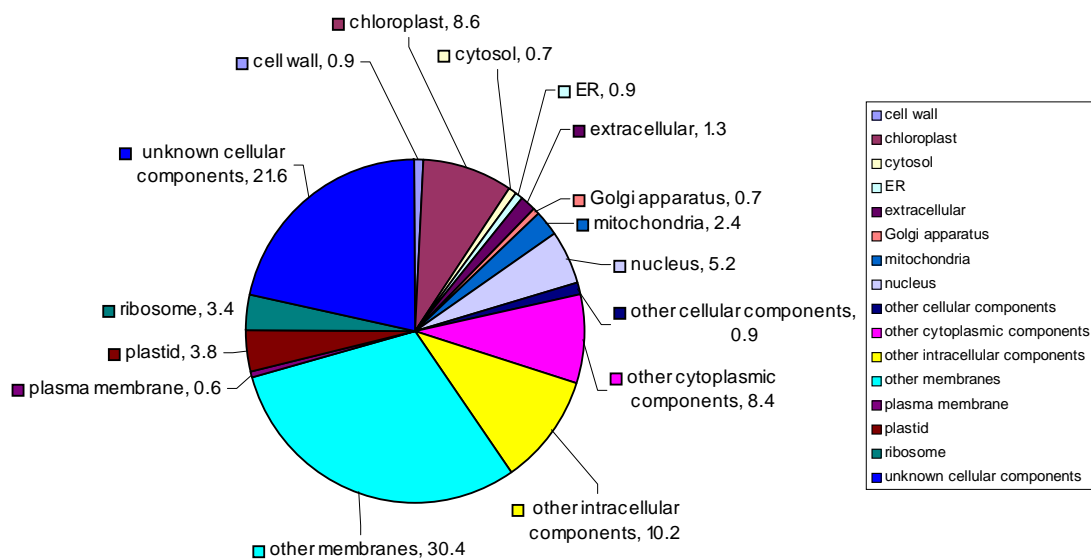


Figure 2.5: Cellular component analysis of the genes down-regulated in *ams*. The GO annotation was analyzed on TAIR8 website. Percentage of each category is shown.

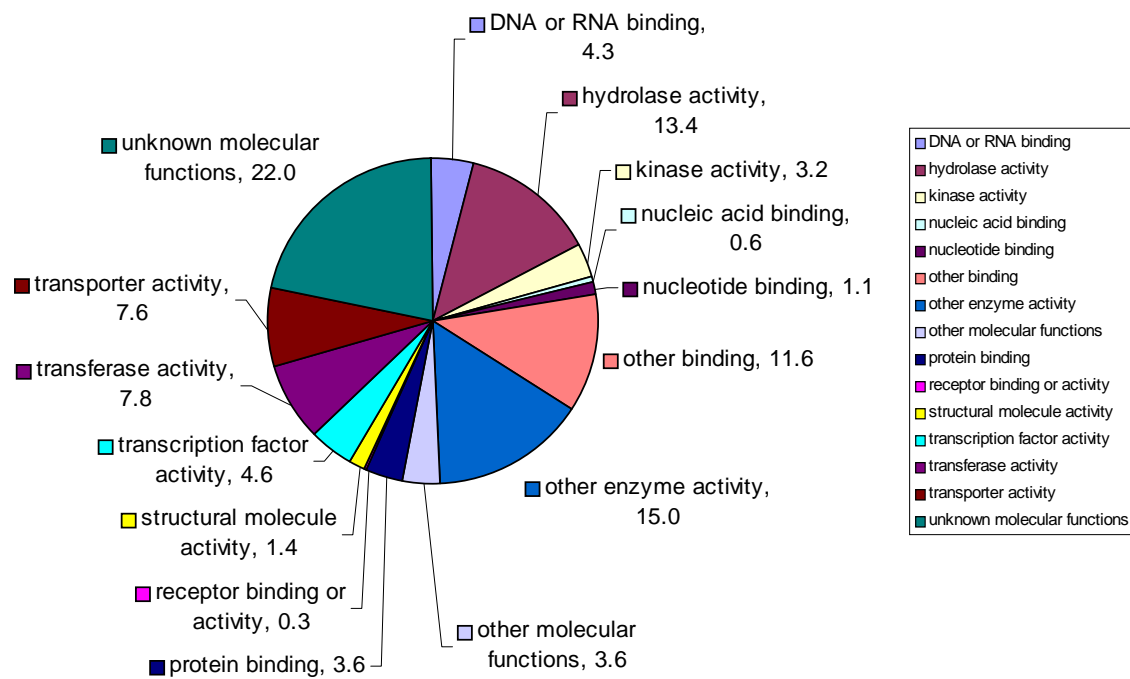


Figure 2.6 Molecular function analysis of the genes down-regulated in *dyt1-1*. The GO annotation was analyzed on TAIR8 website. Percentage of each category is shown.

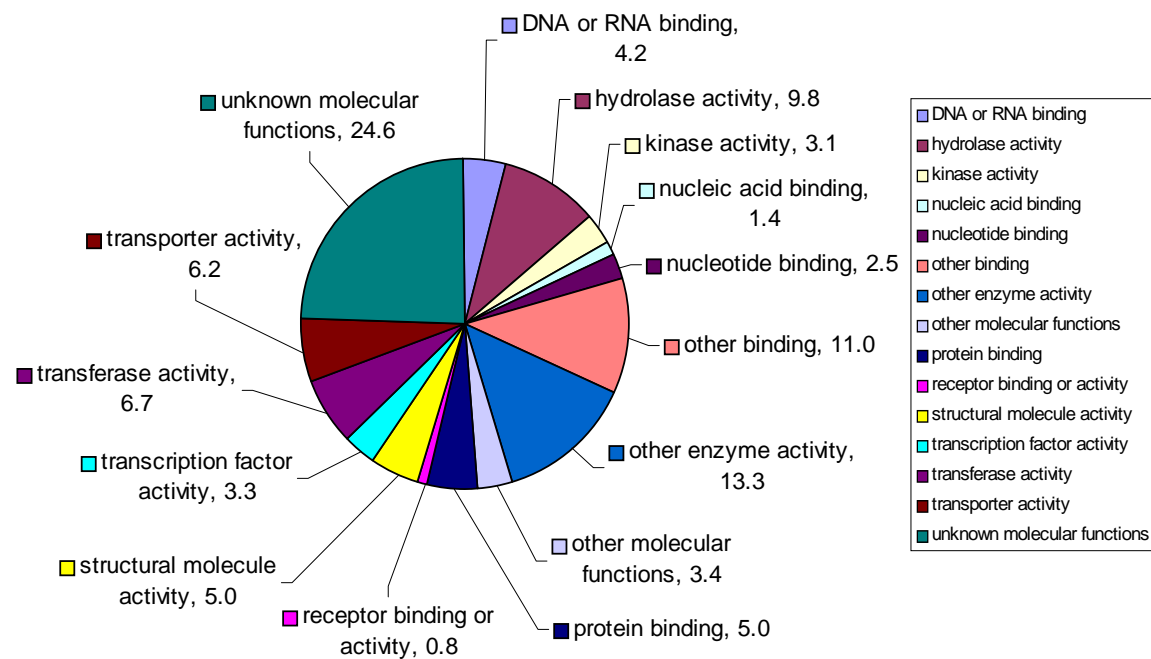


Figure 2.7 Molecular function analysis of the genes down-regulated in *ams*. The GO annotation was analyzed on TAIR8 website. Percentage of each category is shown.

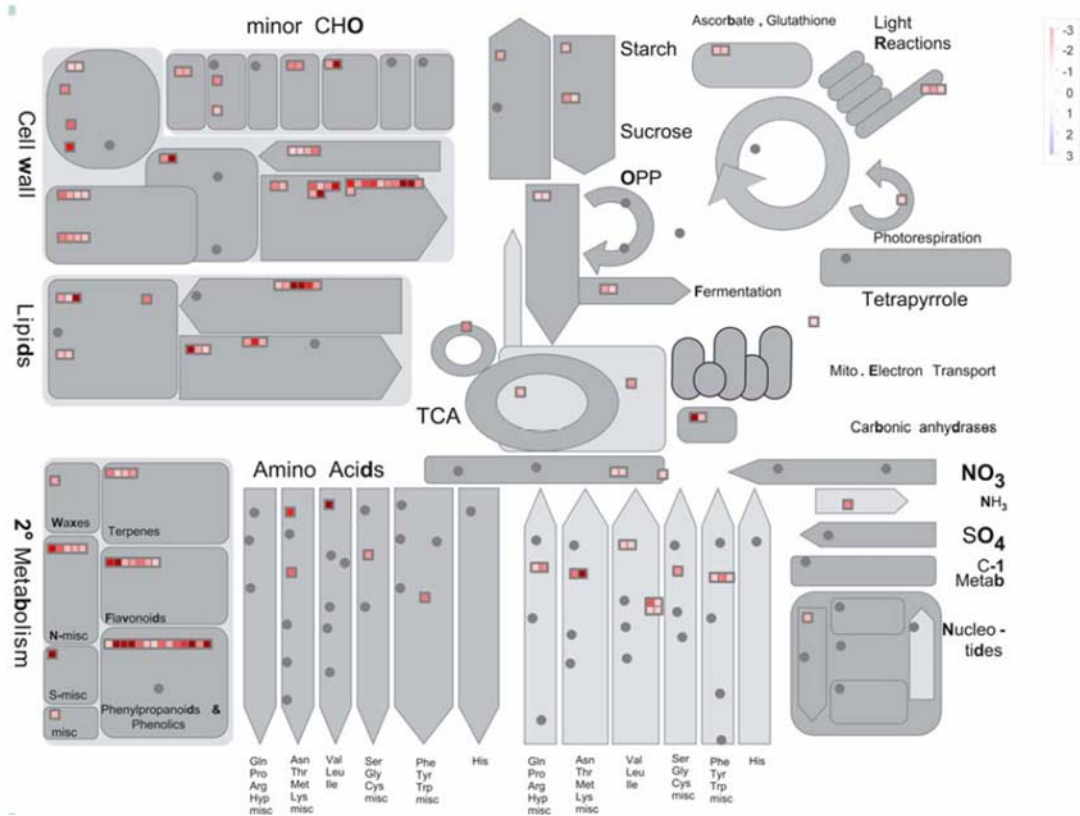


Figure 2.8: A MAPMAN visualization of the genes down-regulated in *dyl1-1*. The metabolism is shown as big box. Within each box, each small square box represents one gene. The shade of red color reflects the fold changes of gene expression ($\log_2 dyl1-1/ler$).

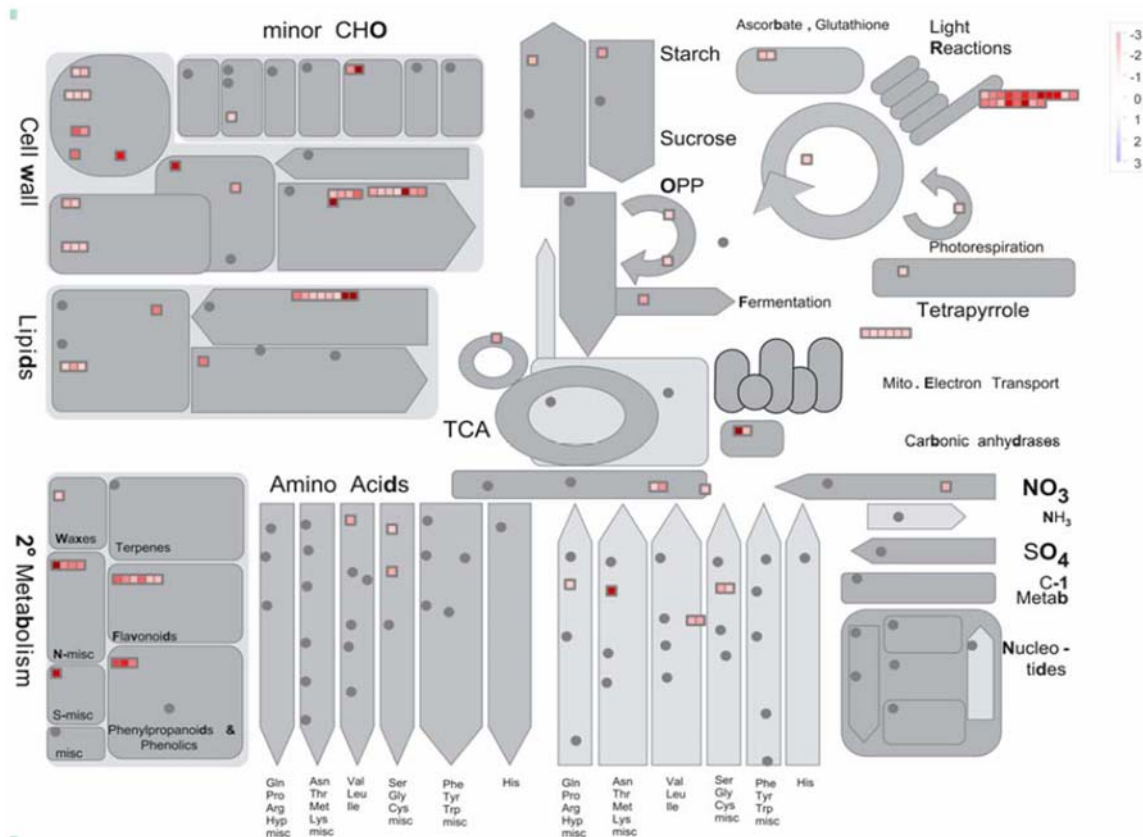


Figure 2.9: A MAPMAN visualization of the genes down-regulated in *ams*. The metabolism is shown as big box. Within each box, each small square box represents one gene. The shade of red color reflect the fold changes of gene expression ($\log_2 \text{ams/ler}$).

Table 2.1 List of transcription factor genes down-regulated in the *dyl1-1* and *ams* mutant anther revealed by ATH1 microarray analysis

AGI	ams	dyl1	Ler	dyl1-Ler	ams-Ler	Domains
At1g19230	4.8	4.9	8.2	-3.3	-3.34865	AP2
At3g48360	6.8	6.6	10.0	-3.3	-3.18194	TAZ
At1g49490	4.6	4.7	6.6	-1.9	-1.9732	Trihelix
At5g10140	6.3	5.1	8.1	-2.9	-1.7873	FLC
At5g56110	5.7	5.8	7.5	-1.6	-1.77759	Myb103
At1g34310	6.8	7.3	8.6	-1.2	-1.77045	ARF1
At5g52170	4.4	4.7	6.2	-1.4	-1.7632	ANL2
At4g36590	4.1	3.8	5.8	-2.0	-1.74652	AGL40
At2g27990	4.7	5.2	6.4	-1.2	-1.72387	Homeodomain
At3g61230	4.5	4.5	6.2	-1.7	-1.6795	LIM
At4g37610	9.3	8.6	10.6	-2.0	-1.3302	TAZ
At1g44830	4.6	4.5	5.9	-1.4	-1.32019	AP2
At2g22800	7.5	7.3	8.7	-1.5	-1.25055	HD-ZIP9
At4g39070	6.2	5.8	7.2	-1.4	-1.02724	CO like

Table 2.2 List of 18 transcription factor genes down-regulated in the *dyt1-1* mutant anther

AGI	ams	dyt1	Ler	dyt1-Ler	ams-Ler	Domain
At3g28470	8.5	5.1	8.6	-3.4	0.0	MYB35
At5g62320	8.3	5.3	8.6	-3.3	-0.3	MYB99
At2g32460	6.7	4.8	7.1	-2.3	-0.4	MYB101
At1g61110	9.8	5.7	7.9	-2.2	1.9	NAC
At2g18550	9.0	7.0	9.1	-2.1	-0.1	Homeodomain
At5g49450	8.3	7.2	8.7	-1.5	-0.4	AtbZIP1
At5g53550	6.7	5.7	7.2	-1.5	-0.5	TFs
At1g69490	9.6	8.3	9.8	-1.5	-0.2	NAP/NAC
At4g34990	7.9	7.1	8.5	-1.4	-0.6	MYB32
At4g10920	7.4	6.6	8.0	-1.4	-0.5	unknown
At1g65620	8.3	6.9	8.3	-1.4	0.0	AS2
At1g68840	6.8	5.5	6.8	-1.3	0.0	AP2
At4g37790	7.3	6.8	8.1	-1.3	-0.8	HD-ZIP22
At3g15500	7.0	6.2	7.4	-1.3	-0.4	NAC
At5g46640	7.3	6.1	7.3	-1.2	0.0	TFs
At1g48000	5.7	5.0	6.1	-1.1	-0.4	MYB112
At1g69570	6.7	4.8	5.8	-1.0	0.9	Dof
At1g69310	5.0	4.4	5.5	-1.0	-0.5	WRKY

Table 2.3 List of transcription factor genes down-regulated in the *ams* mutant anther

AGI	ams	dvt1	Ler	dvt1-Ler	ams-Ler	Domain
At5g62165	5.4	8.3	7.9	0.3	-2.6	MADS
At5g58610	5.9	6.8	7.7	-1.0	-1.9	PHD
At2g23740	5.6	7.3	7.2	0.1	-1.6	Zinc finger
At3g60580	5.6	6.3	7.2	-0.8	-1.5	Zinc finger
At1g77300	5.2	6.9	6.6	0.3	-1.4	SET domain
At3g63070	6.3	7.9	7.7	0.2	-1.4	HUA2/PWWP
At4g09980	5.1	7.1	6.5	0.6	-1.3	DNA binding
At5g12400	7.1	8.3	8.3	0.0	-1.2	PHD
At2g31210	6.0	7.1	7.1	-0.1	-1.1	bHLH
At5g57660	7.4	7.6	8.5	-0.9	-1.1	CO like/B-box

Table 2.4 Known genes that are down-regulated in *dyl1-1*

AGI	ams	dyl	WTAnther	dyl- WTA	ams- WTA	
At3g07450	7.6	9.0	12.8	-3.8	-5.2	5B
At4g20420	8.6	8.7	11.7	-3.0	-3.1	A3
At4g14080	8.7	8.3	12.5	-4.1	-3.8	A6
At5g07230	10.3	8.9	13.0	-4.1	-2.7	A9/LTP
At5g62080	8.9	10.1	13.0	-3.0	-4.2	LTP3
At3g42960	10.2	7.8	11.9	-4.1	-1.7	ATA1
At3g15400	12.2	8.8	12.1	-3.4	0.1	ATA20
At4g28395	10.1	5.5	9.6	-4.1	0.5	ATA7/LTP
At3g11980	10.0	6.5	10.6	-4.0	-0.6	MS2/ Fatty Acid Reductase
At2g47040	5.9	4.6	6.4	-1.8	-0.5	VANGUARD
At5g07550	9.4	4.3	9.0	-4.7	0.4	PUTG1

Table 2.5 Acyl lipid metabolism genes that are down-regulated in *dyl1-1*

AGI	ams	dyl	WTAnther	dyl-WTA	ams-WTA	
At1g71160	9.2	6.6	10.2	-3.6	-1.0	Acyl lipid
At1g06520	9.2	6.5	9.7	-3.3	-0.5	Acyl lipid
At4g14815	9.9	5.7	8.9	-3.2	1.0	Acyl lipid
At3g52160	9.0	5.0	8.2	-3.2	0.8	Acyl lipid
At5g08030	6.4	3.7	6.0	-2.3	0.4	Acyl lipid
At5g49070	6.9	3.9	6.2	-2.3	0.7	Acyl lipid
At5g14980	6.7	4.9	6.8	-2.0	-0.2	Acyl lipid
At1g03090	9.6	9.0	10.9	-1.9	-1.3	Acyl lipid
At1g18280	7.3	5.4	7.0	-1.7	0.2	Acyl lipid
At5g60500	6.2	3.7	5.3	-1.6	0.9	Acyl lipid
At4g38530	7.5	5.5	7.0	-1.5	0.5	Acyl lipid
At4g28130	5.3	4.2	5.7	-1.5	-0.4	Acyl lipid
At5g55320	5.3	5.0	6.4	-1.5	-1.1	Acyl lipid
At3g23790	7.7	7.6	8.8	-1.2	-1.1	Acyl lipid
At3g11980	10.0	6.5	10.6	-4.0	-0.6	MS2/ Fatty Acid Reductase
At1g06250	5.8	4.6	7.6	-3.0	-1.8	Lipase
At3g48460	10.4	8.0	10.8	-2.8	-0.4	Lipase
At1g75920	7.0	4.7	6.6	-1.9	0.4	Lipase
At4g18550	8.4	5.6	7.1	-1.5	1.3	Lipase
At1g74460	8.9	8.0	9.3	-1.3	-0.4	Lipase
At3g62860	8.1	6.4	7.6	-1.2	0.5	Lipase
At4g10850	7.8	5.6	8.6	-3.0	-0.8	MTN3
At3g48740	6.8	5.5	6.9	-1.4	-0.1	MTN3
At5g52160	9.1	5.6	10.7	-5.2	-1.7	LTP
At1g66850	10.8	7.5	11.8	-4.4	-1.0	LTP
At3g52130	9.6	8.9	12.9	-4.1	-3.3	LTP
At1g36150	9.7	6.1	9.2	-3.0	0.5	LTP
At5g59310	6.6	5.9	8.7	-2.7	-2.1	LTP
At4g08670	7.4	4.4	6.3	-1.9	1.1	LTP
At5g59320	6.8	6.4	7.7	-1.2	-0.9	LTP
At2g48130	5.4	5.3	6.5	-1.2	-1.1	LTP

Table 2.6 Genes for cell wall modification/degradation down-regulated in *dyl1-1*

AGI	ams	dyl	WT Anther	dyl- WTA	ams- WTA	annotation
At3g23770	7.1	5.8	9.9	-4.1	-2.7	b-1,3-glucanase
At5g56870	6.6	6.8	10.1	-3.2	-3.4	b-galactosidase
At4g26140	6.9	6.6	8.2	-1.6	-1.3	b-galactosidase
At1g75940	10.9	3.9	8.6	-4.8	2.3	b-glucosidase
At3g60140	4.1	3.7	5.9	-2.1	-1.8	b-glucosidase
At5g09730	6.3	5.6	8.2	-2.7	-1.9	b-xylosidase
At5g49360	8.3	7.8	9.8	-2.0	-1.5	Xylosidase
At3g25050	5.9	4.5	5.9	-1.5	0.0	XTH3
At3g48690	9.8	9.3	11.1	-1.8	-1.3	Carbohydrate easterase
At3g26140	6.4	4.3	6.0	-1.7	0.4	cellulase
At5g07830	9.8	8.3	10.5	-2.2	-0.7	Glycoside hydrolase
At1g02640	10.2	9.4	11.4	-1.9	-1.2	Glycoside hydrolase
At3g57520	11.5	10.9	12.3	-1.5	-0.8	Glycoside hydrolase
At5g20250	11.7	11.0	12.4	-1.4	-0.7	Glycoside hydrolase
At5g57550	5.0	4.8	6.1	-1.3	-1.1	Glycoside hydrolase
At4g24040	9.0	8.8	10.1	-1.3	-1.1	Glycoside hydrolase
At1g77410	8.5	8.1	9.3	-1.3	-0.8	Glycoside hydrolase
At4g25000	7.8	5.8	7.0	-1.2	0.8	Glycoside hydrolase
At5g20710	8.4	6.1	7.5	-1.4	0.9	Glycosyl hydrolase
At5g25820	5.0	4.9	6.1	-1.3	-1.1	Glycosyltransferase
At1g33430	11.1	9.1	12.7	-3.7	-1.6	Glycosyltransferase
At1g22015	7.3	7.0	10.0	-3.0	-2.7	Glycosyltransferase
At5g57500	5.2	4.0	6.0	-2.1	-0.9	Glycosyltransferase
At1g14110	6.6	4.4	6.0	-1.6	0.6	Glycosyltransferase
At4g15750	4.3	4.6	6.4	-1.8	-2.1	invertase/pection methylesterase inhibitor
At5g50030	5.1	4.7	6.5	-1.8	-1.4	invertase
At4g22080	6.9	4.2	8.5	-4.3	-1.5	Pectate Lyase
At3g24230	4.3	4.0	7.6	-3.6	-3.2	Pectate Lyase
At1g11920	4.8	4.2	5.7	-1.5	-0.9	Pectate Lyase
At1g10770	5.5	4.7	6.3	-1.6	-0.8	Pectin easterase
At2g47050	5.0	4.6	6.0	-1.5	-1.0	Pectin easterase
At4g28090	5.4	4.9	6.4	-1.4	-0.9	Pectin easterase
At5g07430	4.5	4.0	5.4	-1.4	-0.9	pectin methylesterase
At1g75790	8.3	6.4	10.2	-3.8	-1.9	pectineasterase
At3g01270	5.2	5.0	6.2	-1.2	-1.1	petate lyase
At1g02790	4.7	3.8	6.2	-2.3	-1.5	Polygalacturonase
At1g60390	8.2	7.2	9.4	-2.2	-1.2	Polygalacturonase
At1g23760	6.8	5.8	7.9	-2.1	-1.2	Polygalacturonase
At3g07850	5.8	3.9	5.7	-1.8	0.2	Polygalacturonase
At1g32170	6.2	5.6	7.4	-1.7	-1.2	Polygalacturonase
At5g17200	4.5	3.9	5.5	-1.6	-1.0	Polygalacturonase
At3g07820	4.7	4.0	5.6	-1.6	-0.9	Polygalacturonase

At5g39910	5.0	5.2	6.6	-1.4	-1.7	Polygalacturonase
At2g47040	5.9	4.6	6.4	-1.8	-0.5	VANGUARD

Table 2.7 Genes involved in secondary metabolism down-regulated in *dyt1-1*

AGI	dyt	ams	WT Anther	dyt- WTA	ams- WTA	
At1g62940	7.7	10.6	11.6	-3.9	-1.0	4CL
At3g21230	5.5	5.9	7.7	-2.2	-1.8	4CL
At1g09500	5.2	6.8	6.6	-1.4	0.2	CAD
At1g76470	5.7	8.0	10.3	-4.5	-2.3	CCR
At4g30470	7.8	9.6	9.7	-1.9	-0.1	CCR
At4g34850	8.2	9.8	11.4	-3.3	-1.6	CHS-like
At1g02050	9.0	9.6	11.6	-2.6	-1.9	CHS-like
At1g67990	4.6	10.3	9.5	-4.9	0.8	COMT
At1g68540	8.7	9.4	11.5	-2.8	-2.1	DFR
At4g35420	10.4	10.0	11.9	-1.5	-1.9	DFR/anthocyanin
At5g54010	4.2	6.1	6.0	-1.8	0.0	Flavonol
At3g10340	5.3	7.4	7.3	-2.0	0.1	PAL
At2g37040	8.0	9.0	9.3	-1.4	-0.3	PAL1
At2g41290	6.6	9.9	8.5	-1.9	1.4	strictosidine synthase

Table 2.8 Genes encoding pollen wall or coat proteins down-regulated in *dyt1-1*

AGI	ams	dyt	WTAnther	dyt-WTA	ams-WTA	Annotation
At1g75910	8.0	4.6	7.3	-2.7	0.8	APG
At1g20120	8.6	4.4	7.1	-2.7	1.5	APG
At1g75930	7.7	4.5	7.0	-2.5	0.7	APG
At3g57690	8.7	7.0	8.7	-1.7	0.0	AGP23
At5g14380	5.1	4.3	5.7	-1.3	-0.6	AGP6
At3g63100	9.7	6.3	8.1	-1.8	1.6	Glycin rich
At2g05540	11.9	10.0	11.8	-1.8	0.1	Glycin rich
At5g07520	5.8	3.9	5.7	-1.8	0.2	Glycin rich/Oleosin
At5g07530	10.0	6.2	7.3	-1.1	2.7	glycine rich
At5g07560	7.5	4.7	7.5	-2.8	0.0	Oleosin
At5g45880	5.9	5.5	7.1	-1.7	-1.3	Oleosin
At5g38760	5.7	3.8	5.7	-1.9	0.0	Pollen coat
At5g07550	9.4	4.3	9.0	-4.7	0.4	PUTG1
At3g02480	5.1	5.1	7.0	-2.0	-2.0	Pollen coat

Table 2.9 Transporter genes that are down-regulated in *dyt1-1*

AGI	ams	dyt	WTAnther	dyt-WTA	ams-WTA	
At5g63850	9.2	5.4	7.3	-1.9	1.8	amino acid
At3g13220	7.9	6.3	10.0	-3.8	-2.2	ABC
At2g39350	10.0	8.3	10.5	-2.2	-0.5	ABC
At4g30110	8.5	6.1	8.2	-2.2	0.3	Cd
At5g52340	7.4	6.2	7.6	-1.4	-0.2	exocyt
At3g13400	5.7	4.4	6.3	-1.9	-0.7	Fe
At4g19690	5.4	3.8	5.2	-1.3	0.3	Fe
At1g24400	7.9	5.2	7.2	-2.0	0.7	Lys/His
At1g68570	8.6	6.0	8.6	-2.5	0.1	Peptide
At1g32450	6.8	4.9	7.2	-2.3	-0.4	Peptide
At1g07340	9.5	6.4	9.2	-2.7	0.4	Sugar
At4g04760	6.8	4.4	6.8	-2.4	0.0	Sugar
At2g20670	10.0	8.7	11.1	-2.3	-1.1	Sugar
At2g47160	7.3	7.5	8.7	-1.2	-1.3	anion
At3g25260	7.3	6.0	9.7	-3.7	-2.4	Nitrate
At5g46050	5.9	5.6	8.7	-3.2	-2.8	Peptide
At1g44800	7.5	5.7	8.6	-3.0	-1.1	Drug/Metabolite
At1g30840	5.8	5.5	8.0	-2.5	-2.2	Drug/Metabolite
At2g26690	7.2	7.1	9.5	-2.3	-2.3	Nitrate/Peptide/Organic
At5g25400	4.3	4.8	7.0	-2.2	-2.7	Organic
At5g43340	9.4	5.8	8.0	-2.2	1.4	Phosphate
At1g77210	8.0	5.8	7.9	-2.1	0.1	Sugar
At5g28470	5.9	4.4	6.5	-2.1	-0.6	Peptide
At4g21120	6.6	6.0	8.0	-2.0	-1.4	amino acid
At5g53510	7.9	5.4	7.2	-1.8	0.7	Peptide
At1g51500	8.4	6.8	8.5	-1.7	-0.2	ABC
At1g11260	9.8	9.9	11.6	-1.7	-1.7	Sugar
At5g59040	6.1	5.1	6.8	-1.7	-0.6	Cu
At3g45060	6.4	5.5	7.2	-1.7	-0.8	Nitrate
At5g47560	5.6	5.0	6.7	-1.7	-1.0	Na ₂ CO ₃
At1g61800	6.2	6.0	7.4	-1.4	-1.2	Drug/Metabolite
At4g28040	6.4	5.3	6.6	-1.4	-0.3	Drug/Metabolite
At1g09930	7.9	5.4	6.8	-1.3	1.2	Peptide
At2g30070	7.6	7.3	8.5	-1.3	-0.9	K
At5g26220	7.1	4.5	5.7	-1.2	1.4	Cation
At1g27080	6.9	5.7	6.8	-1.2	0.1	Nitrate
At5g26340	6.4	5.1	6.2	-1.1	0.2	Sugar

2.6 References

- Balasubramanian, S. and Schneitz, K.** (2000) NOZZLE regulates proximal-distal pattern formation, cell proliferation and early sporogenesis during ovule development in *Arabidopsis thaliana*. *Development*, **127**, 4227-4238.
- Coimbra, S., Almeida, J., Junqueira, V., Costa, M.L. and Pereira, L.G.** (2007) Arabinogalactan proteins as molecular markers in *Arabidopsis thaliana* sexual reproduction. *J Exp Bot*, **58**, 4027-4035.
- Hernandez-Pinzon, I., Ross, J.H., Barnes, K.A., Damant, A.P. and Murphy, D.J.** (1999) Composition and role of tapetal lipid bodies in the biogenesis of the pollen coat of *Brassica napus*. *Planta*, **208**, 588-598.
- Hsieh, K. and Huang, A.H.** (2005) Lipid-rich tapetosomes in *Brassica* tapetum are composed of oleosin-coated oil droplets and vesicles, both assembled in and then detached from the endoplasmic reticulum. *Plant J*, **43**, 889-899.
- Hsieh, K. and Huang, A.H.** (2007) Tapetosomes in *Brassica* tapetum accumulate endoplasmic reticulum-derived flavonoids and alkanes for delivery to the pollen surface. *Plant Cell*, **19**, 582-596.
- Ito, T., Nagata, N., Yoshida, Y., Ohme-Takagi, M., Ma, H. and Shinozaki, K.** (2007) *Arabidopsis* MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development. *Plant Cell*, **19**, 3549-3562.
- Ito, T. and Shinozaki, K.** (2002) The MALE STERILITY1 gene of *Arabidopsis*, encoding a nuclear protein with a PHD-finger motif, is expressed in tapetal cells and is required for pollen maturation. *Plant Cell Physiol*, **43**, 1285-1292.
- Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J.L. and Meyerowitz, E.M.** (2004) The homeotic protein AGAMOUS controls microsporogenesis by regulation of SPOROCTELESS. *Nature*, **430**, 356-360.
- Jung, K.H., Han, M.J., Lee, Y.S., Kim, Y.W., Hwang, I., Kim, M.J., Kim, Y.K., Nahm, B.H. and An, G.** (2005) Rice Undeveloped Tapetum1 is a major regulator of early tapetum development. *Plant Cell*, **17**, 2705-2722.

- Kapoor, S., Kobayashi, A. and Takatsuji, H.** (2002) Silencing of the tapetum-specific zinc finger gene TAZ1 causes premature degeneration of tapetum and pollen abortion in petunia. *Plant Cell*, **14**, 2353-2367.
- Li, N., Zhang, D.S., Liu, H.S., Yin, C.S., Li, X.X., Liang, W.Q., Yuan, Z., Xu, B., Chu, H.W., Wang, J., Wen, T.Q., Huang, H., Luo, D., Ma, H. and Zhang, D.B.** (2006) The rice tapetum degeneration retardation gene is required for tapetum degradation and anther development. *Plant Cell*, **18**, 2999-3014.
- Ma, H.** (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annu Rev Plant Biol*, **56**, 393-434.
- Napoli, C.A., Fahy, D., Wang, H.Y. and Taylor, L.P.** (1999) white anther: A petunia mutant that abolishes pollen flavonol accumulation, induces male sterility, and is complemented by a chalcone synthase transgene. *Plant Physiol*, **120**, 615-622.
- Pollak, P.E., Vogt, T., Mo, Y. and Taylor, L.P.** (1993) Chalcone Synthase and Flavonol Accumulation in Stigmas and Anthers of *Petunia hybrida*. *Plant Physiol*, **102**, 925-932.
- Preston, J., Wheeler, J., Heazlewood, J., Li, S.F. and Parish, R.W.** (2004) AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. *Plant J*, **40**, 979-995.
- Roberts, M.R., Hodge, R., Ross, J.H., Sorensen, A., Murphy, D.J., Draper, J. and Scott, R.** (1993) Characterization of a new class of oleosins suggests a male gametophyte-specific lipid storage pathway. *Plant J*, **3**, 629-636.
- Ross, J.H. and Murphy, D.J.** (1996) Characterization of anther-expressed genes encoding a major class of extracellular oleosin-like proteins in the pollen coat of Brassicaceae. *Plant J*, **9**, 625-637.
- Ruiter, R.K., Van Eldik, G.J., Van Herpen, R.M., Schrauwen, J.A. and Wullems, G.J.** (1997) Characterization of oleosins in the pollen coat of *Brassica oleracea*. *Plant Cell*, **9**, 1621-1631.
- Sorensen, A.M., Krober, S., Unte, U.S., Huijser, P., Dekker, K. and Saedler, H.** (2003) The *Arabidopsis* ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. *Plant J*, **33**, 413-423.

- Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y. and Stitt, M.** (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J*, **37**, 914-939.
- Ting, J.T., Wu, S.S., Ratnayake, C. and Huang, A.H.** (1998) Constituents of the tapetosomes and elaioplasts in *Brassica campestris* tapetum and their degradation and retention during microsporogenesis. *Plant J*, **16**, 541-551.
- Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O.E., Palacios-Rojas, N., Selbig, J., Hannemann, J., Piques, M.C., Steinhauser, D., Scheible, W.R., Gibon, Y., Morcuende, R., Weicht, D., Meyer, S. and Stitt, M.** (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. *Plant Physiol*, **138**, 1195-1204.
- van Eldik, G.J., Reijnen, W.H., Ruiter, R.K., van Herpen, M.M., Schrauwen, J.A. and Wullems, G.J.** (1997) Regulation of flavonol biosynthesis during anther and pistil development, and during pollen tube growth in *Solanum tuberosum*. *Plant J*, **11**, 105-113.
- Wang, T.W., Balsamo, R.A., Ratnayake, C., Platt, K.A., Ting, J.T. and Huang, A.H.** (1997) Identification, subcellular localization, and developmental studies of oleosins in the anther of *Brassica napus*. *Plant J*, **11**, 475-487.
- Wijeratne, A.J., Zhang, W., Sun, Y., Liu, W., Albert, R., Zheng, Z., Oppenheimer, D.G., Zhao, D. and Ma, H.** (2007) Differential gene expression in *Arabidopsis* wild-type and mutant anthers: insights into anther cell differentiation and regulatory networks. *Plant J*, **52**, 14-29.
- Wilson, Z.A., Morroll, S.M., Dawson, J., Swarup, R. and Tighe, P.J.** (2001) The *Arabidopsis* MALE STERILITY1 (MS1) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. *Plant J*, **28**, 27-39.
- Wu, S.S., Moreau, R.A., Whitaker, B.D. and Huang, A.H.** (1999) Steryl esters in the elaioplasts of the tapetum in developing *Brassica* anthers and their recovery on the pollen surface. *Lipids*, **34**, 517-523.

- Yang, C., Vizcay-Barrena, G., Conner, K. and Wilson, Z.A.** (2007) MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. *Plant Cell*, **19**, 3530-3548.
- Yang, S.L., Jiang, L., Puah, C.S., Xie, L.F., Zhang, X.Q., Chen, L.Q., Yang, W.C. and Ye, D.** (2005) Overexpression of TAPETUM DETERMINANT1 alters the cell fates in the Arabidopsis carpel and tapetum via genetic interaction with excess microsporocytes1/extra sporogenous cells. *Plant Physiol*, **139**, 186-191.
- Yang, S.L., Xie, L.F., Mao, H.Z., Puah, C.S., Yang, W.C., Jiang, L., Sundaresan, V. and Ye, D.** (2003) Tapetum determinant1 is required for cell specialization in the Arabidopsis anther. *Plant Cell*, **15**, 2792-2804.
- Yang, W.C., Ye, D., Xu, J. and Sundaresan, V.** (1999) The SPOROCTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev*, **13**, 2108-2117.
- Ylstra, B., Touraev, A., Moreno, R.M., Stoger, E., van Tunen, A.J., Vicente, O., Mol, J.N. and Heberle-Bors, E.** (1992) Flavonols Stimulate Development, Germination, and Tube Growth of Tobacco Pollen. *Plant Physiol*, **100**, 902-907.
- Zhang, W., Sun, Y., Timofejeva, L., Chen, C., Grossniklaus, U. and Ma, H.** (2006) Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development*, **133**, 3085-3095.
- Zhang, X., Feng, B., Zhang, Q., Zhang, D., Altman, N. and Ma, H.** (2005) Genome-wide expression profiling and identification of gene activities during early flower development in Arabidopsis. *Plant Mol Biol*, **58**, 401-419.
- Zhao, D.Z., Wang, G.F., Speal, B. and Ma, H.** (2002) The excess microsporocytes1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the Arabidopsis anther. *Genes Dev*, **16**, 2021-2031.
- Zhao, X., de Palma, J., Oane, R., Gamuyao, R., Luo, M., Chaudhury, A., Herve, P., Xue, Q. and Bennett, J.** (2008) OsTDL1A binds to the LRR domain of rice receptor kinase MSP1, and is required to limit sporocyte numbers. *Plant J*, **54**, 375-387.

Zhu, J., Chen, H., Li, H., Gao, J.F., Jiang, H., Wang, C., Guan, Y.F. and Yang, Z.N.

(2008) Defective in Tapetal Development and Function1 (TDF1) is Essential for Anther Development and Tapetal Function for Microspore Maturation in *Arabidopsis*. *Plant J.*

Chapter 3

Identification of the binding sites of DYT1 by Selex and ChIP

Dr. Yiben Peng and Xuan Ma sequenced the oligos selected from the Selex procedure and helped analyze the binding sites based on the sequencing results. Dihong Lu helped with the sample collection in ChIP experiments. Dr. Gangping Xue made suggestions on the Selex protocol and Dr. Mengjuan Guo at the Cold Spring Harbor Laboratory provided useful suggestions about the ChIP procedure.

3.1 ABSTRACT

The transcriptome analysis described in Chapter 2 uncovered a list of genes downstream of *DYT1*. It suggests that *DYT1* might regulate the expression of these genes either directly or indirectly by activating at least 32 transcription factors. However, it was not clear which genes are the direct targets of *DYT1*. This is a key question to address in order to decipher the structure of the transcriptional network downstream of *DYT1*. One approach to identify direct targets of a transcription factor is to identify the *in vitro* binding sites. Based on the binding site sequence, a list of candidate target genes can be obtained. DYT1 is a bHLH protein which is suggested to bind to E-boxes, however, this hypothesis has not been tested experimentally. Therefore, I took an unbiased selection procedure (Selex) to identify the *in vitro* binding sites for DYT1. I found that the binding sites of DYT1 are E-boxes. Then I confirmed the binding of DYT1 to specific E-box containing oligos by using gel-shift assays. By searching the promoter sequences of the downstream genes of *DYT1*, I found that a list of genes have E-boxes in promoter and could be direct targets of DYT1. I further tested the *in vivo* binding of DYT1 to those genes by using chromatin immuno precipitation (ChIP). My results provide evidence for the direct binding of DYT1 to these target genes including *DYT1* itself. An updated model about the transcriptional regulation of tapetum functions is proposed, which suggest multiple mechanisms of regulation by DYT1. These results will help us to resolve the structure of the transcriptional cascade in the tapetum.

3.2 INTRODUCTION

The sequence of the DYT1 protein suggests that it is a bHLH transcription factor (Zhang *et al.*, 2006). In *Arabidopsis*, bHLH proteins comprise a large family with more than 140 members (Li *et al.*, 2006; Toledo-Ortiz *et al.*, 2003). The bHLH domain is about 60 amino acids long, which can be further divided into two secondary structures: a 15-aa basic region at the N-terminus that recognizes DNA and a Helix-Loop-Helix region at the C-terminus for protein dimerization (Li *et al.*, 2006; Toledo-Ortiz *et al.*, 2003). The DNA binding properties of bHLH proteins have been studied. In *Arabidopsis*, except for 27 bHLHs which do not have the basic region for DNA binding, all other bHLHs can bind to

either non E-boxes (11) or E-boxes (5'-CANNTG-3') (109). Among the E-boxes, the G-boxes (5'-CACGTG-3') are the preferred binding sites for a large group of bHLHs in *Arabidopsis*. Many bHLHs in *Arabidopsis* are predicted to be G-box binders because they have the conserved residues of His/Lys9, Glu13 and Arg17. Glu13 recognizes the conserved "CA" in E-box, and Arg17 specifically recognizes "G" in G-box. In addition, His/Lys9 recognizes "G" that is complementary to "C" in the middle of G-box. The bHLHs that lack these residues are predicted to bind to non G-box sequences (Toledo-Ortiz *et al.*, 2003). The HLH domain is not directly needed for DNA binding. However, the HLH region is required for interactions between bHLH proteins and interactions with other proteins such as R2R3 MYB, and the formation of complex is needed for the stable binding to E-boxes, which are palindrome structures. (Toledo-Ortiz *et al.*, 2003).

Many bHLH proteins have important functions in *Arabidopsis* development and responses to environment. For example, PHYTOCHROME INTERACTING FACTOR 1 (PIF1) is critical to chlorophyll biosynthesis and PIF3, PIF4 and PIF7 are involved in the phytochrome B signaling (Al-Sady *et al.*, 2006; Huq *et al.*, 2004; Huq *et al.*, 2002; Leivar *et al.*, 2008; Shen *et al.*, 2005). The bHLH proteins are also involved in developmental processes such as seed germination (PIL5), root vascular formation (LONESOME HIGHWAY), stomata differentiation (MUTE and SPEECHLESS), trichome and root hair differentiation (GL3 and EGL3), fruit development and dehiscence (ALCATRAZ and INDEHISCENT) (Bernhardt *et al.*, 2003; Bernhardt *et al.*, 2005; Liljegren *et al.*, 2004; MacAlister *et al.*, 2007; Oh *et al.*, 2007; Ohashi-Ito *et al.*, 2007; Pillitteri *et al.*, 2007; Rajani *et al.*, 2001; Schiefelbein, 2003; Zhao *et al.*, 2008). Recently, DYT1 was identified as a critical transcription factor in the specification of tapetum functions as described in Chapters 1 and 2 (Zhang *et al.*, 2006).

The transcriptome analysis of the *dyt1-1* mutant in Chapter 2 suggests that a transcriptional cascade is regulated by *DYT1*. However, it is not clear what the structure of this cascade is, especially which genes are directly regulated by DYT1. After comparing the sequence of DYT1 to that of E-box binding bHLHs, I predict that DYT1 might also bind to an E-box *in vivo*. Because DYT1 lacks the conserved Arg17 residue

for binding to G-box, I expected that DYT1 might not specifically bind to a G-box (DYT1 has Arg18). It is necessary to experimentally determine the binding sites of DYT1. Previous studies suggest that the binding sites of a transcription factor determined by *in vitro* experiments could be used to predict *in vivo* binding sites. For example, the binding sites for the AG and AG-like proteins were identified as a CArG box by *in vitro* experiments (Huang *et al.*, 1993; Huang *et al.*, 1996). In such experiments, the proteins purified from a heterogeneous system such as bacteria are used to select binding sites from a random DNA oligo pools and the consensus binding site is determined by summarizing the sequences of selected oligos. The CArG boxes were identified by this procedure and were also confirmed to be the *in vivo* binding site of the AG protein. For example, a CArG box at the 3'UTR of *SPL* can be bound by AG and is critical for the establishment of the *SPL* expression in the anther (Ito *et al.*, 2004). In this study, I first used an *in vitro* procedure called Selex to select DYT1 binding DNAs from a pool of random oligos (Xue, 2005). The empirically determined binding sites for DYT1 were then used to search for candidate targets of DYT1. However, transcription regulation is a complex process and the occurrence of binding sites is not always sufficient for a gene to be activated by corresponding transcription factors. It might involve multiple levels of regulation in addition to the binding of transcription factors, and other factors such as chromatin remodeling factors can also significantly affect the transcription by increasing or decreasing the accessibility of transcription factors to binding sites. To determine the *in vivo* binding sites, I used chromatin immuno-precipitation. Using this assay, I identified several genes as the direct targets of DYT1. This information will help us further resolve the structure of the DYT1 transcriptional cascade.

3.3 MATERIALS and METHODS

3.3.1 Transgenic plant construction

All plants were grown under the same conditions as described in Chapter 2. To identify the direct targets of DYT1 by ChIP, a genomic fragment of *DYT1* (*gDYT1*) was amplified without the stop codon and cloned into the plant expression vector pGWB19. The plasmid harboring the *gDYT1*: FLAG construct was sequenced to confirm the fidelity of

cloning and in-frame fusion. The transgene was then transformed into the *dyt1*^{+/-} plants via *Agrobacterium* mediated transformation. The T1 transgenic plants were genotyped by PCR and the rescued lines (*gDYT1: FLAG/dyt1*^{-/-}) were selected, from which the T2 transgenic plants were used for ChIP.

3.3.2 Selex

To identify the consensus binding sites for DYT1, a Selex procedure was used (figure 3.4). A 6×His:DYT1 construct was made by cloning the *DYT1* coding region into the pDEST17 vector (figure 3.1). The recombinant protein was then expressed, extracted and partially purified from the *E.coli* strain BL21 according to the manual for pDEST17 (Invitrogen). The DNA oligos were designed with 30bp random nucleotides in the middle region and two primer sequences at the flanking regions: 5'-CGGATCCCTCGAGCTGCAGC-(N30)-GCTAGCCGATCGGAGCTCGG-3'. The primer oMC2718 was used to make dsDNA probes by using Klenow enzyme according to the manufacture's manual (Promega). The binding site selection was done based on a procedure from Dr. Gangping Xue (Xue, 2005) (figure 3.4). The cobalt resin from PIERCE was used to do the Selex. After four rounds of Selex, the selected DNA oligos on the cobalt resin were amplified and cloned into pGEM-T vector (Promega) for sequencing.

3.3.3 Gel-Shift assay

The binding consensus sites for DYT1 determined by the Selex procedure were tested by using gel-shift assay. Protein purification was done with the same procedure as that for Selex. The biotin labeled dsDNA probes were made by Klenow or Taq DNA polymerase reactions. A synthesized 5'-Biotin labeled primer, oMC2746, was used to initiate the Klenow or Taq reaction on a synthesized probe or the clones in pGEM-T selected from Selex. The *in vitro* binding reaction was carried out according to the procedures of Huang *et al.* (Huang *et al.*, 1993; Huang *et al.*, 1996). The free and bound probes were then separated in 6% PAGE gel in 0.5×TBE at 100V×1hour. The probes were transferred to Hybond N membrane (Amersham) by using a semi-dry transfer apparatus (Bio-Rad).

After transferring, the probes were cross-linked to the membrane and the membrane was developed by using a Light Shift kit from PIERCE and X-film from Kodak.

3.3.4 ChIP

Young buds (up to flower stage 10) from the T2 generation of transgenic lines (*gDYT1:FLAG/dyt1-/-*) were collected. The chromatin preparation and immunoprecipitation procedure were performed according the published protocols (Gendrel *et al.*, 2002; Haring *et al.*, 2007; Lee *et al.*, 2007; Wang *et al.*, 2002). A Branson sonifier was used to break chromatin into 200-1000bp fragments. In my experiments, Agarose-Protein A beads (Upstate) or Dynal Protein G beads (Invitrogen) were used. Anti-FLAG monoclonal antibody was purchased from Sigma. Amplicon was designed as 200-300bp in length in most cases. Amplicons spanning the E-box region in the promoter of putative target genes were chosen to test the binding of DYT1 to this element. The amplicon containing no E-box was used as negative control.

3.3.5 Yeast two hybrid assay

Yeast two hybrid assay was used to test the protein-protein interactions among DYT1, AMS, three bHLH proteins (At1g06170, At2g31210, and At2g31220), MYB33, and MYB65. The open reading frames of these genes were amplified by using the cDNAs prepared from wild type young buds and the Pfx DNA polymerase from Invitrogen. The PCR products were cloned into the pDEST-32 and pDEST22 vectors. Yeast host strain PJ694A was used to test the interactions.

3.4 RESULTS

3.4.1 DYT1 binds to E-boxes (CANNTG)

The first step to identify the targets of DYT1 is to determine the DYT1 binding sites. By using Selex, I sequenced about 140 clones which contained the dsDNAs selected by 6×His:DYT1. The nucleotide sequences of the selected dsDNAs were aligned to find the consensus sites. It was found that DYT1 binds to E-boxes (figure 3.2). To confirm this result, a gel shift assay was used to test the specificity of the binding of DYT1 to an E-

box fragment from the *MSI* promoter. It was shown that DYT1 can bind to this fragment strongly. This binding can be eliminated by adding competitor probes, which suggests that the binding is specific. One point mutant (C→A at the first position or A→G at the second position) can abolish the binding. This suggests that the first two nucleotides are critical (figure 3.3-A). The binding activity is proportional to the amount of the DYT1 protein in the binding reaction. This further confirmed the binding specificity (figure 3.3-B).

3.4.2 DYT1 binds to its own promoter *in vivo*

It is notable that *DYT1* shows a specific peak expression during anther stages 5-6. A fact is that the *DYT1* promoter has E-boxes. Both lines of evidence suggest that DYT1 might regulate itself. To test this hypothesis, I carried out ChIP assay, and found that an E-box containing fragment in the *DYT1* promoter was enriched in the ChIP-DNA pools from the *gDYT1: FLAG* transgenic plants contrasting to the fragments that did not span E-boxes (figure 3.5).

3.4.3 DYT1 binds to the promoters of *MYB35*, *MS* and other transcription factor genes except *AMS*

Previous analysis on the *dyt1-1* transcriptome revealed a list of transcription factor genes that are downstream of *DYT1*. E-boxes were found in the promoters of these transcription factors. I then tested the *in vivo* binding of DYT1 to the promoters of these genes. *MYB35* was recently identified as an important gene in anther development (Zhu *et al.*, 2008). It is downstream of *DYT1* and upstream of *AMS* and *MYB103* (Sorensen *et al.*, 2003; Zhang *et al.*, 2007). 5 E-boxes were found in its promoter. ChIP-PCR analysis suggests that DYT1 can bind to these E-boxes. Since the amplicon is about 200 bps and spans several E-boxes, it is hard to differentiate which E-box is involved and which is not. More work is needed to narrow down the exact binding sites for DYT1. By using the same method, I found that a set of transcription factor genes might be the direct targets of DYT1. These include *MSI*, a gene encoding a PHD domain transcription factor important for pollen wall formation, and other genes encoding the following transcriptional factors: *MYB103*, *FLC*, *TAZ/At3g48360*, and *MYB32* (figure 3.5) (Ito *et al.*, 2007; Kapoor *et al.*,

2002; Preston *et al.*, 2004; Searle *et al.*, 2006; Yang *et al.*, 2007). Contrasting to our expectation, the E-boxes of *AMS* could not be bound by DYT1.

3.4.4 DYT1 binds to the promoter of chalcone synthase gene, a key gene in flavonol biosynthesis pathway.

In order to test if DYT1 can also directly bind to the promoter of enzyme genes, a chalcone synthase gene was selected. Chalcone synthase is a key enzyme in flavonoids biosynthesis pathway which is affected in *dyt1-1* (Atanassov *et al.*, 1998; Napoli *et al.*, 1999; Pollak *et al.*, 1993; van der Meer *et al.*, 1992; van Eldik *et al.*, 1997). The occurrence of E-boxes in the *CHS* promoter suggests it is a direct target of DYT1. In the ChIP assay, this fragment was enriched in the pool of ChIP-DNAs from the *gDYT1:FLAG* transgenic plants (figure 3.5). This result suggests that DYT1 also directly regulates enzyme genes.

3.4.5 DYT1, AMS and three anther preferential bHLHs interact in yeast two hybrid assays

In *Arabidopsis*, there are 5 bHLH genes that are tapetum preferential as suggested by the microarray analysis of the *ems* and *spl* mutants: *DYT1*, *AMS*, At1g06170, At2g31210 and At2g31220. The last three bHLH genes form a clade on the phylogenetic tree of the bHLH gene family. Previous studies have shown that protein-protein interactions are required for the formation of transcriptional complexes of bHLH proteins. Our microarray results revealed a partial overlap between the *dyt1-1* and *ams* transcriptome, which is consistent with a hypothesis that DYT1 and AMS interact in a transcriptional complex. The other bHLHs might function in different complexes and control different subset of genes in the tapetum. In a yeast two-hybrid assay, it was found that DYT1 forms homodimer, and that DYT1 can also form heterodimers with AMS and the other 3bHLHs (figure 3.6). Because of the self-activation of the activation domain fused with AMS and bHLHs, the potential interactions among AMS and 3bHLHs are not tested. Two R2R3 MYB proteins, MYB33 and MYB65, were also tested in yeast two-hybrid

assays. The *myb33*^{-/-}*myb65*^{-/-} double mutant exhibit tapetum defects at similar stages to that of *dyt1-1*. MYB-bHLH heterodimerization has been suggested by previous studies. However, MYB33 and MYB65 do not interact with DYT1 in yeast two-hybrid assays.

3.5 DISCUSSION

3.5.1 DYT1 might bind to E-boxes as dimer

I determined that E-box (CANNTG) is the binding site for DYT1 by performing Selex and gel-shift. By mutating the sites in E-box, I found that the first two nucleotides in E-boxes are critical for the DYT1 binding. The sequence of E-box is palindrome which suggests that DYT1 might bind to E-boxes as dimer. This is consistent with the yeast two hybrid result that DYT1 form dimer with its self and with other bHLH proteins including AMS. Further experiments should be done to test the occupancy of the target DNA by the DYT1 complexes. The formation of dimers might provide a mechanism of the regulation of the DYT1 activity. Because dimer is suggested to be a functional form of DYT1, I speculate that inhibition of dimer formation might be critical to repress the DYT1 function. The formation of heterodimers with DYT1 might affect the specificity of target genes. It will be interesting to determine if the complexes with different bHLHs have difference in binding properties.

3.5.2 The establishment of the *DYT1* expression requires self regulation and the functions of *SPL* and *EMS*

The finding by ChIP assays that DYT1 binds to its own promoter suggests the existence of a self regulation mechanism of DYT1 (Lee *et al.*, 2007; Wang *et al.*, 2002). I speculate that the self regulation of DYT1 is required to establish the specific expression pattern of DYT1 in the tapetum. It is expressed at a background level before stage 5. Then at stages 5-6, it is highly expressed in the tapetal cells (Zhang *et al.*, 2006). Because dimerization might be required for the transcriptional activity of DYT1, I speculate that the DYT1 dimerization might be inhibited by unknown mechanisms before stage 5 so that only basal level of transcription is detected. Whereas at stage 5, DYT1 forms dimers and

activates its own expression by binding to the E-boxes in its own promoter (Yang *et al.*, 2005; Yang *et al.*, 2003; Zhao *et al.*, 2002). Through such a self-activation, *DYT1* acquires a high expression level which might be required to reinforce the commitment of transcriptome changes in the tapetum. Similar theme can also be found in other development processes. For example, the formation of heterodimers by the floral homeotic proteins AP3 and PI is required for the establishment of high expression of both genes. Without the formation of dimers, neither AP3 nor PI can enter nuclei to activate themselves and downstream genes (McGonigle *et al.*, 1996).

The genes important for early anther development, *SPL* and *EMS1*, might be required to establish a basal and peak expression of *DYT1* respectively. The fact that the *DYT1* expression is not detected in the *spl* mutant suggests that *SPL* is required for the expression of *DYT1*. In my study, I first identified the binding sites of *SPL* because its DNA binding properties have not been studied. By performing Selex and gel shift, I found that *SPL* binds to a 6-bp site which also exists in the *DYT1* and *EMS1* promoters. This finding suggests that *SPL* could directly activate both *DYT1* and *EMS1* (Yang *et al.*, 1999). However, a peak expression of *DYT1* in the tapetum is dependent on *TPD1/EMS1* because the *DYT1* peak expression is abolished in either *tpd1* or *ems1*. *TPD1* is produced and release by meiocytes as a signal peptide to activate the *EMS1* signaling pathways. The signal from the *EMS1* kinase might finally promote the dimerization of *DYT1* and the establishment of the peak expression of *DYT1*. Similar mechanisms also exist in other signaling processes such as the Brassinosteroid-BRI signaling (Belkhadir *et al.*, 2006; Belkhadir *et al.*, 2006; He *et al.*, 2000; Kinoshita *et al.*, 2005; Li *et al.*, 1997; Nam *et al.*, 2002; Vert *et al.*, 2006; Yin *et al.*, 2005; Yin *et al.*, 2002). In that process, phosphorylation and dimerization of transcription factors are the key switch to activating BR response genes. The *DYT1* self activation under the regulations of both *SPL* and *TPD1/EMS1* might be an important mechanism to coordinate tapetum functions with the development of meiocyte/microspore.

3.5.3 *DYT1* directly regulates downstream transcription factor genes and enzyme genes

Previous microarray analysis suggests that *DYT1* regulates at least 32 transcription factor genes. Based on ChIP results, I found that some of them are direct targets of *DYT1* and that only the DNA fragments containing E-boxes can be bound by *DYT1*. The fragments without E-boxes can not be bound. This suggests that E-box is necessary for the binding by *DYT1*. However, occurrence of E-boxes is not sufficient for a promoter to be bound by *DYT1*. For example, although the *AMS* promoter contains E-boxes, it can not be bound by *DYT1* in ChIP assays. The fact that a genomic fragment of *AMS* without E-boxes at the promoter can still rescue the *ams* mutants suggests that those E-boxes are not essential and might not be bound by *DYT1* (Sorensen *et al.*, 2003). Other factors such as modifications of chromatin and interactions with other proteins can also affect the binding of *DYT1* to E-boxes *in vivo*.

The ChIP results also suggest that both direct and indirect regulation by *DYT1* might be important. For example, *AMS* is indirectly regulated by *DYT1*. In ChIP assays, *DYT1* can directly regulate *MYB35* but not *AMS* (Zhu *et al.*, 2008). Therefore, in the *DYT1-MYB35-AMS* transcriptional cascade, *AMS* is indirectly regulated by *DYT1* through *MYB35*. However, *MYB103* is regulated by *DYT1* both directly and indirectly. In my ChIP analysis, it was found that *DYT1* directly binds to the *MYB103* promoter (Zhang *et al.*, 2007; Zhu *et al.*, 2008). In the *DYT1-MYB35-AMS-MYB103* transcriptional cascade, *DYT1* regulates *MYB103* both by direct binding and by indirect regulation through the activation of *AMS* as an intermediate factor. *DYT1* also directly regulates enzyme genes which are positioned at the end of transcriptional cascade. Here the regulation of downstream genes by *DYT1* is not a linear, uni-directional process; instead, co-regulation and cross-regulation are very important so that the transcription network exhibits more complexity. A deeper study of the structure of the *DYT1* transcriptional network is needed to obtain a better understanding, and genome wide approaches such as ChIP on Chip or ChIP-sequencing should be applied in order to obtain more information to facilitate this study (Lee *et al.*, 2007).

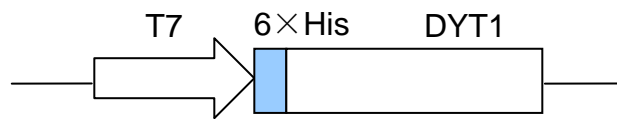
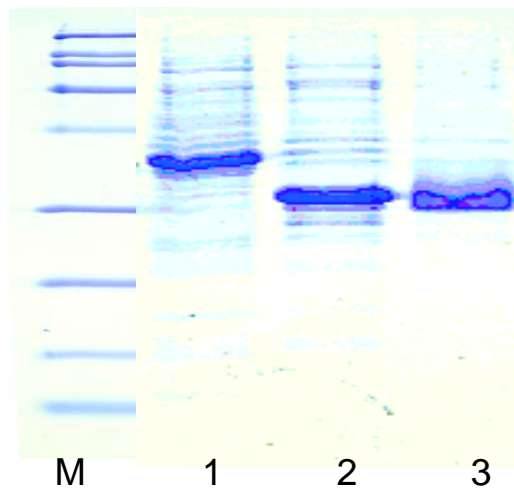
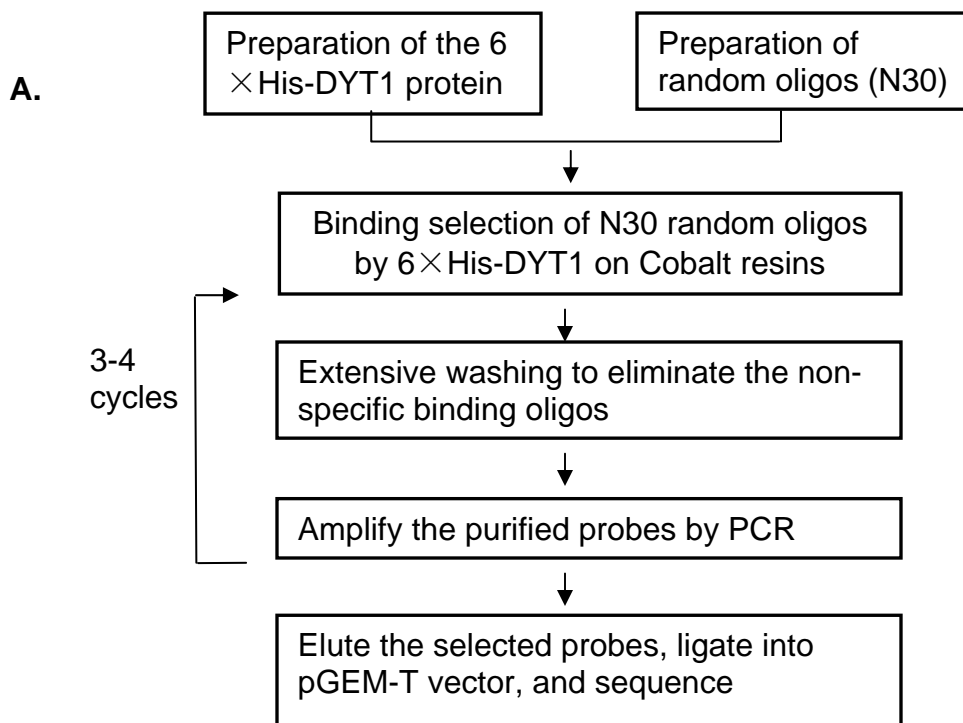
A**B**

Figure 3.1: Expression and purification of the 6×His-DYT1 recombinant protein in BL21. A: Construct for 6×His-DYT1 expression in *E.coli*. B. Extraction and purification of 6×His-DYT1: Lane M, protein marker; lane 1, SPL lysate; lane 2, 6×His-DYT1 lysate; lane 3, partially purified 6×His-DYT1.



B.

	-2	-1	1	2	3	4	5	6	+1	+2
A	14	10	4	32	15	18	1	2	10	11
C	10	20	40	8	1	15	9	0	11	9
T	16	3	4	7	6	1	29	1	4	17
G	10	17	4	3	28	17	12	47	25	13
	T/a	C/g	C	A	G/a	A/g,c	T/g	G	G/a,c	T/g,a

Figure 3.2 Identification of DYT1 binding sites by Selex procedure. A. A flow chart of Selex experiments; B. Summary of the sequencing results from the selected probes.

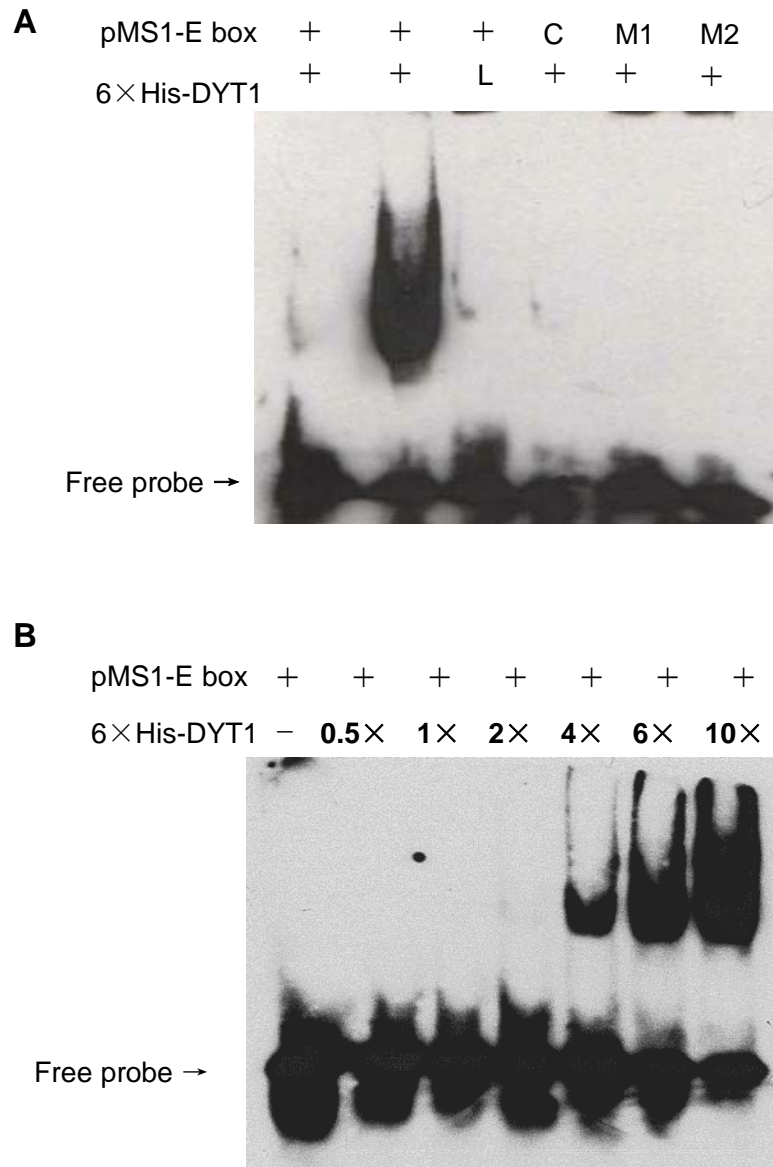
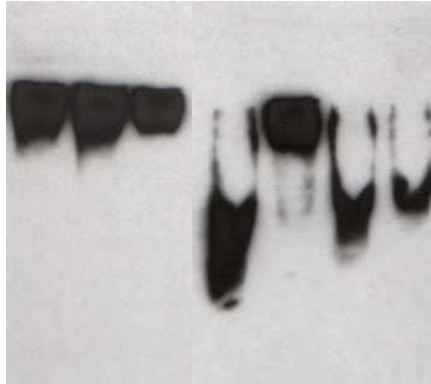


Figure 3.3: DYT1 binds to the E-box of the MS1 promoter. **A.** DYT1 bind to the E-box of MS1 promoter specifically. C: competitor probe with the same sequence with pMS1-E box; M1, mutant probe; M2, mutant probe. L, lysate from the bacteria expressing 6×His-SPL. **B.** The intensity of shift band is proportional to the amount of 6×His-DYT1 in binding reaction.

A

SPL	+	+	+	-	+	L	+
Probe- Selex	#45	#57	#70	#45	#45	#45	#45+C



B

#45	CGGATCCCT	<u>CGAGCT</u>	GCAGCGCTAAGGGTA
#57	GGCGAAAAGAAT	<u>GGAGCT</u>	AGCCGATCGGAG
#70	GCACTAGTGATT	CGGATCCCT	<u>CGAGCT</u> GCA

Figure 3.4: SPL bind to a 6bp motif specifically: A. SPL binds to three different probes selected from Selex procedure. The SPL binding to probe #45 is specific. B. The nucleotide sequence of three selected probes from Selex. The consensus site is highlighted.

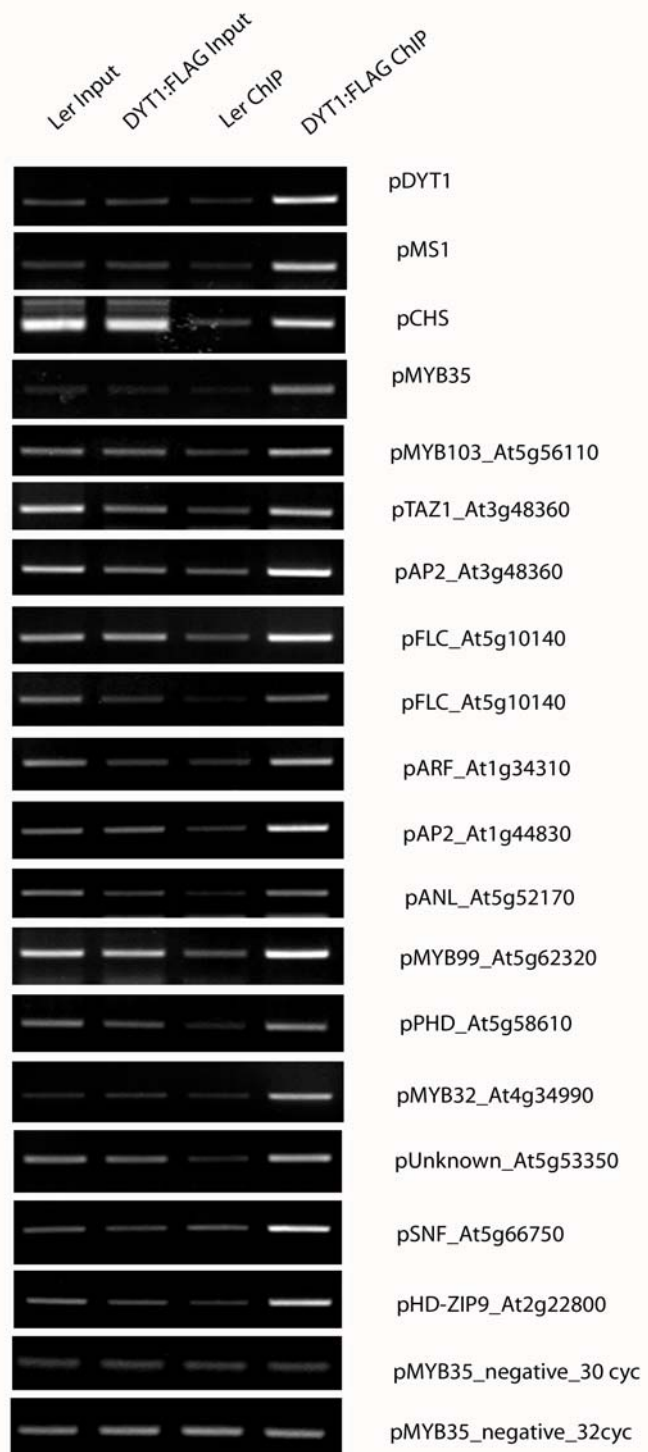
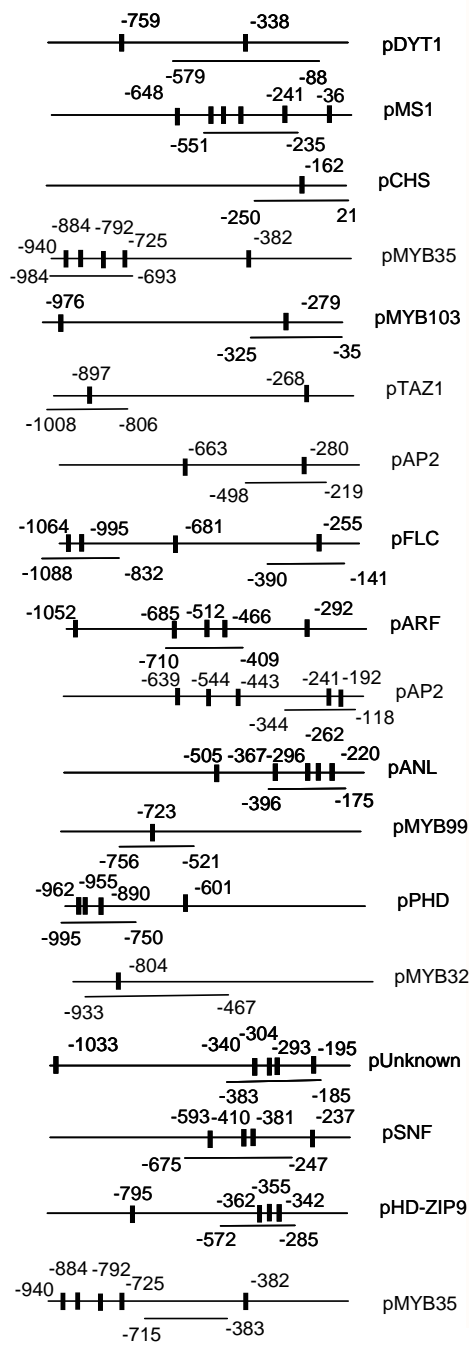


Figure 3.5: DYT1 binds to the promoters of downstream genes as revealed by ChIP: The locations of E-boxes at each promoter is shown as black rectangular box. The amplicon designed for ChIP-PCR is shown as a line below the promoter. For the gel image, each gel slice has four lanes. From the left to the right, they are ler input, gDYT1:FLAG input, ler ChIP, gDYT1:FLAG ChIP. An amplicon at the MYB35 promoter that does not span E-box was used as a negative control.

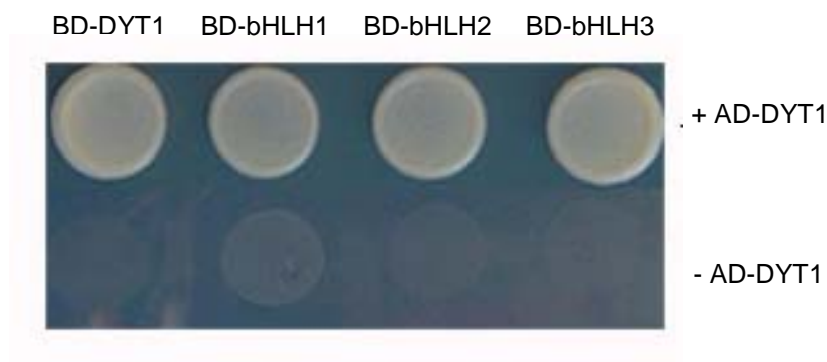


Figure 3.6: DYT1 interacts with 3bHLH and itself in yeast two hybrid assay. The minimal plate minus histidine and adenine was used for stringent assay.

3.6 References

- Al-Sady, B., Ni, W., Kircher, S., Schafer, E. and Quail, P.H.** (2006) Photoactivated phytochrome induces rapid PIF3 phosphorylation prior to proteasome-mediated degradation. *Mol Cell*, **23**, 439-446.
- Atanassov, I., Russinova, E., Antonov, L. and Atanassov, A.** (1998) Expression of an anther-specific chalcone synthase-like gene is correlated with uninucleate microspore development in *Nicotiana sylvestris*. *Plant Mol Biol*, **38**, 1169-1178.
- Belkhadir, Y. and Chory, J.** (2006) Brassinosteroid signaling: a paradigm for steroid hormone signaling from the cell surface. *Science*, **314**, 1410-1411.
- Belkhadir, Y., Wang, X. and Chory, J.** (2006) Brassinosteroid signaling pathway. *Sci STKE*, **2006**, cm4.
- Bernhardt, C., Lee, M.M., Gonzalez, A., Zhang, F., Lloyd, A. and Schiefelbein, J.** (2003) The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root. *Development*, **130**, 6431-6439.
- Bernhardt, C., Zhao, M., Gonzalez, A., Lloyd, A. and Schiefelbein, J.** (2005) The bHLH genes GL3 and EGL3 participate in an intercellular regulatory circuit that

- controls cell patterning in the Arabidopsis root epidermis. *Development*, **132**, 291-298.
- Gendrel, A.V., Lippman, Z., Yordan, C., Colot, V. and Martienssen, R.A.** (2002) Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. *Science*, **297**, 1871-1873.
- Haring, M., Offermann, S., Danker, T., Horst, I., Peterhansel, C. and Stam, M.** (2007) Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods*, **3**, 11.
- He, Z., Wang, Z.Y., Li, J., Zhu, Q., Lamb, C., Ronald, P. and Chory, J.** (2000) Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. *Science*, **288**, 2360-2363.
- Huang, H., Mizukami, Y., Hu, Y. and Ma, H.** (1993) Isolation and characterization of the binding sequences for the product of the Arabidopsis floral homeotic gene AGAMOUS. *Nucleic Acids Res*, **21**, 4769-4776.
- Huang, H., Tudor, M., Su, T., Zhang, Y., Hu, Y. and Ma, H.** (1996) DNA binding properties of two Arabidopsis MADS domain proteins: binding consensus and dimer formation. *Plant Cell*, **8**, 81-94.
- Huq, E., Al-Sady, B., Hudson, M., Kim, C., Apel, K. and Quail, P.H.** (2004) Phytochrome-interacting factor 1 is a critical bHLH regulator of chlorophyll biosynthesis. *Science*, **305**, 1937-1941.
- Huq, E. and Quail, P.H.** (2002) PIF4, a phytochrome-interacting bHLH factor, functions as a negative regulator of phytochrome B signaling in Arabidopsis. *Embo J*, **21**, 2441-2450.
- Ito, T., Nagata, N., Yoshida, Y., Ohme-Takagi, M., Ma, H. and Shinozaki, K.** (2007) Arabidopsis MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development. *Plant Cell*, **19**, 3549-3562.
- Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J.L. and Meyerowitz, E.M.** (2004) The homeotic protein AGAMOUS controls microsporogenesis by regulation of SPOROCTELESS. *Nature*, **430**, 356-360.

- Kapoor, S., Kobayashi, A. and Takatsuji, H.** (2002) Silencing of the tapetum-specific zinc finger gene TAZ1 causes premature degeneration of tapetum and pollen abortion in petunia. *Plant Cell*, **14**, 2353-2367.
- Kinoshita, T., Cano-Delgado, A., Seto, H., Hiranuma, S., Fujioka, S., Yoshida, S. and Chory, J.** (2005) Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature*, **433**, 167-171.
- Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W., Zhao, H., Lee, I. and Deng, X.W.** (2007) Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell*, **19**, 731-749.
- Leivar, P., Monte, E., Al-Sady, B., Carle, C., Storer, A., Alonso, J.M., Ecker, J.R. and Quail, P.H.** (2008) The Arabidopsis phytochrome-interacting factor PIF7, together with PIF3 and PIF4, regulates responses to prolonged red light by modulating phyB levels. *Plant Cell*, **20**, 337-352.
- Li, J. and Chory, J.** (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell*, **90**, 929-938.
- Li, X., Duan, X., Jiang, H., Sun, Y., Tang, Y., Yuan, Z., Guo, J., Liang, W., Chen, L., Yin, J., Ma, H., Wang, J. and Zhang, D.** (2006) Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiol*, **141**, 1167-1184.
- Liljegren, S.J., Roeder, A.H., Kempin, S.A., Gremski, K., Ostergaard, L., Guimil, S., Reyes, D.K. and Yanofsky, M.F.** (2004) Control of fruit patterning in Arabidopsis by INDEHISCENT. *Cell*, **116**, 843-853.
- MacAlister, C.A., Ohashi-Ito, K. and Bergmann, D.C.** (2007) Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. *Nature*, **445**, 537-540.
- McGonigle, B., Bouhidel, K. and Irish, V.F.** (1996) Nuclear localization of the Arabidopsis APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression. *Genes Dev*, **10**, 1812-1821.
- Nam, K.H. and Li, J.** (2002) BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell*, **110**, 203-212.

- Napoli, C.A., Fahy, D., Wang, H.Y. and Taylor, L.P.** (1999) white anther: A petunia mutant that abolishes pollen flavonol accumulation, induces male sterility, and is complemented by a chalcone synthase transgene. *Plant Physiol*, **120**, 615-622.
- Oh, E., Yamaguchi, S., Hu, J., Yusuke, J., Jung, B., Paik, I., Lee, H.S., Sun, T.P., Kamiya, Y. and Choi, G.** (2007) PIL5, a phytochrome-interacting bHLH protein, regulates gibberellin responsiveness by binding directly to the GAI and RGA promoters in Arabidopsis seeds. *Plant Cell*, **19**, 1192-1208.
- Ohashi-Ito, K. and Bergmann, D.C.** (2007) Regulation of the Arabidopsis root vascular initial population by LONESOME HIGHWAY. *Development*, **134**, 2959-2968.
- Pillitteri, L.J., Sloan, D.B., Bogenschutz, N.L. and Torii, K.U.** (2007) Termination of asymmetric cell division and differentiation of stomata. *Nature*, **445**, 501-505.
- Pollak, P.E., Vogt, T., Mo, Y. and Taylor, L.P.** (1993) Chalcone Synthase and Flavonol Accumulation in Stigmas and Anthers of *Petunia hybrida*. *Plant Physiol*, **102**, 925-932.
- Preston, J., Wheeler, J., Heazlewood, J., Li, S.F. and Parish, R.W.** (2004) AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. *Plant J*, **40**, 979-995.
- Rajani, S. and Sundaresan, V.** (2001) The Arabidopsis myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. *Curr Biol*, **11**, 1914-1922.
- Schiefelbein, J.** (2003) Cell-fate specification in the epidermis: a common patterning mechanism in the root and shoot. *Curr Opin Plant Biol*, **6**, 74-78.
- Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasino, R.A. and Coupland, G.** (2006) The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. *Genes Dev*, **20**, 898-912.
- Shen, H., Moon, J. and Huq, E.** (2005) PIF1 is regulated by light-mediated degradation through the ubiquitin-26S proteasome pathway to optimize photomorphogenesis of seedlings in Arabidopsis. *Plant J*, **44**, 1023-1035.
- Sorensen, A.M., Krober, S., Unte, U.S., Huijser, P., Dekker, K. and Saedler, H.** (2003) The Arabidopsis ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. *Plant J*, **33**, 413-423.

- Toledo-Ortiz, G., Huq, E. and Quail, P.H.** (2003) The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell*, **15**, 1749-1770.
- van der Meer, I.M., Stam, M.E., van Tunen, A.J., Mol, J.N. and Stuitje, A.R.** (1992) Antisense inhibition of flavonoid biosynthesis in petunia anthers results in male sterility. *Plant Cell*, **4**, 253-262.
- van Eldik, G.J., Reijnen, W.H., Ruiter, R.K., van Herpen, M.M., Schrauwen, J.A. and Wullems, G.J.** (1997) Regulation of flavonol biosynthesis during anther and pistil development, and during pollen tube growth in *Solanum tuberosum*. *Plant J*, **11**, 105-113.
- Vert, G. and Chory, J.** (2006) Downstream nuclear events in brassinosteroid signalling. *Nature*, **441**, 96-100.
- Wang, H., Tang, W., Zhu, C. and Perry, S.E.** (2002) A chromatin immunoprecipitation (ChIP) approach to isolate genes regulated by AGL15, a MADS domain protein that preferentially accumulates in embryos. *Plant J*, **32**, 831-843.
- Xue, G.P.** (2005) A CELD-fusion method for rapid determination of the DNA-binding sequence specificity of novel plant DNA-binding proteins. *Plant J*, **41**, 638-649.
- Yang, C., Vizcay-Barrena, G., Conner, K. and Wilson, Z.A.** (2007) MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. *Plant Cell*, **19**, 3530-3548.
- Yang, S.L., Jiang, L., Puah, C.S., Xie, L.F., Zhang, X.Q., Chen, L.Q., Yang, W.C. and Ye, D.** (2005) Overexpression of TAPETUM DETERMINANT1 alters the cell fates in the Arabidopsis carpel and tapetum via genetic interaction with excess microsporocytes1/extra sporogenous cells. *Plant Physiol*, **139**, 186-191.
- Yang, S.L., Xie, L.F., Mao, H.Z., Puah, C.S., Yang, W.C., Jiang, L., Sundaresan, V. and Ye, D.** (2003) Tapetum determinant1 is required for cell specialization in the Arabidopsis anther. *Plant Cell*, **15**, 2792-2804.
- Yang, W.C., Ye, D., Xu, J. and Sundaresan, V.** (1999) The SPOROCTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev*, **13**, 2108-2117.

- Yin, Y., Vafeados, D., Tao, Y., Yoshida, S., Asami, T. and Chory, J.** (2005) A new class of transcription factors mediates brassinosteroid-regulated gene expression in Arabidopsis. *Cell*, **120**, 249-259.
- Yin, Y., Wang, Z.Y., Mora-Garcia, S., Li, J., Yoshida, S., Asami, T. and Chory, J.** (2002) BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. *Cell*, **109**, 181-191.
- Zhang, W., Sun, Y., Timofejeva, L., Chen, C., Grossniklaus, U. and Ma, H.** (2006) Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development*, **133**, 3085-3095.
- Zhang, Z.B., Zhu, J., Gao, J.F., Wang, C., Li, H., Zhang, H.Q., Zhang, S., Wang, D.M., Wang, Q.X., Huang, H., Xia, H.J. and Yang, Z.N.** (2007) Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis. *Plant J*, **52**, 528-538.
- Zhao, D.Z., Wang, G.F., Speal, B. and Ma, H.** (2002) The excess microsporocytes1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the Arabidopsis anther. *Genes Dev*, **16**, 2021-2031.
- Zhao, M., Morohashi, K., Hatlestad, G., Grotewold, E. and Lloyd, A.** (2008) The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. *Development*, **135**, 1991-1999.
- Zhu, J., Chen, H., Li, H., Gao, J.F., Jiang, H., Wang, C., Guan, Y.F. and Yang, Z.N.** (2008) Defective in Tapetal Development and Function1 (TDF1) is Essential for Anther Development and Tapetal Function for Microspore Maturation in Arabidopsis. *Plant J*.

Chapter 4

Analysis of the *Arabidopsis* floral proteome: detection of over 2000 proteins and evidence for post translational modifications

Baomin Feng¹, Lianchao Li², Xiaofan Zhou³, Bruce Stanley⁴, and Hong Ma^{1*}

Jason Ma (Stanford University, CA) assisted with C scripts for analyzing the data. Jiong Wang (technician, Ma Lab, PSU, PA) helped with plant care, Dr. John Diller and Dihong Lu (Ma lab, PSU, PA) provided their comments on the manuscript, and Anne Stanley (the College of Medicine, PSU, Hershey, PA) assisted in processing the MudPIT samples. I also appreciate suggestions on proteomics procedures from Zhixin Zhao (Assmann lab, PSU, PA). This work is supported by a grant from the US Department of Energy (DE-FG02-02ER15332) to H.M. and by Biology Department at the Pennsylvania State University.

4.1 ABSTRACT

The proteome of the *Arabidopsis* flower has not been extensively studied previously. Here, we report a proteomic analysis of the wild type *Arabidopsis* flower. Using both 2-dimensional electrophoresis/mass spectrometry (2-DGE/MS) and multi-dimensional protein identification technology (MudPIT) approaches, we identified 2446 proteins. While a single experiment or analysis uncovered only a subset of the proteins we identified, a combination of multiple experiments and analyses facilitated the detection of a greater number of proteins. When proteins are grouped according the RNA expression levels revealed by microarray experiments, we found that proteins encoded by genes with relatively high levels of expression were detected with greater frequencies. On the other hand, at the level of individual gene/protein, there was not a good correlation between protein spot intensity and microarray values. We also obtained strong evidence for post-translational modification from 2-DGE and MudPIT data. We detected proteins that are annotated to function in protein synthesis, folding, modification, and degradation, as well as the presence of regulatory proteins such as transcription factors and protein kinases. Finally, sequence and evolutionary analysis of genes for active methyl group metabolisms suggest that these genes are highly conserved. Our results allow the formulation of hypotheses regarding post-translational regulation of proteins in the flower, providing new understanding about *Arabidopsis* flower development and physiology.

4.2 INTRODUCTION

The proteome represents the complete set of proteins that are present in an entire organism, a specific organ, tissue, cells or even sub-cellular compartments (Jorin et al., 2007). Proteins play crucial roles in most cellular processes, as fundamental structural components and catalysts of most biochemical reactions, including metabolism, signaling, and regulation. Meanwhile their modifications provide a very dynamic and flexible mechanism for regulating biological processes. Therefore, investigating the proteome is a key to better understanding the mechanism of cellular processes. However, due to the variable physical and chemical properties of proteins, the need for sufficient samples, and other difficulties of protein characterization, the analysis of the proteome has been more challenging than that of the transcriptome.

Recently, rapid advances techniques for protein separation and identification have created opportunities for deeper investigation of the proteome. Two-dimensional gel electrophoresis (2-DGE) separates proteins according to molecular weight and iso-electric point with relatively high precision (Jorin et al., 2007). In addition, the abundance of proteins analyzed using 2-DGE can be estimated using staining intensity on the gel. Proteins separated by a 2-DGE can be further analyzed by mass spectrometry (MS) for identification. Other separation techniques such as liquid chromatography (LC) have been used to separate small peptides (Washburn et al., 2001; Delahunty and Yates, 2007). This is the basis for a for whole-proteome analysis called multi-dimensional protein identification technology (MudPIT). In this method, peptides generated by trypsin digestion are separated by two- or multiple- dimension LC and their masses determined by MS (Florens and Washburn, 2006; Ru et al., 2006; Lee et al., 2007; Maor et al., 2007). Tandem MS/MS methods such as those performed by mass spectrometers called matrix assisted laser desorption /ionization- time of flight (MALDI-TOF/TOF) or quadrupole TOF (Q-TOF) are used to acquire both peptide masses and the specific fragmentation patterns of peptides, which can provide very reliable identification of proteins when combined with proper statistical methods for estimating false discovery rates. Today with the complete genome sequence of model organisms, large scale protein identification can be achieved by matching the peptide mass spectra with those predicted from the sequence database, including protein modifications (AGI, 2000). These techniques and resources allow new opportunities to investigate the

proteome, such as studies of the plant proteome from various tissues and cells or after treatment (Gallardo et al., 2001; Peck et al., 2001; Carter et al., 2004; Jarvis, 2004; Nuhse et al., 2004; Holmes-Davis et al., 2005; Majeran et al., 2005; Peck, 2005; Bayer et al., 2006; Chibani et al., 2006; Rajjou et al., 2006; Lee et al., 2007; Thelen and Peck, 2007). However, the *Arabidopsis* floral proteome has not been extensively studied previously.

The *Arabidopsis* flower has four types of floral organs (sepals, petals, stamens and pistils) arranged in four concentric whorls. According to morphological landmarks, the *Arabidopsis* flower development can be divided into 12 stages from the initiation of the floral meristem to the opening of the flower (Smyth et al., 1990). During stages 1-9, a number of key developmental events take place, such as the initiation of floral organ primordia, specification of organ type, meiosis and gametophyte initiation. At stage 12, all floral organs are mature, followed by later stages with pollen germination, double fertilization, and the early embryo development (Smyth et al., 1990). Our previous transcript profiling using microarray indicated that there are approximately 14,000 genes present in the stage-12 flower (Zhang et al., 2005b). In the last two decades, dozens of genes, such as those determining flowering time and floral organ identity, have been characterized by molecular genetic studies (Ma, 2005; Zahn et al., 2006). Those genetically identified genes represent only a small portion of the genes that are believed to be involved in flower development.

Proteomic studies can provide valuable information on floral proteins and enhance the understanding the molecular mechanism of flower development. After the completion of mRNA biogenesis, there are additional levels of regulation that affect the quantity and activity of the protein products. Translational control, protein modification and turnover can all contribute to the regulation of protein abundance and activities. Such information can only obtained by studies at the protein level. In yeast and mammals, although there is a general correlation between mRNA and protein levels, a substantial number of genes exhibit a discrepancy between mRNA and protein abundance (Futcher et al., 1999; Gygi et al., 1999; Pradet-Balade et al., 2001; Griffin et al., 2002; Schmidt et al., 2007). Because little is known about protein levels and protein modifications in the flower, we are interested in obtaining information about floral proteins using a proteomics approach.

Furthermore, we would like to investigate protein modifications that are present in the flower. The critical roles of protein modifications have been suggested by previous genetic studies. For example, the *CLV1* gene determines the size of floral meristem and encodes a receptor-like protein kinase, suggesting that it functions by phosphorylating downstream proteins (Clark et al., 1996; Brand et al., 2000; Schoof et al., 2000; DeYoung et al., 2006). Another form of protein modification, ubiquitination, can affect the stability and localization of a protein, thus affecting the corresponding biological functions. Defects in ubiquitination pathways can result in abnormal floral organ identity and meiosis as suggested by the phenotypes caused by a mutation in *ASK1*, which encodes a key component of SCF complexes that facilitate ubiquitination of target proteins (Samach et al., 1999; Yang et al., 1999a; Zhao et al., 1999; Zhao et al., 2001; Ni et al., 2004). However, there is relatively little knowledge about protein modification in the flower. A proteomic characterization of protein modifications will provide crucial clues about the regulation of a biological process and be valuable in engineering plants with altered protein activity.

In this study, we used two approaches to investigate the *Arabidopsis* floral proteome and obtained several interesting findings. First, we used 2-DGE and MS to identify 216 proteins in the flower proteome at stage 12. We also found some proteins at 2 or more spots by 2-DGE, suggesting that these proteins had alternative forms possibly due to modifications. Second, we identified nearly 2400 floral proteins from two MudPIT experiments. Moreover, we recovered further evidence for modified peptides using the MudPIT analyses. The identification of a large number of proteins provides direct evidence for the presence of proteins with putative function in different biochemical pathways. Additionally, our proteomic data suggests that methylation pathways play an important role in the flower. In general, both 2-DGE and MudPIT analyses allow the sampling of the floral proteome; in addition, relatively large percentages of protein products were identified for genes with high detected mRNA levels, particularly by two or more experiments. In contrast, low percentages of proteins encoded by genes with relatively low levels of mRNA were identified by proteomics. Nevertheless, we were able to detect proteins that correspond to genes with very low microarray intensities, suggesting that proteomics methods are able to detect products of very low abundance transcripts. Our analyses indicate that proteomics is an informative approach to study protein properties in a high throughput manner.

4.3 Materials and methods

4.3.1 Plant materials and growth conditions

The *Arabidopsis* wild type plants of the Columbia ecotype were used in this study. All plants were grown at 23°C under long day conditions of 16 hours in light and 8 hours in darkness. Flower buds at stage 12 were collected and frozen immediately in liquid nitrogen.

4.3.2 Protein isolation and separation by 2-D electrophoresis

To separate and identify floral proteins, we used 2-DGE gels with MALDI-TOF/TOF and/or Q-TOF identification of trypsin-digested gel-spots. Total protein was isolated from the flower buds using acetone/TCA precipitation methods. Briefly, about 0.5-1g flower buds were ground into fine powder in liquid nitrogen. The powder was re-suspended in 15% TCA with 0.07% β -mercaptoethanol in acetone (pre-chilled at -70°C), and precipitated at -20°C for at least 45 minutes. The precipitated proteins were collected by spinning at 11,000 \times g at 4°C for 20 minutes. The pellet was washed twice with cold acetone and dried under vacuum to form crude protein powder.

The crude protein extracts were dissolved in rehydration buffer (8M Urea, 2% CHAPS, 0.002% bromophenol blue) containing 0.5% Bio-Lyte and 50mM DTT. Quantification of the protein concentration was done by using the Bradford assay (Bio-Rad Protein Assay reagents). 200 μ g of total proteins were loaded onto the first dimension IPG (immobilized pH gradient) strips by soaking for 16 hours. The proteins were then focused on Bio-Rad IEF instruments. The SDS-PAGE electrophoresis was done according to the user's manual from Bio-Rad, with the Bio-Rad broad range protein markers used as molecular weight standards. The 2-D PAGE gels were then stained with Coomassie Brilliant Blue and scanned with a GC-800 densitometer. The acquired images were analyzed using Bio-Rad PDQuest software.

4.3.3 Identification of proteins by MALDI-TOF MS and/or Q-TOF MS

To identify the proteins, 265 visible protein spots on 2-D gel were cut out, digested by trypsin and put to MS analysis. The protein spots of interests were cut out from the 2-D gel either manually or using the Spot Cutter from Bio-Rad. Gel spots were subjected to in-gel trypsin digestion according to the user's manual (Trypsin Gold, Promega). Briefly, the gel slices were

washed three times with 100 μ L of 25 mM ammonium bicarbonate and dehydrated with 100 μ L of 50% acetonitrile, and dried in a Speed-Vac. They were then incubated overnight at 37 °C with trypsin (12.5 ng/ μ L in 25 mM ammonium bicarbonate). Peptides were then extracted twice with 25 μ L of 5% formic acid for 20 min. The extracts were dried in a Speed-Vac and resuspended in 35 μ L of 5% acetonitrile with formic acid (0.1%).

For Q-TOF (nano-LC-ESI-MS/MS) analysis, tryptic digests were analyzed by capillary liquid chromatography–nanoelectrospray ionization–tandem mass spectrometry (CapLC–ESI-MS/MS). A Micromass Q-TOF Premier mass spectrometer coupled with a Waters CapLC HPLC unit was used for the analysis. Peptides (1–5 μ L) were injected into solvent A (acetonitrile/ water/formic acid, 5/95/0.1) supplied by the auxiliary pump of the capillary HPLC unit and trapped in a Dionex μ -Precolumn (C₁₈ PepMap 100, 5 μ m, 300 μ m \times 5 mm) for on-line desalting and concentration. After washing for 5 min with solvent A at 10 μ L /min, trapped peptides were then back-flushed with the gradient solvent flow on to the analytical column, a Dionex PepMap fused silica capillary column (C₁₈, 5 μ m, 75 μ m \times 150 mm), using a 10-port switching valve. The analytical column was run with a gradient (0–100 % solvent B; acetonitrile/ water/formic acid; 95/5/0.1; in 45 min). The mass spectrometry was calibrated using Glu-Fib product ion.

The Q-TOF mass spectrometer was operated to acquire MS/MS of tryptic peptides in data-dependent acquisition mode for precursor ion selection using charge-state recognition and intensity threshold as selection criteria using MassLynx 4.0 SP1. In order to carry out the tandem mass spectrometric data acquisition, a survey scan (2 s) over the m/z of 400–1500 was performed. From each survey scan, up to four most intense precursor ions based on the selection criteria were selected for tandem mass spectrometry to obtain the production spectra resulting from collision-induced dissociation in the presence of argon. The product ion spectra (6–8 s) collected were processed using Protein Lynx Global Server 2.1 and were converted to peak list text files for database searching. For protein identification, the Internet-based MASCOT MS/MS Ions Search tool (http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS) was used to manually search against the National Center for Biotechnology Information (NCBI) non-redundant protein database. Search parameters were set as follows: taxonomy, All entries; enzyme, trypsin; variable modifications, acetyl (N-term), oxidation (M), phospho (ST), phospho (Y), pyro-glu (N-term E), pyro-glu (N-term Q); mass values, monoisotopic; protein mass,

unrestricted peptide; mass tolerance, 150 ppm; fragment mass tolerance, 0.2 D; max missed cleavages, 1. Proteins with a minimum of one peptide match (with a significant ion score of 25 according to MASCOT) and containing a sequence tag with at least three amino acids in a row in either Y- or B-ion series, but not matching to any other proteins, were considered as positive and listed in the final summary. All proteins with only a single peptide match were verified by manual inspection of Y- and B-ion series to identify the minimal sequence tag of three consecutive ions. Each of these proteins had multiple numbers of two amino acid sequence tags, significant Mascot scores > 25). For matches representing protein families, only one member is presented. All proteins identified through the NCBI database (and corresponding annotations) will be verified by BLAST searches against the public available database. To obtain the corresponding AGI identifier from SwissProt protein identifier, a C script is used to search the NCBI database. Corresponding microarray data from the Stage 12 flower tissue was used to evaluate the corresponding genes' transcripts level (Zhang et al., 2005b).

4.3.4 Multi-dimensional protein identification (MudPIT)

For MudPIT experiments, 4.5mg dried protein powder was suspended in 1mL of 100mM NH_4HCO_3 , then mixed by vigorous vortexing and sonication to improve the dissolution of proteins. The amount of proteins in this sample was estimated by the Bradford assay. Approximately 800-1000ug of protein was digested with about 20-30 μg of trypsin (Gold, Promega, Madison, WI, USA). DTT was added to the digestion reaction to a final concentration 10mM, and the sample was kept at 37°C for 1 hour. Then iodoacetamide (IDA) was added to a final concentration of 50mM, and the sample was kept in dark for 1 hour. About 20-30 μg of Trypsin Gold (Promega) was added to the sample. It was then incubated at 37°C for 18 hours. After clearing the sample by centrifugation at top speed for 10 minutes, the supernatant was adjusted to pH3-4 by adding 10% trifluoroacetic acid (TFA). This mixture of tryptic peptides was sent to the Penn State College of Medicine Mass Spec/Proteomics facility for MudPIT analysis. After 2D-LC separation of the tryptic peptide mix, the spectra from each spot were analyzed by using Mascot or ProteinPilot software searching against either the SwissProt or NCBI databases. The SwissProt identifier and NCBI accession number were converted to the *Arabidopsis* genome identifier by using a script which can search the online NCBI databases.

Some *Arabidopsis* genome identifiers were obtained by doing a local batch blast against *Arabidopsis* whole genome sequences (Altschul et al., 1997; AGI, 2000).

4.3.5 Go annotation and pathway analysis

To categorize the proteins identified by different approaches, and find the important pathways for flower development, we searched the GO annotation of each protein on TAIR website and the pathway information from a database from TAIR website. The percentage of each category is calculated and represented in a pie graph. A script was written for searching the pathways from the database Aracyc. Manually sorting and finding pathways were also done for some important pathways such as the methylation pathway.

4.3.6 Phylogenetic Analysis

The protein sequences of the genes selected from the list were used as queries to perform TBLASTN search against plant genomes on NCBI (National Center of Biotechnology Information) and JGI (Joint Genome Institute), or the TIGR Plant Transcript Assemblies Database for those species without complete genome sequences. An e-value cut-off at 1e-05 was used. Each new result was used to perform a second round of search until no new result was found. For each gene family, the alignment was carried out by MUSCLE version 3.6 (Edgar, 2004). The alignment results were manually adjusted by using GeneDoc version 2.6.002. The Neighbor-Joining trees were constructed by using MEGA 4.0 with a bootstrap analysis of 1000 replicates (Tamura et al., 2007). The Maximum Likelihood trees were constructed by using PHYML 2.4.4 with a bootstrap analysis of 100 replicates (Guindon and Gascuel, 2003). The substitution model and parameters for each family was chosen according to the test results of ProtTest 1.4 (Abascal et al., 2005) for protein tree or MrModelTest 1.32 (Posada and Crandall, 1998) for DNA tree. Only the NJ trees were shown with bootstrap values for NJ.

4.4 RESULTS

4.4.1 Identification of Floral Proteins by 2-DGE/MS and MudPIT

To investigate the *Arabidopsis* floral proteome, we first used 2-DGE electrophoresis and mass spectrometry to analyze proteins extracted with trichloroacetic acid (TCA) from stage 12 floral buds, as defined by Smyth et al (Smyth et al., 1990). Two dimension gel electrophoresis was carried out in two steps: isoelectric focusing (first dimension) and SDS-page (second dimension). After preliminary 2-DGE experiments, we chose to use pH4-7 IPG gel strips for the first dimension because this range can resolve most protein spots than can be resolved by a pH3-10 IPG strip and other single strips. On the 2-DGE, we were able to detect more than 300 spots by staining with Coomassie blue R-250 (Figure 1). A total of 265 spots were cut from the 2-DGE for protein identification by mass spectrometry and database search using the Mascot software (Perkins et al., 1999). Among these spots, at least one protein was identified from each of 185 spots by Q-TOF MS analysis (Supplemental Table 1). A subset of identified proteins was further tested and confirmed by MALDI-TOF/TOF MS analysis. In total, 216 non-redundant proteins were identified from the 185 spots on 2-DGE by using MALDI-TOF/TOF and/or Q-TOF MS/MS experiments.

Although 2-DGE provided estimates of relative protein abundance, molecular weight, and pI, the total number of proteins identified using 2-DGE was limited. To sample the flower proteome more deeply, we also used the LC-MALDI MudPIT approach. We performed two independent MudPIT experiments using two biological replicates of stage 12 flowers. The flowers were ground and extracted with TCA/acetone, then about 1 mg of the TCA/acetone extracted crude protein powder was digested with trypsin and analyzed by using MudPIT.

The procedure for MudPIT analysis involves the generation, separation, and detection of a large collection of peptides from the entire proteome. This process results in the detection of highly abundant proteins with relatively large probability and in principle uncover low abundance proteins with reduced probability. Therefore, MudPIT samples a subset of all peptides generated from the proteome, rather similar to EST (expressed sequence tag) analysis. In contrast, MudPIT is unlike microarray analysis, which can reproducibly detect gene expression above some threshold levels. To detect a maximum number of proteins with limited flower sample and resources, we analysis MS data from the two MudPIT experiments using two databases (SwissProt or NCBI) and two algorithms (Mascot or ProteinPilot) (Shilov et al., 2007). The combination of all of the analyses of both MudPIT experiments resulted in the identification of

2397 proteins (Supplemental Table 2). Among these, 1556 proteins were identified with the data from the first MudPIT experiment and 2220 proteins were identified from the second experiment, with an overlap of 1373 proteins identified by both experiments.

4.4.2 Multiple Sampling and Bioinformatic Analyses Increase Protein Detection

Among the 216 proteins identified by 2-DGE/MS analysis, 167 (77%) were identified by at least one of the MudPIT experiments (Figure 2), indicating that as few as two MudPIT experiments can detect most, albeit not all, proteins uncovered by 2-DGE analysis. As mentioned above, a large number of proteins were detected by only one of the two MudPIT experiments, a regularly observed phenomenon noted by other scientists, including researchers in the John Yates III lab where original MudPIT procedures were developed (Washburn et al., 2001; Washburn et al., 2003; Delahunty and Yates, 2007). While identifications of the most abundant proteins tend to show good overlap between replicates, each experiment is only able to sample a fraction of the peptides that are in a complex mixture, and the less abundant proteins that are identified tend to differ between experiments. – this has led to the estimation, again by John Yates, that one would have to run at least 9 replicates to have identified 95% of all peptides detectable by MS at least once, with each additional replicate experiment identifying additional peptides/proteins (Washburn et al., 2003; Delahunty and Yates, 2007).

To obtain further understanding of the MudPIT results, we sorted the identified proteins on the basis of whether a protein was detected (1) in one or the other of the MudPIT experiments; (2) in searches with either the SwissProt or NCBI databases; and (3) using either the Mascot or Paragon (ProteinPilot) algorithms, resulting in 8 sets of protein identifications (Supplemental Table 2). When we compared these 8 sets of proteins, we found that similar to the partial overlap between the results of the two MudPIT experiments, the two databases also supported the detection of partially overlapping protein sets, as did the two algorithms (Figure 2). Therefore, the combination of multiple analyses of the same MS/MS data by different algorithms yielded a greater number of protein identifications than any single analysis of one database with one algorithm. In fact, such multiple-algorithm –searches have been advocated as a better approach to MS/MS database searching by other investigators (e.g., (Resing et al., 2004).

Because the proteins identified by MudPIT analyses are likely to result from a sampling process, in which more abundant proteins are represented by more peptide spectra than are less abundant proteins, we wondered what might be the relationship between the level of gene expression and the frequency of detection of the corresponding protein. We divided the *Arabidopsis* genes into categories according to the microarray data we have previously obtained for expression in the flower using the Affymetrix ATH1 gene chip (Zhang et al., 2005b). We identified those proteins that are encoded by the genes in each category, and calculated the percentage of proteins detected by our proteomics experiments. We found that the general trend is that protein products of genes with higher expression levels tend to be detected at a higher frequency, particularly by two or more experiments (Figure 3). However, even for genes expressed at very high levels, not all encoded proteins were identified by the 2-DGE/MS and MudPIT experiments. At the same time, for the very large number of genes with relatively low values of microarray data, even a small percentage represented the detection of many proteins. Specifically, some of the identified proteins correspond to genes whose expression was below the level of reliable detection by the microarray experiment. This lack of good overall correlation between mRNA level and protein levels has previously been found in yeast and in human kidney cell lines, where observed changes in mRNA level can account for no more than 40% of the observed changes (up or down) in protein levels (Gygi et al., 1999; Griffin et al., 2002).

4.4.3 Comparison between protein and transcript levels

Although genome-wide mRNA profiling techniques such as microarray experiments can estimate relative gene expression levels, they do not provide information regarding the protein level. In addition, the above described relatively high protein detection frequencies corresponding to genes with high microarray values suggest that genes with high transcript levels tend to have high protein levels. However, it is not clear whether at the individual gene/protein level, whether there is a good correlation between mRNA and protein levels. Therefore, we have examined the 2-DGE spot intensity using the PDQuest software, and compared the spot intensity with the microarray value of the corresponding gene.

As mentioned before, sometimes two or more proteins were identified from the same 2-DGE spot. In those cases, we used the staining intensity from the PDQuest software as a crude

estimate for the protein with the highest ion score and peptides number detected by MS. If one protein was detected in two or more spots, we used the sum of the intensities from all spots with the protein for an estimate of the total quantity of the protein. We compared the estimated protein amounts with microarray data and found a poor correlation ($R^2=0.0642$) between protein and RNA. Because the estimates for proteins found in the same spots were clearly inaccurate, we also compared the result for 2-DGE spots that only had a single detected protein identified; the resulting correlation was still very poor ($R^2=0.0339$) (Figure 4).

4.4.4 Protein modifications were suggested by 2-D image and MudPIT

Protein modifications can affect the conformation of proteins, thereby affecting protein activities and stability. To obtain evidence for protein modification at a global scale, we took two approaches. First, we examined the proteins identified from 2-DGE/MS and found that 30 proteins were detected in two or more 2-DGE spots (Table 1), suggesting that, for each of these proteins, at least one spot potentially represented a modified form of the protein, although changes *in vitro* cannot be ruled out without further analysis. Among these 30 proteins, some were found at two spots with similar molecular weights but different pI values. For other proteins, the two spots had close pI values but different molecular weights; others corresponded to spots that differed in both molecular weight and pI. In addition, some spots yielded multiple proteins, indicating that several proteins migrated to a small area of the 2-DGE. This is not likely due to a failure to focus proteins in general, because the majority of the protein spots resulted in the identification of a single protein per spot; however, with an estimated 15,000 to 20,000 proteins expressed in any cell type, and even more protein forms expected when posttranslational modifications are included, it is little wonder that at least a good number of the 1000-2000 spots visible on a 2D gel would contain more than one protein. Although some proteins from the same spot showed expected molecular weight and pI, in more cases, proteins with different expected molecular weights and pI values migrated to a same position on the 2-DGE, again suggesting that some of these proteins were modified (Table 1).

Our second approach to obtaining evidence for protein modifications relied on the MudPIT results, which revealed many types of peptide modifications with statistical significance (Confidence > 99%, estimated false discovery rate below 5%). Among these, several could have occurred *in vivo*: including methylation, acetylation, formylation, phosphorylation and

palmitoylation (Table 2); others, such as oxidation, are likely to have occurred during sample preparation and are not represented here. Altogether, 237 modified peptides representing 100 proteins were detected with confidence values of >99% (Table 2), suggesting that these proteins were modified. This provides valuable information for hypothesizing potential regulations of the proteins. Interestingly, among the 30 proteins that were detected at 2 or more spots on the 2-DGE, 12 proteins also had peptide modifications from the MudPIT analysis, further supporting the idea that these proteins were modified.

4.4.5 Putative functions of the detected floral proteins

To obtain clues about possible functions of the proteins identified by our proteomic experiments, we examined their GO (gene ontology) annotations (Figure 5) (Rhee et al., 2003; Poole, 2007). Compared with the GO distributions of the predicted *Arabidopsis* proteome, we found that the percentage of hydrolases, structural proteins and proteins with other enzyme activities (not clearly predicted) are significantly increased. On the other hand, transcription factors, receptors and other molecular functions were under-represented compared to the whole genome. Nevertheless, our results indicate that the proteomic approaches were able to detect some of these putative regulatory proteins.

To further analyze the proteomic data, we looked for biochemical pathways that are significantly represented by the proteomic data using the Aracyc database on the TIR website (Table 3) (Zhang et al., 2005a). Many of the detected pathways are for amino acid synthesis, the TCA cycle, and photosynthesis. In addition, components of methylation pathways and one-carbon unit metabolism are significantly enriched, including a complete set of the SAM (S-Adenosyl Methionine) pathway components, many methyltransferase. The importance of methylation is also supported by the detection of many methylated peptides. Furthermore, enzymes for farnesylation and geranylgeranylation were also found, suggesting that lipid modifications of proteins might be important for flower development (Supplemental Table 3) (Pei et al., 1998; Yalovsky et al., 2000; Caldelari et al., 2001; Brady et al., 2003).

Our results also showed enrichment for proteins involved in protein synthesis, folding and degradation. A large number of ribosomal proteins were identified, including subunits of cytosolic and chloroplast ribosomes. In addition, chaperones important for proper folding and

unfolding of proteins were uncovered, suggesting that they might be important in floral tissues. The components of ubiquitin-dependent proteasome pathway for protein degradation were also highly represented. Besides proteasome subunits, floral proteins involved in ubiquitination included ubiquitin and ubiquitin-like proteins, ubiquitin-specific proteases, E1-ubiquitin activation enzyme, E2-ubiquitin conjugation enzyme, and subunits of ubiquitin ligase such as SKP1 proteins (ASK1 and ASK2) and some F-box proteins.

We also detected several types of structural proteins, including cytoskeleton and motor proteins such as tubulins, actins, dynamin, and kinesins. It was previously thought that membrane proteins are hard to extract and identify. Our results showed that the proteomic procedures were able to detect membrane associated transporters, proteins in the endo-membrane systems and associated with vesicle (SNARE, clathrin etc). Moreover, proteins with putative functions in signaling processes, such as receptors, GTPases and kinases/phosphatase, and in protein-protein interactions were also found. Another group of proteins are those involved in the synthesis and signal transduction of hormones such as auxin, brassinosteroids, JA and ethylene.

Finally, it is worth noting that we detected a number of proteins with genetically defined functions in controlling light signaling and flowering time (CONSTANS, COP1, COP8) (Deng et al., 1991; Chamovitz et al., 1996; Valverde et al., 2004; Liu et al., 2008), floral organ identity (AP2, SUP) (Kunst et al., 1989; Drews et al., 1991; Sakai et al., 1995; Jacobsen and Meyerowitz, 1997; Wurschum et al., 2006), meristem size and floral organ number (ASK1, CAF, CLV1, SUP) (Clark et al., 1997; Yang et al., 1999a; Brand et al., 2000; Zhao et al., 2001; Park et al., 2002; Zhao et al., 2003; Ni et al., 2004; Ogawa et al., 2008), organ shape and polarity (SERRATE) (Lobbes et al., 2006; Yang et al., 2006) and early anther development (SPL) (Schiefthaler et al., 1999; Yang et al., 1999b; Ito et al., 2004). Because mutations in many of these genes dramatically alter early flower development, their functions at late floral stages have not been carefully studied. The proteomic results here indicate that these proteins are still expressed at significant levels during later stages of floral development, suggesting that they function at these stages as well, and further studies are warranted.

4.4.6 Phylogenetic analysis of SAM synthesis pathway components

Our 2-DGE/MS analysis identified a number of proteins involved in the SAM (S-adenosyl methionine) pathway. In particular, we detected all four SAM synthetases encoded by the *Arabidopsis* genome, each of which was supported by isoform-specific peptides. SAM synthetase (SAMS) is a key enzyme in the active methyl cycle, catalyzing the formation of SAM, the major donor of methyl groups for numerous methylation reactions (Loenen, 2006). Although the SAMS enzyme activity is expected to be important and conserved, it is not clear specific SAMS genes are conserved or divergent among plants. It is possible that the four *Arabidopsis* SAMS genes might have different functions. To investigate the evolutionary history of the SAMS genes, we performed phylogenetic analysis of the SAMS genes and their homologs. We used the *Arabidopsis* SAMS protein sequences as queries to search against other plant genomes, including *Populus trichocarpa*, rice, maize, *Selaginella moellendorffii* and *Physcomitrella patens*, and the EST sequences of *Pinus pinaster*. We identified 6 SAMS genes in *Populus*, 3 in rice, 2 in maize, 3 in *Selaginella*, 2 in *Physcomitrella* and 1 in *Chlamydomonas*.

The alignment of protein sequences showed that members of this family have highly similar sequences, with the lowest observed identity of 85% and similarity of 91% (between AT2G36880 and Os01g18860). A preliminary Neighbor-Joining (NJ) tree (not shown) was constructed by using amino acid sequences, but the resolution of the tree is very poor due to the highly conserved protein sequences. Thus the cDNA sequences of these genes were retrieved and used for further analysis. Both the NJ and Maximum Likelihood (ML) tree generated have the same topology (Figure 6). The tree topology indicates that genes from eudicots, monocots, pine, fern and moss all form well supported lineage-specific clades, respectively, suggesting that the most recent common ancestors of angiosperms, seed plants, vascular plants, and land plants all had a single copy of SAMS, which was derived from a single copy in the common ancestor of green plants. A number of duplications are supported by the topology, in eudicots, *Selaginella*, and *Physcomitrella*. In addition, recent duplication events have occurred after the divergence of *Arabidopsis* and *Populus*.

Therefore, the multiple copies in different species have had relatively short history, suggesting that they are still functionally related, consistent with the observed high levels of sequence similarity. At the same time, the two clades in the eudicots have been maintained since before the

divergence of *Arabidopsis* and poplar, estimated to be approximately 100 million years or more, suggesting that these two clades might have diverged functionally distinct, although they still have very highly similar coding regions. In addition, we examine our microarray data (Zhang et al., 2005b) and public data for expression patterns of the four *Arabidopsis* genes and found that they are all highly expressed in flower and also other organs.

SAM is a methyl donor and we also detected two other enzymes in the active methyl cycle, homocysteinase (Zhu et al., 2003) and methionine synthase (Eichel et al., 1995; Ravanel et al., 2004) in our proteomic experiments. To investigate whether genes encoding these two enzymes are also conserved, we also performed phylogenetic studies of the gene families for these two enzymes. Database searches identified, respectively, 2 homocysteinase and 3 methionine synthase encoding genes in *Arabidopsis*, 2 and 4 in *Populus*, 1 and 2 in rice, 1 and 3 in maize, 1 and 1 in *Asparagus*, 2 and 1 in pine, 1 and 2 in *Selaginella*, 2 and 3 in *Physcomitrella* and 1 of each in *Chlamydomonas*. Similar to SAM synthetase, these two protein families are also quite well conserved. The NJ and ML trees of both families showed similar topologies to the SAM synthetase family, but the homocysteinase family has fewer duplication events within monocots and eudicots compared with the other two. The microarray data also indicated a high expression level of all the *Arabidopsis* genes in the two families except the methionine synthase encoding gene AT5G20980, which was expressed at a very low level. The phylogenetic tree showed that AT5G20980 has an obvious longer branch compared with other genes, suggesting that it has evolved relatively rapidly.

In summary, our phylogenetic analyses indicate that genes encoding key enzymes in the active methyl cycle have been conserved through much of the history of land plants; it is likely that these genes have important functions that have been maintained during land plant evolution. Although there also have been some recent duplications within the eudicots and/or monocots lineages, most of the duplicated genes are highly similar in sequence and expression, suggesting that they still have similar functions.

4.5 DISCUSSION

4.5.1 A combination of proteomics methods and analyses allows increased detection of floral proteins.

We showed here that 2-DGE and MudPIT experiments can uncover a portion of the floral proteome and that the identification results from these proteomic approaches are partially overlapping. The results indicate that both proteomic methods sample the proteome, but cannot detect all or even most proteins in one experiment (it is estimated that any one cell type in higher eukaryotic cells expresses about 15,000 – 20,000 different proteins, not counting post-translational modifications). Although 2-DGE is thought to favor abundant proteins, it is intriguing to note that some of the proteins detected by the 2-DGE analysis were missed by both of the MudPIT experiments, suggesting that they might be present at moderate levels. This is supported by the observation that even proteins of very faint 2-DGE spots can be detected by MS, indicating that this approach is quite sensitive. One advantage of 2-DGE is that it provides information about the size and pI of the whole protein, as well as a rough estimate of the abundance. However, on a per protein basis, 2-DGE is more laborious and expensive than MudPIT, which allows the detection and identification of more than a thousand proteins in one experiment, albeit without the whole protein level information, such as pI, approximate molecular weight, and abundance that can be provided by whole protein gel separation methods. In addition, 2-DGE requires the proteins to be extracted from the cells and separated on the 2-DGE; in contrast, MudPIT experiments allow crude extracts to be digested with trypsin and the total peptide pools are then analyzed by 2-D LC and MS. The procedural difference might also contribute to the difference in the identified proteins.

The MudPIT approach is quite similar in a general way to the analysis of the transcriptome using Expressed Sequence Tag (EST) analysis, which samples from a cDNA library by sequencing random clones and genes with larger transcript numbers are sampled more frequently. Therefore, the fact that there is considerable difference in the sets of proteins identified by the two MudPIT experiments strongly suggests that the total floral proteome is much larger than the ~2400 proteins we have uncovered. From microarray studies, it was estimated that more than 14,000 genes are expressed in the Arabidopsis flower. Many of these genes are expressed at low levels; our results indicate that even the products of genes with very low mRNA levels can be detected but at much lower frequencies than the products of highly expressed genes. Therefore, additional fractionation and enrichment procedures, as well as multiple MudPIT experiments with the different enriched fractions, would likely be needed to detect the lower abundance proteins.

We also found that analysis of the MS data from MudPIT experiments using two different algorithms (ProteinPilot and Mascot) with two protein databases (SwissProt and NCBI) yielded overlapping but non-identical protein sets. With either algorithm, significantly greater numbers of non-redundant protein identifications were possible when searching against the *A. thaliana* entries in the NCBI database than when searching against the *A. thaliana* entries in the SwissProt database, suggesting that *A. thaliana* sequences may be under-represented in the SwissProt database; there may also be some difference in the way that information regarding the predicted *Arabidopsis* proteome is organized in the databases, and of course the two algorithms differ in the way the MS/MS data are analyzed and scored. Our experience from the analysis of the *Arabidopsis* floral proteome indicate that when performing a proteome-wide survey, it is beneficial to analyze the data using multiple algorithms/databases, as previously suggested by Resing and others (Resing et al., 2004), in order to maximize the detection of the proteins present in the tissue. Further studies will likely provide better understanding of both the experimental and the bioinformatic components of proteomic efforts.

To learn whether the proteins we detected are also expressed in other parts of the plant, we compared our results with previous proteomic data from *Arabidopsis*. Holmes-Davis et al. (2005) reported the identification of 135 proteins from mature pollen using 2-DG/MS; 82 of these proteins were detected by our MudPIT experiments, consistent with the fact that our samples of stage-12 flowers contained mature pollen. In a recently study, 2342 proteins were identified from leaves (Lee et al., 2007). Among these, 1109 were also identified by our MudPIT experiments, indicating that many proteins are shared between leaves and flowers. In addition, more than a third (145 out of 402; Carter et al., 2004) of detected vegetative vacuolar proteins and about one sixth (119/689; Chibani et al., 2006) of seed dormancy-related proteins were found in our studies. Therefore, the floral proteome observed in our studies showed substantial overlap with proteins uncovered from other organs, suggesting that many proteins perform functions that are common in the flower and other parts of the plant, in agreement with mRNA expression data (e.g., Zhang et al., 2005).

4.5.2 Evidence for post-translational modifications in the *Arabidopsis* floral proteome

Protein modifications are important for cellular processes and cannot be detected by mRNA level analysis. We have obtained strong evidence for multiple protein modifications in the *Arabidopsis*

flower. First, the distribution pattern of some of the proteins on 2-DGE suggests that they had multiple forms showing different molecular weights and pI values. Second, the MS/MS analysis of the peptides released from these proteins directly suggest the type and position of the covalent modifications (Tables 1 and 2). The MudPIT data revealed many modifications that potentially had occurred *in vivo*, such as formylation, methylation, myristoylation, and acetylation, all with strong confidence scores (> 99%) indicating in the modified peptides. Thus, proteomic methods can directly detect such modifications and provide information regarding the possible regulation of modified proteins, which might play important roles in many pathways.

As an example of one observed post-translational modification, methylation is a covalent attachment of a methyl group to target molecules. In plants, the primary donor of methyl group is SAM, which is generated by SAM cycle (Loenen, 2006). A broad range of methyltransferases recognize different target molecules and transfer the methyl group from SAM to the target sites. In plants, methylation on DNA, RNA, protein or small molecules has been shown to be important. DNA methylation is the most well studied and is implicated in the remodeling of chromatin structure and repression of gene expression. In the *Arabidopsis* flower, DNA methylation has been shown to be involved in the vernalization pathway for flowering (Amasino, 2004; Schmitz et al., 2008), the regulation of the *SUPERMAN* gene (Sakai et al., 1995; Jacobsen and Meyerowitz, 1997; Lindroth et al., 2001; Cao and Jacobsen, 2002; Jackson et al., 2002), genomic imprinting and embryo developments (Grossniklaus et al., 2001; Jullien et al., 2006; Baroux et al., 2007; Huh et al., 2007). Reduction in DNA methylation can cause complex abnormal phenotypes (Jacobsen and Meyerowitz, 1997; Jacobsen et al., 2000). RNA methylation is related to RNA stability. For example, microRNA stability is regulated by the *HEN1* RNA methyltransferase (Li et al., 2005; Yu et al., 2005). Protein methylation such as histone methylation is also involved many regulatory processes (Johnson et al., 2002; Schmitz et al., 2008). In addition, small molecules such as plant hormones (auxin, gibberilline, salicylic acid and jasmonic acid) are also modified by methylation (Seo et al., 2001; Cheong and Choi, 2003; Zubieta et al., 2003; Qin et al., 2005; Varbanova et al., 2007).

We have detected several SAM cycle enzymes, including all four SAMS. Sequence similarity and phylogenetic analysis suggest that these proteins have highly conserved functions

during plant evolution. At the same time, there are several duplications in major angiosperm groups, including one that occurred before the split of *Arabidopsis* and poplar, suggesting that they might have non-identical functions. Similarly, the genes encoding methionine synthases also have diverged prior to the *Arabidopsis*-poplar divergence, but homocysteinase genes only have more recent duplications. It is thought that duplicated genes allow functional divergence, which in turn is important for the maintenance of duplicated genes. However, the duplicated SAM cycle genes remain highly similar in sequence and expression, suggesting that they might still have the same or very similar functions. One reason multiple genes with very similar functions might be maintained is that additional copies can promote very high level expression, as supported by the observation that they are expressed at high levels in different organs. If future experiments can show that these SAM cycle enzymes indeed have the same or very similar functions, and multiple genes are needed to achieve high level expression, this would be an excellent example of maintenance of functionally similar duplicated genes.

We have also detected different types of methyltransferases. The presence of multiple methyltransferases suggests that many types of molecules are potentially modified by methylation. Moreover, we have also detected many methylated peptides, with one, two, or three methyl groups, suggesting that protein methylation in the flower is quite complex and highly active. Study of the biological significance of the methylations of specific proteins will likely lead to new insights about the regulation of protein activities.

In conclusion, we present a proteomic study of the *Arabidopsis* flower. We illustrate that multiple experiments and analyses can lead to relatively deep sampling of the proteome, and that proteins with varying abundance as estimated from staining intensity can be detected. We further present strong evidence for protein modification, including methylation, and suggest that they are important for normal flower development and physiology. Our study opens a new door for the understanding of the flower and should stimulate future research in this fertile area.

Table 4.1 Protein with multiple spots on 2-D and modifications detected by MudPIT

AGI#	Spot #	Modification shown by MudPIT
AT1G01050	22, 23	
AT1G04410	14, 1C03	Methyl(E)@2
AT1G07890	23, 1C01	
AT1G19570	38, 1B07	
AT1G53240	10, 1G04	Trimethyl(K)@3; Dethiomethyl(M)@10
AT1G67090	68, 29, 67, 79	Formyl@N-term; Methyl(L)@5; Methyl(S)@6; Methyl(Q)@2
AT2G21660	60, 1B05	Methyl(S)@8; Methyl(S)@13, Protein Terminal Acetyl@N-term; Dimethyl(R)@16, Formyl@N-term; Dimethyl(R)@9
AT2G24270	55-1, 11, 55-3, 1F11	Protein Terminal Acetyl@N-term; Formyl@N-term
AT2G37220	17, 18	Acetyl(S)@20; Arg->GluSA(R)@23; Formyl@N-term
AT2G39730	1A02, 3, 56-58, 1E08	Formyl@N-term
AT2G46860	22, 23	
AT2G47470	23, 1B11	
AT2G47730	65, 1G11	
AT3G09200	61, 2C01	
AT3G25050	15, 1A04	Formyl@N-term; Phospho(S)@8;
AT3G52930	1C04, 2B07	
AT3G53460	60, 61	
AT3G55440	23, 28, 2C02	
AT4G03280	48, 1A06	Formyl@N-term
AT4G04460	64, 1G10	
AT4G08390	63-1, 63-2	
AT4G18480	79, 56-60	
AT4G31300	25, 80	
AT4G38680	28, 29, 80	
AT4g38970	5, 1F03	Trimethyl(K)@6
AT5G08690	52, 56-60, 1B08	Dimethyl(K)@18, Oxidation(P)@2; Methyl(S)@3; Formyl(K)@12
AT5G38410	67, 50, 73, 68	Formyl(K)@1, Formyl@N-term
AT5G48480	36, 60	Protein Terminal Acetyl@N-term
AT5G66570	62, 1A11, 2B08, 2B09	Formyl@N-term
ATCG00490	12, 16, 19, 78, 11, 23, 31, 32, 60, 61, 1A05, 1B02, 1B12, 1C02, 1D10, 1F06, 2B11, 1E04	Methyl(L)@10; Methyl(H)@2; Methyl(D)@5 Methyl(H)@7 Acetyl@N-term; Methyl(T)@3; Methyl(H)@3 Methyl(D)@6; Hex(1)HexNAc(1)(T)@9; Methyl(L)@5; Formyl@N-term; Methyl(R)@11 Methyl(N)@6; Methyl(S)@7 Methyl(T)@1

Table 4.2 Number of peptide modifications detected by MudPIT with 99% confidence

Type of peptide modifications	Number of peptides
Formylation	52
Acetylation	7
Trimethylation	7
Dimethylation	24
Methylation	141
HexNAC	5
Myristoylation	3

Table 4.3 Important categories of proteins identified by MudPIT

Category	Molecular function	Number of proteins identified
Proteolysis	UBQ	8
	E1	4
	E2	5
	F-box	2
	SKIP	2
	Cullin	1
	20S proteasome	7
	26S proteasome	16
	Proteases	21
	30S subunits	4
Ribosome	50S subunits	4
	40S subunits	58
	60S subunits	91
Membrane	Plasma and other membrane transporters / integral proteins (ABC transporter, aquaporin, ion channel or transporter (H ⁺ , K ⁺ , Ca ²⁺), organic molecule transporter etc.)	58
	nuclear pore, importin etc.	4
	Vacuole	6
	Golgi, Vesicle transport	19
	actin	9
Cytoskeleton and motors	tubulin	13
	dynamitin	6
	kinesin	3
	myosin	5
	villin	2
Kinase/phosphatase	Receptor like	5
	MAP Kinases	4
	PI pathway kinases	3
	Other kinases	14
	phosphatases	8
Hormone related	Auxin response	4
	ABA	2
	Ethylene	2
	JA	1
	Transcriptional factor	5
Important known floral regulators	small RNA biogenesis	2
	light signalling/protein degradation	3
	Receptor like kinase	1

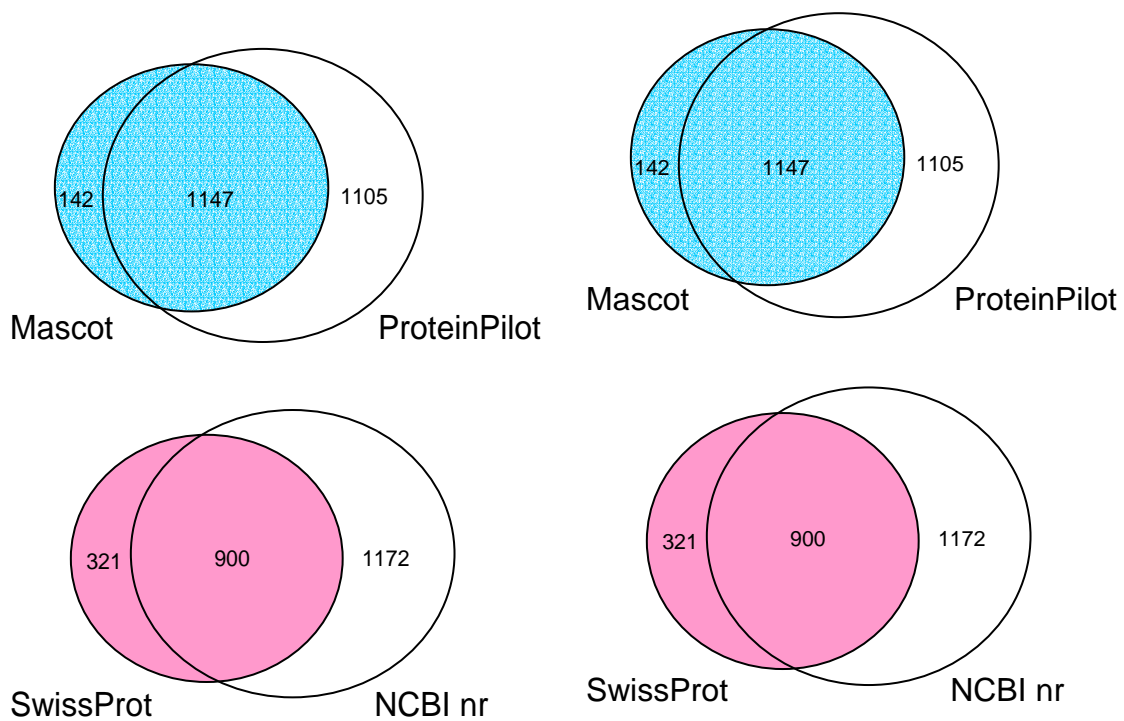


Figure 4.2 Overlap between the lists of protein identifications by 2-D/MS and MudPITs and the comparison between two MudPIT experiments. The number at each area represents the number of proteins identified.

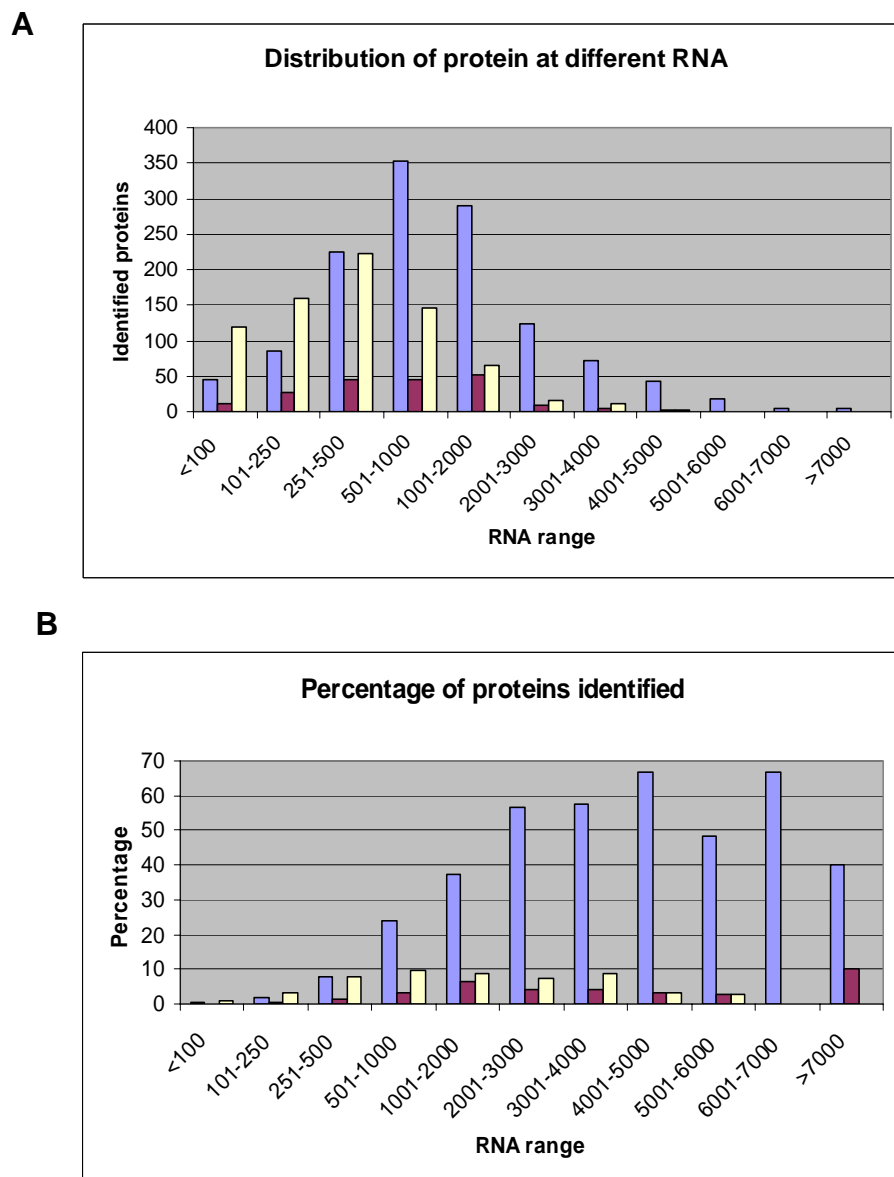


Figure 4.3 The distribution of proteins identified by MudPIT at different RNA ranges. The RNA abundance was determined by ATH1 microarray experiments. A. the actual number of proteins identified at different RNA ranges were compared among three type of identifications: the identifications by both MudPIT experiments, by only one MudPIT and two or more analysis, and by only one analysis. B. the percentage of proteins identified at different RNA ranges were compared as above. The percentage is taken by dividing the number of identified protein with the total number of genes of which the mRNA is present as suggested by microarray.

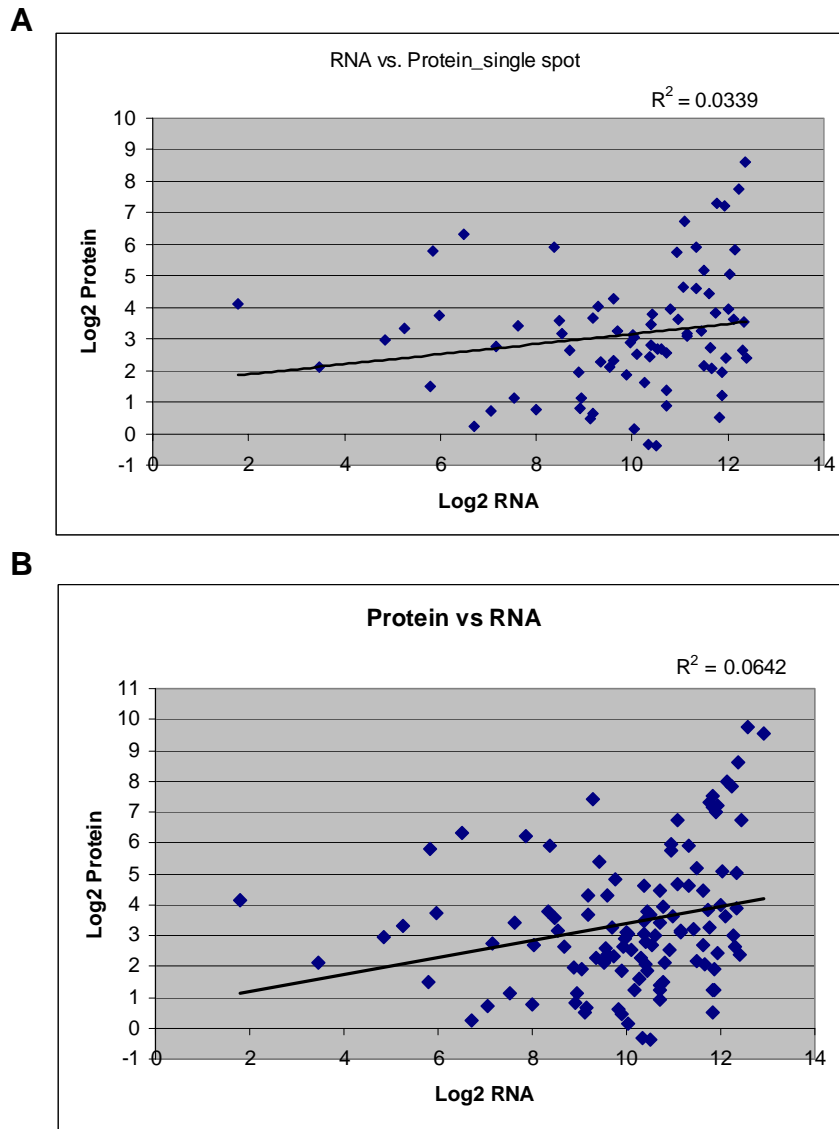


Figure 4.4 Comparisons between the RNA and protein abundance.

A. The log ratio of RNA (detected by microarray) and the log ratio of protein abundance (spot intensity detected by 2-D gel and normalized with molecular weight) was compared for the spots with single protein identified.

B. A similar comparison was done by using all spots from 2-D gel image. For the spots with multiple protein identified, the spot intensity was used as an estimation of the highest represented protein in that spot (highest identified peptide number and ion score value).

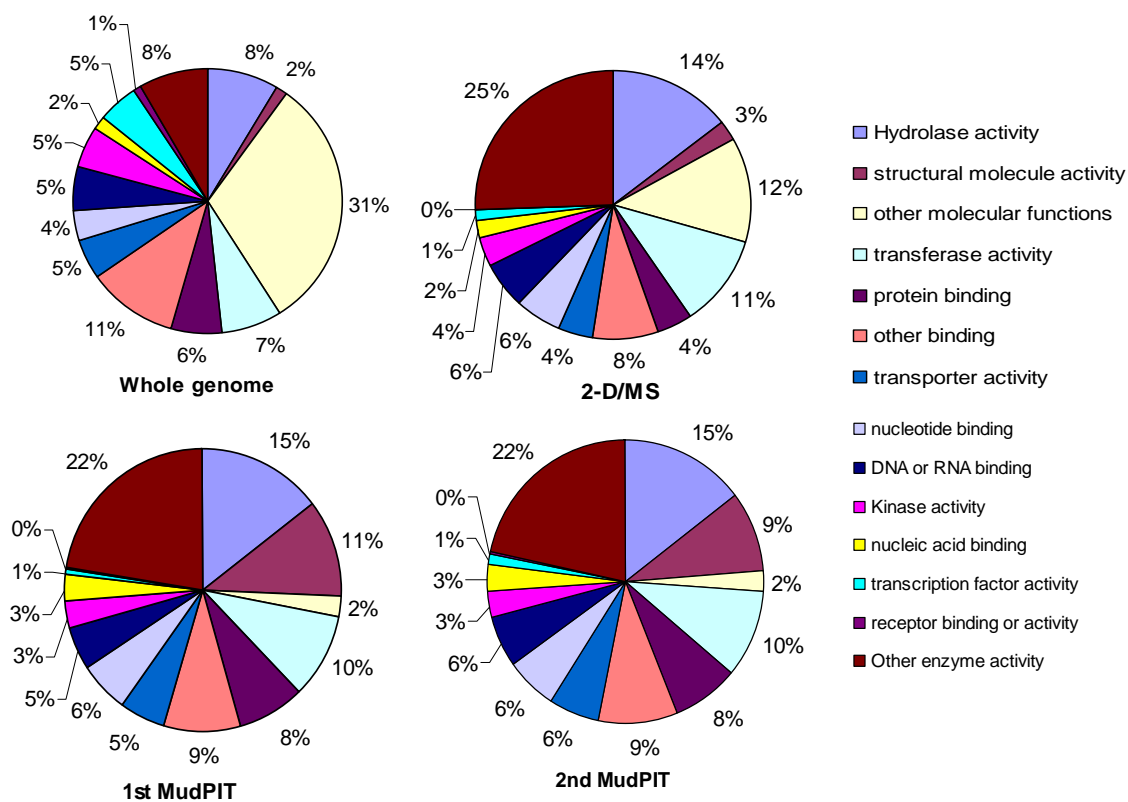
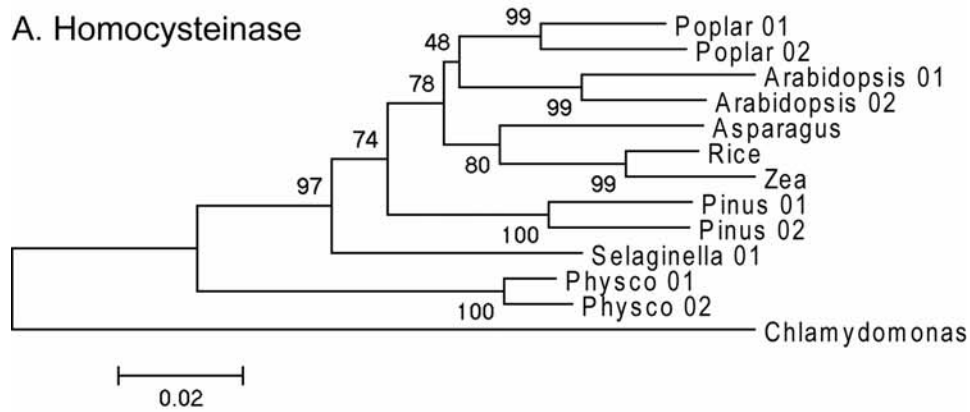
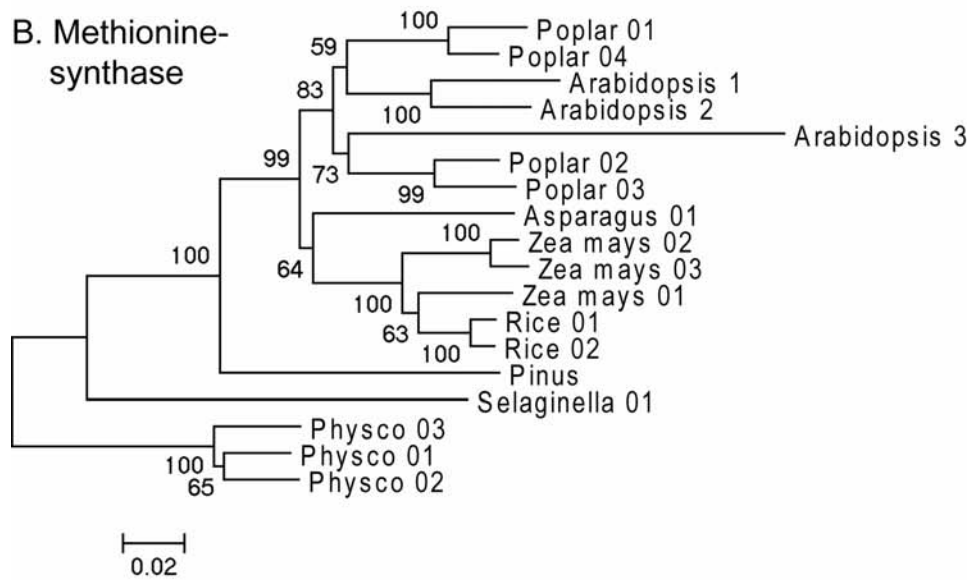


Figure 4.5 GO categories of the proteins identified by 2-D MS and MudPIT. The GO categories were based on molecular function as suggested by the TAIR website.

A. Homocysteinase



B. Methionine-synthase



C. SAM synthetase

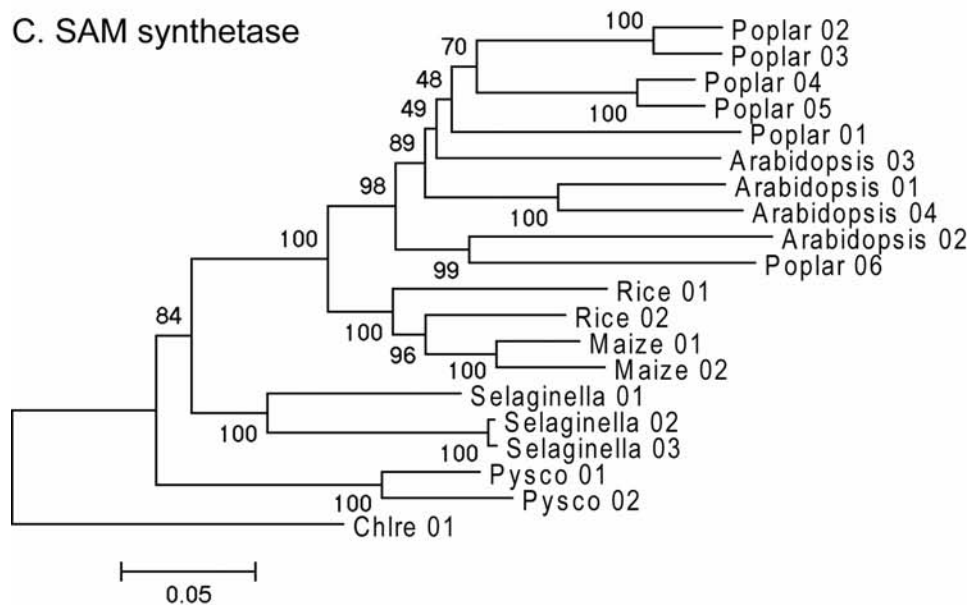


Figure 4.6 Phylogenetic trees of SAM synthetase (A), homocysteinase (B), and methionine synthase (C) gene families. The trees are constructed by NJ method based on the alignment of CDS sequences.

4.6 References

- Abascal, F., R. Zardoya and D. Posada** (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman** (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
- Amasino, R.** (2004). Vernalization, competence, and the epigenetic memory of winter. *Plant Cell* **16**, 2553-2559.
- Arabidopsis Genome Initiative.** (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815.
- Baroux, C., S. Pien and U. Grossniklaus** (2007). Chromatin modification and remodeling during early seed development. *Curr. Opin. Genet. Dev.* **17**, 473-479.
- Bayer, E. M., A. R. Bottrill, J. Walshaw, M. Vigouroux, M. J. Naldrett, C. L. Thomas and A. J. Maule** (2006). *Arabidopsis* cell wall proteome defined using multidimensional protein identification technology. *Proteomics* **6**, 301-311.
- Brady, S. M., S. F. Sarkar, D. Bonetta and P. McCourt** (2003). The *ABSCISIC ACID INSENSITIVE 3 (ABI3)* gene is modulated by farnesylation and is involved in auxin signaling and lateral root development in *Arabidopsis*. *Plant J.* **34**, 67-75.
- Brand, U., J. C. Fletcher, M. Hobe, E. M. Meyerowitz and R. Simon** (2000). Dependence of stem cell fate in *Arabidopsis* on a feedback loop regulated by CLV3 activity. *Science* **289**, 617-619.
- Caldelari, D., H. Sternberg, M. Rodriguez-Concepcion, W. Gruissem and S. Yalovsky** (2001). Efficient prenylation by a plant geranylgeranyltransferase-I requires a functional CaaL box motif and a proximal polybasic domain. *Plant Physiol.* **126**, 1416-1429.
- Cao, X. and S. E. Jacobsen** (2002). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr. Biol.* **12**, 1138-1144.

- Carter, C., S. Pan, J. Zouhar, E. L. Avila, T. Girke and N. V. Raikhel** (2004). The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. *Plant Cell* **16**, 3285-3303.
- Chamovitz, D. A., N. Wei, M. T. Osterlund, A. G. von Arnim, J. M. Staub, M. Matsui and X. W. Deng** (1996). The COP9 complex, a novel multisubunit nuclear regulator involved in light control of a plant developmental switch. *Cell* **86**, 115-121.
- Cheong, J. J. and Y. D. Choi** (2003). Methyl jasmonate as a vital substance in plants. *Trends Genet.* **19**, 409-413.
- Chibani, K., S. Ali-Rachedi, C. Job, D. Job, M. Jullien and P. Grappin** (2006). Proteomic analysis of seed dormancy in *Arabidopsis*. *Plant Physiol.* **142**, 1493-1510.
- Clark, S. E., S. E. Jacobsen, J. Z. Levin and E. M. Meyerowitz** (1996). The *CLAVATA* and *SHOOT MERISTEMLESS* loci competitively regulate meristem activity in *Arabidopsis*. *Development* **122**, 1567-1575.
- Clark, S. E., R. W. Williams and E. M. Meyerowitz** (1997). The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* **89**, 575-585.
- Delahunty, C. M. and J. R. Yates, 3rd** (2007). MudPIT: multidimensional protein identification technology. *Biotechniques* **43**, 563, 565, 567 passim.
- Deng, X. W., T. Caspar and P. H. Quail** (1991). *cop1*: a regulatory locus involved in light-controlled development and gene expression in *Arabidopsis*. *Genes Dev.* **5**, 1172-1182.
- DeYoung, B. J., K. L. Bickle, K. J. Schrage, P. Muskett, K. Patel and S. E. Clark** (2006). The *CLAVATA1*-related *BAM1*, *BAM2* and *BAM3* receptor kinase-like proteins are required for meristem function in *Arabidopsis*. *Plant J.* **45**, 1-16.
- Drews, G. N., J. L. Bowman and E. M. Meyerowitz** (1991). Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell* **65**, 991-1002.
- Edgar, R. C.** (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797.
- Eichel, J., J. C. Gonzalez, M. Hotze, R. G. Matthews and J. Schroder** (1995). Vitamin-B12-independent methionine synthase from a higher plant (*Catharanthus roseus*). Molecular characterization, regulation, heterologous expression, and enzyme properties. *Eur. J. Biochem.* **230**, 1053-1058.

- Florens, L. and M. P. Washburn** (2006). Proteomic analysis by multidimensional protein identification technology. *Methods Mol. Biol.* **328**, 159-175.
- Futcher, B., G. I. Latter, P. Monardo, C. S. McLaughlin and J. I. Garrels** (1999). A sampling of the yeast proteome. *Mol. Cell Biol.* **19**, 7357-7368.
- Gallardo, K., C. Job, S. P. Groot, M. Puype, H. Demol, J. Vandekerckhove and D. Job** (2001). Proteomic analysis of *Arabidopsis* seed germination and priming. *Plant Physiol.* **126**, 835-848.
- Griffin, T. J., S. P. Gygi, T. Ideker, B. Rist, J. Eng, L. Hood and R. Aebersold** (2002). Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol. Cell Proteomics* **1**, 323-333.
- Grossniklaus, U., C. Spillane, D. R. Page and C. Kohler** (2001). Genomic imprinting and seed development: endosperm formation with and without sex. *Curr. Opin. Plant Biol.* **4**, 21-27.
- Guindon, S. and O. Gascuel** (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696-704.
- Gygi, S. P., Y. Rochon, B. R. Franza and R. Aebersold** (1999). Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **19**, 1720-1730.
- Holmes-Davis, R., C. K. Tanaka, W. H. Vensel, W. J. Hurkman and S. McCormick** (2005). Proteome mapping of mature pollen of *Arabidopsis thaliana*. *Proteomics* **5**, 4864-4884.
- Huh, J. H., M. J. Bauer, T. F. Hsieh and R. Fischer** (2007). Endosperm gene imprinting and seed development. *Curr. Opin. Genet. Dev.* **17**, 480-485.
- Ito, T., F. Wellmer, H. Yu, P. Das, N. Ito, M. Alves-Ferreira, J. L. Riechmann and E. M. Meyerowitz** (2004). The homeotic protein AGAMOUS controls microsporogenesis by regulation of *SPOROCTELESS*. *Nature* **430**, 356-360.
- Jackson, J. P., A. M. Lindroth, X. Cao and S. E. Jacobsen** (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**, 556-560.
- Jacobsen, S. E. and E. M. Meyerowitz** (1997). Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. *Science* **277**, 1100-1103.
- Jacobsen, S. E., H. Sakai, E. J. Finnegan, X. Cao and E. M. Meyerowitz** (2000). Ectopic hypermethylation of flower-specific genes in *Arabidopsis*. *Curr. Biol.* **10**, 179-186.
- Jarvis, P.** (2004). Organellar proteomics: chloroplasts in the spotlight. *Curr. Biol.* **14**, R317-319.

- Johnson, L., X. Cao and S. Jacobsen** (2002). Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr. Biol.* **12**, 1360-1367.
- Jorin, J. V., A. M. Maldonado and M. A. Castillejo** (2007). Plant proteome analysis: A 2006 update. *Proteomics* **7**, 2947-2962.
- Jullien, P. E., A. Katz, M. Oliva, N. Ohad and F. Berger** (2006). Polycomb group complexes self-regulate imprinting of the Polycomb group gene *MEDEA* in *Arabidopsis*. *Curr. Biol.* **16**, 486-492.
- Kunst, L., J. E. Klenz, J. Martinez-Zapater and G. W. Haughn** (1989). *AP2* Gene Determines the Identity of Perianth Organs in Flowers of *Arabidopsis thaliana*. *Plant Cell* **1**, 1195-1208.
- Lee, J., W. M. Garrett and B. Cooper** (2007). Shotgun proteomic analysis of *Arabidopsis thaliana* leaves. *J. Sep. Sci.* **30**, 2225-2230.
- Li, J., Z. Yang, B. Yu, J. Liu and X. Chen** (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr. Biol.* **15**, 1501-1507.
- Lindroth, A. M., X. Cao, J. P. Jackson, D. Zilberman, C. M. McCallum, S. Henikoff and S. E. Jacobsen** (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* **292**, 2077-2080.
- Liu, L. J., Y. C. Zhang, Q. H. Li, Y. Sang, J. Mao, H. L. Lian, L. Wang and H. Q. Yang** (2008). COP1-Mediated Ubiquitination of CONSTANS Is Implicated in Cryptochrome Regulation of Flowering in *Arabidopsis*. *Plant Cell*.
- Lobbes, D., G. Rallapalli, D. D. Schmidt, C. Martin and J. Clarke** (2006). SERRATE: a new player on the plant microRNA scene. *EMBO Rep.* **7**, 1052-1058.
- Loenen, W. A.** (2006). S-adenosylmethionine: jack of all trades and master of everything? *Biochem. Soc. Trans.* **34**, 330-333.
- Ma, H.** (2005). Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annu. Rev. Plant Biol.* **56**, 393-434.
- Majeran, W., Y. Cai, Q. Sun and K. J. van Wijk** (2005). Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. *Plant Cell* **17**, 3111-3140.

- Maor, R., A. Jones, T. S. Nuhse, D. J. Studholme, S. C. Peck and K. Shirasu** (2007). Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. *Mol. Cell Proteomics* **6**, 601-610.
- Ni, W., D. Xie, L. Hobbie, B. Feng, D. Zhao, J. Akkara and H. Ma** (2004). Regulation of flower development in *Arabidopsis* by SCF complexes. *Plant Physiol.* **134**, 1574-1585.
- Nuhse, T. S., A. Stensballe, O. N. Jensen and S. C. Peck** (2004). Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database. *Plant Cell* **16**, 2394-2405.
- Ogawa, M., H. Shinohara, Y. Sakagami and Y. Matsubayashi** (2008). *Arabidopsis* CLV3 peptide directly binds CLV1 ectodomain. *Science* **319**, 294.
- Park, W., J. Li, R. Song, J. Messing and X. Chen** (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**, 1484-1495.
- Peck, S. C., T. S. Nuhse, D. Hess, A. Iglesias, F. Meins and T. Boller** (2001). Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. *Plant Cell* **13**, 1467-1475.
- Peck, S. C.** (2005). Update on proteomics in *Arabidopsis*. Where do we go from here? *Plant Physiol.* **138**, 591-599.
- Pei, Z. M., M. Ghassemian, C. M. Kwak, P. McCourt and J. I. Schroeder** (1998). Role of farnesyltransferase in ABA regulation of guard cell anion channels and plant water loss. *Science* **282**, 287-290.
- Perkins, D. N., D. J. Pappin, D. M. Creasy and J. S. Cottrell** (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567.
- Poole, R. L.** (2007). The TAIR Database. *Methods Mol. Biol.* **406**, 179-212.
- Posada, D. and K. A. Crandall** (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.
- Pradet-Balade, B., F. Boulme, H. Beug, E. W. Mullner and J. A. Garcia-Sanz** (2001). Translation control: bridging the gap between genomics and proteomics? *Trends Biochem. Sci.* **26**, 225-229.

- Qin, G., H. Gu, Y. Zhao, Z. Ma, G. Shi, Y. Yang, E. Pichersky, H. Chen, M. Liu, Z. Chen and L. J. Qu** (2005). An indole-3-acetic acid carboxyl methyltransferase regulates *Arabidopsis* leaf development. *Plant Cell* **17**, 2693-2704.
- Rajjou, L., M. Belghazi, R. Huguet, C. Robin, A. Moreau, C. Job and D. Job** (2006). Proteomic investigation of the effect of salicylic acid on *Arabidopsis* seed germination and establishment of early defense mechanisms. *Plant Physiol.* **141**, 910-923.
- Ravanel, S., M. A. Block, P. Rippert, S. Jabrin, G. Curien, F. Rebeille and R. Douce** (2004). Methionine metabolism in plants: chloroplasts are autonomous for *de novo* methionine synthesis and can import S-adenosylmethionine from the cytosol. *J. Biol. Chem.* **279**, 22548-22557.
- Resing, K. A., K. Meyer-Arendt, A. M. Mendoza, L. D. Aveline-Wolf, K. R. Jonscher, K. G. Pierce, W. M. Old, H. T. Cheung, S. Russell, J. L. Wattawa, G. R. Goehle, R. D. Knight and N. G. Ahn** (2004). Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal. Chem.* **76**, 3556-3568.
- Rhee, S. Y., W. Beavis, T. Z. Berardini, G. Chen, D. Dixon, A. Doyle, M. Garcia-Hernandez, E. Huala, G. Lander, M. Montoya, N. Miller, L. A. Mueller, S. Mundodi, L. Reiser, J. Tacklind, D. C. Weems, Y. Wu, I. Xu, D. Yoo, J. Yoon and P. Zhang** (2003). The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res.* **31**, 224-228.
- Ru, Q. C., L. A. Zhu, R. A. Katenhusen, J. Silberman, H. Brzeski, M. Liebman and C. D. Shriver** (2006). Exploring human plasma proteome strategies: high efficiency in-solution digestion protocol for multi-dimensional protein identification technology. *J. Chromatogr. A.* **1111**, 175-191.
- Sakai, H., L. J. Medrano and E. M. Meyerowitz** (1995). Role of SUPERMAN in maintaining *Arabidopsis* floral whorl boundaries. *Nature* **378**, 199-203.
- Samach, A., J. E. Klenz, S. E. Kohalmi, E. Risseuw, G. W. Haughn and W. L. Crosby** (1999). The *UNUSUAL FLORAL ORGANS* gene of *Arabidopsis thaliana* is an F-box protein required for normal patterning and growth in the floral meristem. *Plant J.* **20**, 433-445.
- Schieffthaler, U., S. Balasubramanian, P. Sieber, D. Chevalier, E. Wisman and K. Schneitz** (1999). Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early

- sporogenesis during sex organ development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **96**, 11664-11669.
- Schmidt, M. W., A. Houseman, A. R. Ivanov and D. A. Wolf** (2007). Comparative proteomic and transcriptomic profiling of the fission yeast *Schizosaccharomyces pombe*. *Mol. Syst. Biol.* **3**, 79.
- Schmitz, R. J., S. Sung and R. M. Amasino** (2008). Histone arginine methylation is required for vernalization-induced epigenetic silencing of FLC in winter-annual *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **105**, 411-416.
- Schoof, H., M. Lenhard, A. Haecker, K. F. Mayer, G. Jurgens and T. Laux** (2000). The stem cell population of *Arabidopsis* shoot meristems is maintained by a regulatory loop between the *CLAVATA* and *WUSCHEL* genes. *Cell* **100**, 635-644.
- Seo, H. S., J. T. Song, J. J. Cheong, Y. H. Lee, Y. W. Lee, I. Hwang, J. S. Lee and Y. D. Choi** (2001). Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proc. Natl. Acad. Sci. USA* **98**, 4788-4793.
- Shilov, I. V., S. L. Seymour, A. A. Patel, A. Loboda, W. H. Tang, S. P. Keating, C. L. Hunter, L. M. Nuwaysir and D. A. Schaeffer** (2007). The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell Proteomics* **6**, 1638-1655.
- Smyth, D. R., J. L. Bowman and E. M. Meyerowitz** (1990). Early flower development in *Arabidopsis*. *Plant Cell* **2**, 755-767.
- Tamura, K., J. Dudley, M. Nei and S. Kumar** (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596-1599.
- Thelen, J. J. and S. C. Peck** (2007). Quantitative proteomics in plants: choices in abundance. *Plant Cell* **19**, 3339-3346.
- Valverde, F., A. Mouradov, W. Soppe, D. Ravenscroft, A. Samach and G. Coupland** (2004). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science* **303**, 1003-1006.
- Varbanova, M., S. Yamaguchi, Y. Yang, K. McKelvey, A. Hanada, R. Borochoy, F. Yu, Y. Jikumaru, J. Ross, D. Cortes, C. J. Ma, J. P. Noel, L. Mander, V. Shulaev, Y. Kamiya, S. Rodermel, D. Weiss and E. Pichersky** (2007). Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. *Plant Cell* **19**, 32-45.

- Washburn, M. P., D. Wolters and J. R. Yates, 3rd** (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242-247.
- Washburn, M. P., R. R. Ulaszek and J. R. Yates, 3rd** (2003). Reproducibility of quantitative proteomic analyses of complex biological mixtures by multidimensional protein identification technology. *Anal. Chem.* **75**, 5054-5061.
- Wurschum, T., R. Gross-Hardt and T. Laux** (2006). APETALA2 regulates the stem cell niche in the *Arabidopsis* shoot meristem. *Plant Cell* **18**, 295-307.
- Yalovsky, S., M. Rodriguez-Concepcion, K. Bracha, G. Toledo-Ortiz and W. Gruissem** (2000). Prenylation of the floral transcription factor *APETALA1* modulates its function. *Plant Cell* **12**, 1257-1266.
- Yang, L., Z. Liu, F. Lu, A. Dong and H. Huang** (2006). SERRATE is a novel nuclear regulator in primary microRNA processing in *Arabidopsis*. *Plant J.* **47**, 841-850.
- Yang, M., Y. Hu, M. Lodhi, W. R. McCombie and H. Ma** (1999a). The *Arabidopsis* *SKP1-LIKE1* gene is essential for male meiosis and may control homologue separation. *Proc. Natl. Acad. Sci. USA* **96**, 11416-11421.
- Yang, W. C., D. Ye, J. Xu and V. Sundaresan** (1999b). The *SPOROCTELESS* gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev.* **13**, 2108-2117.
- Yu, B., Z. Yang, J. Li, S. Minakhina, M. Yang, R. W. Padgett, R. Steward and X. Chen** (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932-935.
- Zahn, L., B. M. Feng and H. Ma** (2006). Beyond the ABC model: regulation of floral homeotic genes. *Advances in Botanical Research*. D. E. Soltis, J. H. Leebens-Mack and P. S. Soltis, Elsevier. **44**, 161-205.
- Zhang, P., H. Foerster, C. P. Tissier, L. Mueller, S. Paley, P. D. Karp and S. Y. Rhee** (2005a). MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiol.* **138**, 27-37.
- Zhang, X., B. Feng, Q. Zhang, D. Zhang, N. Altman and H. Ma** (2005b). Genome-wide expression profiling and identification of gene activities during early flower development in *Arabidopsis*. *Plant Mol. Biol.* **58**, 401-419.

- Zhao, D., M. Yang, J. Solava and H. Ma** (1999). The *ASK1* gene regulates development and interacts with the *UFO* gene to control floral organ identity in *Arabidopsis*. *Dev. Genet.* **25**, 209-223.
- Zhao, D., Q. Yu, M. Chen and H. Ma** (2001). The *ASK1* gene regulates B function gene expression in cooperation with *UFO* and *LEAFY* in *Arabidopsis*. *Development* **128**, 2735-2746.
- Zhao, D., W. Ni, B. Feng, T. Han, M. G. Petrasek and H. Ma** (2003). Members of the *Arabidopsis-SKPI-like* gene family exhibit a variety of expression patterns and may play diverse roles in *Arabidopsis*. *Plant Physiol.* **133**, 203-217.
- Zhu, J., X. Hu, E. Dizin and D. Pei** (2003). Catalytic mechanism of S-ribosylhomocysteinase (LuxS): direct observation of ketone intermediates by ¹³C NMR spectroscopy. *J. Am. Chem. Soc.* **125**, 13379-13381.
- Zubieta, C., J. R. Ross, P. Koscheski, Y. Yang, E. Pichersky and J. P. Noel** (2003). Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. *Plant Cell* **15**, 1704-1716.

CURRICULUM VITAE

Baomin Feng

Department of Biology, The Pennsylvania State University, University Park, Pennsylvania
16802, E-mail: bzf103@psu.edu

Education:

- Oct. 2002- Aug.2008: Ph.D., Department of Biology, the Pennsylvania State University, USA
- Sept. 1998-July 2001: Master, Institute of Botany, CAS, China
- Sept. 1994-July 1998: Bachelor, College of Life sciences, Shandong University, China

Research experience:

- Oct.2002-present: Ph.D. study on floral proteome of *Arabidopsis*, transcriptional regulation of anther development by DYT1, and functions of ASK genes, Department of Biology, The Pennsylvania State University
- Mar 2002-Sept. 2002: functional genomics of heavy metal transporter genes in *Arabidopsis*, AGFP project in Europe, Max Planck Institute, Golm, Germany

Teaching experience:

- Fall 2003-2006 Teaching Assistant for Biol.230W, Department of Biology, The Pennsylvania State University
- Summer 2005, Teaching Assistant for the summer course "Frontier and Techniques in Plant Science" at Cold Spring Harbor Laboratory sponsored by NSF

Publications:

- **Baomin Feng**, Yiben Peng, Dihong Lu, Xuan Ma, Yujin Sun, Greg Ning, and Hong Ma. Regulation of reproductive somatic cell identity by DYT1 via multiple mechanisms. *Gene and Development*, in preparation.
- **Baomin Feng**, Lianchao Li, Xiaofan Zhou, Bruce Stanley, and Hong Ma. Analysis of *Arabidopsis* flower proteome reveals extensive post translational modifications. *Journal of Intergrative Plant Biology*, *Submitted*.
- Laura Zahn, **Baomin Feng** and Hong Ma. (2006) Beyond the ABC model: regulation of floral homeotic genes. *Advances in Botanical Research*, Vol.44, pp.161-205. Elsevier
- Xiaohong Zhang, **Baomin Feng**, Qing Zhang, Diya Zhang, Naomi Altman, and Hong Ma. (2005) Genome-wide expression profiling and identification of gene activities during early flower development in *Arabidopsis*. *Plant Molecular Biology*, Vol. 58(3), pp.401-19
- Weimin Ni, Daoxin Xie, Lawrence Hobbie, **Baomin Feng**, Dazhong Zhao, Joseph Akkara, and Hong Ma. (2004) Regulation of flower development in *Arabidopsis* by SCF complexes. *Plant Physiology*, Vol.134, pp.1574-1585
- Dazhong Zhao, Weimin Ni, **Baomin Feng**, Tianfu Han, Megan G. Petrasek, and Hong Ma. (2003) Members of the *Arabidopsis*-SKP1-like gene family exhibit a variety of expression patterns and may play diverse roles in *Arabidopsis*. *Plant Physiology*, Vol. 133, pp.203-217

Award: Braddock Award for spring 2007