MOLECULAR AND FUNCTIONAL STUDIES OF EARLY ANther DEVELOPMENT IN ARABIDOPSIS

A Thesis in
Biology
by
Yujin Sun

© 2007 Yujin Sun

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2007
The thesis of Yujin Sun was reviewed and approved* by the following:

Hong Ma  
Professor of Biology  
Thesis Advisor  
Chair of Committee

Teh-hui Kao  
Professor of Biochemistry and Molecular Biology  
Chair of the Plant Biology Program

Paula McSteen  
Assistant Professor of Biology

Naomi Altman  
Associate Professor of Statistics

Douglas Cavener  
Professor of Biology  
Head of the Department of Biology

*Signatures are on file in the Graduate School
Abstract

The Arabidopsis anther is a model system for studying fundamental cellular processes of cell division, cell differentiation, and cell-cell communication. Although a number of genes have been identified to be important for early anther development and meiosis, information on the relationships among these genes is very limited and many more genes important for anther development and/or meiosis are waiting to be discovered. I used several different approaches to study anther development, as described in this thesis. To identify anther preferential genes and to gain insights into the regulatory networks of several known important genes, I worked with two collaborators to perform microarray experiments using mRNA materials extracted from anthers (stages 4-6) of the wild type, the solo dancers (sds) and male meiocyte death1 (mmd1) mutants, as well as mRNA from wild-type young inflorescences (floral stages 1-9). The global gene expression profiles of those tissues were compared. Further studies included gene functional categorization and gene cluster analysis. The information gained from these analyses will contribute to the development of anther gene regulatory networks and guide future experiments to understand anther development. I also performed an evolutionary study of a family of Kelch repeat-containing F-box proteins (KFBs). F-box proteins are major components of SCF complexes and can bind to SKP1 and the protein substrates at the N-terminal F-box motif and the C-terminal protein-protein interaction domain(s), respectively. The KFB family is one of the largest subfamily of F-box proteins and has only been reported for human and Arabidopsis previously. To study the possible evolutionary history of KFBs, I performed extensive BLAST searches to identify putative KFBs in selected organisms, and analyzed their relationships phylogenetically. Gene duplication and gene expression of the KFBs were also investigated in rice and Arabidopsis. This study indicated that the origin of KFBs occurred before the
divergence of animals and plants, and that plant KFBs underwent rapid gene duplications. My analyses also suggested that the mechanisms for controlling flower timing and the circadian oscillator might be conserved in flowering plants. In addition, collaborating with a colleague in our laboratory, I performed a functional study of the DYSFUNCTIONAL TAPETETUM1 (DYT1) gene. DYT1 encodes a putative transcription factor and was isolated by map-based cloning using an Arabidopsis male sterile mutant. Phenotypic analysis revealed that DYT1 is required for normal tapetum development and function. The mutant tapetal cells formed excess and/or enlarged vacuoles and lacked the densely stained cytoplasm typical of normal tapetal cells. Although most meiotic processes seemed normal, cytokinesis of meiocytes often failed to occur and meiocytes eventually collapsed within anther locules. Double mutant analysis and gene expression studies also suggested that DYT1 acts downstream of two genes, SPOROCYTELESS/NOZZLE (SPL/NZZ) and EXCESS MICROSPOROCYTES1/EXTRA SPOROGENOUS CELL (EMS1/EXS), which are required for early anther development. The studies also indicated that DYT1 is required for the normal expression of several tapetum genes. These results demonstrated that DYT1 is a crucial component in a genetic network that controls early anther development and function.
Table of contents

List of figures...........................................................................................................................................x
List of tables............................................................................................................................................xii
List of abbreviations................................................................................................................................xiii
Acknowledgements..................................................................................................................................xv

Chapter 1 Introduction.................................................................................................................................1
1.1 Arabidopsis is a powerful model system for molecular and genetic studies......................... 2
1.2 Arabidopsis flower development and the ABC model.......................................................... 2
1.3 Anther development....................................................................................................................... 5
1.4 Molecular and genetic studies of several genes essential for early anther development........ 8
   "SPOROCYTELESS and BAM1/BAM2......................................................................................... 8
   Cell-cell signaling controlling tapetum formation........................................................................ 9
   Regulation of tapetum function by several transcription factors during early anther
development...............................................................................................................................11
1.5 Meiosis...........................................................................................................................................13
   Male meiosis in Arabidopsis........................................................................................................ 14
   "SOLO DANCERS (SDS)...............................................................................................................15
   "MALE MEIOCYTE DEATH1 (MMD1)......................................................................................16
   "Arabidopsis SKP1 like 1 (ASK1)...............................................................................................17
1.6 Protein degradation and the SCF complex ......................................................... 18

    SKP1 homologs .................................................................................................... 19

    The F-box protein family .................................................................................. 20

1.7 Microarray technique ....................................................................................... 20

1.8 Thesis overview .............................................................................................. 21

References: ............................................................................................................. 23

Chapter 2. Exploration of Global Gene Expression Profiles in the Arabidopsis Anther .... 38

2.1 Abstract ............................................................................................................ 39

2.2 Introduction ..................................................................................................... 40

2.3 Materials and Methods ................................................................................... 44

    Plant material collection ................................................................................. 44

    Microarray experiments ................................................................................. 45

    Microarray data analysis ............................................................................... 46

    RT-PCR ........................................................................................................... 46

2.4 Results and Discussion .................................................................................. 47

    Microarray data reproducibility and statistical analysis ................................. 47

    Global gene expression profiles in Arabidopsis anthers and young inflorescences ............................ 47

    Identification of A/I (Anther vs Inflorescence) differential genes .................... 50

    Expression patterns of A/I differential genes ................................................ 53
Identification of anther preferential genes and anther specific genes…………………56
Identification of SDS-dependent genes and MMD1-dependent genes…………………60
Expression patterns of SDS-dependent genes…………………………………………60
Assessing the fidelity of microarray data by RT-PCR…………………………………….63
Functional categorization of A/I differential genes……………………………………….66

2.5 Conclusion………………………………………………………………………………….74
Acknowledgements…………………………………………………………………………77
References………………………………………………………………………………….78

Chapter 3 Genome-wide Analysis of Kelch Repeat-containing F-box Protein Family……91
3.1 Abstract……………………………………………………………………………………93
3.2 Introduction………………………………………………………………………………..94
3.3 Material and Methods……………………………………………………………………97
    Sequence retrieval and protein domain analysis……………………………………….97
    Multiple sequence alignment……………………………………………………………98
    Phylogenetic analysis……………………………………………………………………99
    Chromosome distribution and gene duplication types of AtKFBs and OsKFBs………..99
    Gene expression analysis of KFBs……………………………………………………100
3.4 Results……………………………………………………………………………………101
    Plant genomes encode a large number of KFBs with different numbers of Kelch motif101
    Plant KFB family has expanded dramatically via multiple duplication events………102
    G5 KFBs underwent multiple recent gene duplications in Arabidopsis and Brassica…106
Most of G5 AtKFBs are present as tandem repeats in the Arabidopsis genome………109
Gene expression profiles of AtKFBs and OsKFBs……………………………………..113
3.5 Discussion……………………………………………………………………………....…..117
Rapid gene birth evolution of plant KFBs…………………….................................117
Mechanism for controlling flower timing and circadian oscillator may be conserved in
flowering plants………………………………………………………………………...118
Contribution of tandem duplication to Arabidopsis KFBs.................................119
Acknowledgments…………………………………………….………………………..……….121
References………………………………………………………………………………………122

Chapter 4 Regulation of Arabidopsis Tapetum Development and Function by
DYSFUNCTIONAL TAPETUMI (DYTI) Encoding a Putative bHLH Transcription

Factor…………………………………………………………………………………………128
4.1 Abstract……………………………………………………………………………..130
4.2 Introduction………………………………………………………………………..131
4.3 Material and Methods……………………………………………………………..133
  Plant materials and growth……………………………………………………………..133
  Characterization of mutant phenotypes………………………………………………133
  Mapping and functional complementation…………………………………………134
  RT-PCR and real-time PCR……………………………………………………………137
  RNA in situ hybridization……………………………………………………………..138
  Over-expression of DYTI……………………………………………………………..138
  Phylogenetic analysis of the DYTI subfamily……………………………………….139
4.4 Results ......................................................................................................................... 140

The isolation of a new male-sterile mutant ................................................................. 140

The dyt1 mutant is defective in tapetum development ............................................. 142

DYT1 encodes a putative bHLH transcription factor .............................................. 149

DYT1 is preferentially expressed in tapetal cells ..................................................... 153

DYT1 expression is positively regulated by SPL/NZZ and EMS1/EXS ..................... 156

The expression of anther genes in the dyt1 mutant is altered ................................. 156

DYT1 is not sufficient for tapetum development ...................................................... 160

4.5 Discussion ............................................................................................................... 163

DYT1 is important for tapetum development and function ..................................... 163

DYT1 is required for normal levels of the expression of tapetum genes ................ 164

DYT1 supports completion of meiosis ..................................................................... 166

A model for DYT1 function and the control of tapetum identity ............................. 167

Acknowledgments ...................................................................................................... 170

References .................................................................................................................... 171

Chapter 5 Conclusions and Perspectives ................................................................ 177

References .................................................................................................................... 185
List of Figures

Figure 1.1 Early anther development in the wild type (Ler).........................................................7

Figure 2.1 Expression patterns of 707 A/I differential genes.......................................................55

Figure 2.2 Expression patterns of 231 SDS-dependent genes....................................................62

Figure 2.3 Assessing the fidelity of microarray data by RT-PCR..................................................64

Figure 2.4 Functional categorization of 707 A/I differential genes..............................................65

Figure 3.1 Phylogenetic tree of 113 representative Kelch repeats-containing F-box proteins (KFBs).........................................................................................................................105

Figure 3.2 Phylogenetic tree of 102 Kelch repeats-containing F-box proteins (KFBs) in the plant G5 subfamily..................................................................................................................108

Figure 3.3 Chromosome distribution of 66 G5-Kelch repeats-containing F-box proteins (KFBs) in Arabidopsis.......................................................................................................................110

Figure 3.4 Chromosome distribution of 31 Arabidopsis non-G5 Kelch repeats-containing F-box proteins (KFBs)....................................................................................................................111

Figure 3.5 Chromosome distribution of 28 Kelch repeats-containing F-box proteins (KFBs) in rice...........................................................................................................................................112

Figure 3.6 Expression patterns of Kelch repeats-containing F-box proteins (KFBs).in Arabidopsis.............................................................................................................................................116

Figure 4.1 Phenotypes of the wild type (Ler), dyt1 mutant and transgenic plants for complementation........................................................................................................................................141

Figure 4.2 Meiosis in Ler and dyt1 anthers.......................................................................................143

Figure 4.3 Anther development from stage 4 to stage 8 in the wild type (Ler) and dyt1 mutant.................................................................................................................................145
Figure 4.4 The *DYT1* gene structure and annotated conserved domain............................152

Figure 4.5 The *DYT1* expression pattern.................................................................155

Figure 4.6 The expression of anther and tapetum genes in the wild type and *dyt1* mutant......159

Figure 4.7 The *ems1 dyt1* double mutant anther morphology at late stage 5.....................161

Figure 4.8 Phenotypes of wild-type and mutant stage 5 anthers.................................162

Figure 4.9 A model for *DYT1* function and tapetum specification...............................169
List of Tables

Table 2.1 Pearson correlation coefficient (r-value) of two replicates for each genotype ........49
Table 2.2 Expression patterns of known genes in the list of A/I differential genes ...............52
Table 2.3 Expression patterns of 106 anther specific genes ...........................................57
Table 4.1 PCR primers used in this study ........................................................................135
Table 4.2 Microarray data of the genes selected in RT-PCR ............................................147
Table 4.3 Primers used in real-time PCR ........................................................................148
List of abbreviations

AMS=ABORTED MICROSPORES
ASK1=Arabidopsis SKP1 like 1
CDK=cyclin dependent kinases
ChIP=chromatin immunoprecipitation
DAPI=4'-6-Diamidino-2-phenylindole
DEX=dexamethasone
DYT1=DYSFUNCTIONAL TAPETETUM1
E1=ubiquitin-activating enzyme
E2=ubiquitin-conjugating enzyme
E3=ubiquitin ligase
EMS1/EXS=EXCESS MICROSPOROCYTES1/EXTRA SPOROGENOUS CELL
EST=expressed sequence tag
FDR=false discovery rate
KFB=Kelch repeat-containing F-box protein
Ler=Landsberg erecta
ML=maximum likelihood
MMD1=MALE MEIOCYTE DEATH1
MS1=MALE STERILITY1
NJ=neighbor-joining
PMC=pollen mother cell
RNAi= RNA-interference
$r$-value=Pearson correlation coefficient
SAM=significance analysis of microarrays
SC=synaptonemal complex
SCF=SKP1, Cullin1, F-box protein, and Rbx1
SDS=SOLO DANCERS
SKP1=S-phase kinase-associated protein 1
SPL/NZZ=SPOROCYTELESS/NOZZLE
TPD1=TAPETUM DETERMINANT1
Acknowledgements

I would like to express my deepest gratitude to my advisor Dr. Hong Ma for his great support and guidance in the past five years. He not only taught me a lot of academic knowledge, but also offered great help in life as well. He was always there when I needed help and without that I could not have come this far. I also want to thank my other committee members, Dr. Teh-hui Kao, Dr. Paula McSteen, and Dr. Naomi Altman. They have given me valuable suggestions and comments to improve my research. They also provided great support for my postdoctoral applications in the last two months.

I would like to thank all of the present and past members of the Ma lab for their friendships and various help all the way. Dr. Dazhong Zhao and Dr. Wei Zhang offered me lots of academic instruction. Wang Jiong and Hu Yi have worked hard to make the lab a good place to work. I enjoyed both the academic atmosphere and the family atmosphere in this lab. I will miss them all and definitely will miss those birthday parties. I also thank Ruth Haldeman and Missy Hazen in the electron microscopy (EM) facility. They have patiently helped me when I did EM-related work there.

Finally, I want to thank my wife and my parents for their endless support. Without them, I cannot have achieved this much. I also thank my two sons, Kevin and Andy. They give me pressure as well as encouragement to go on. And I will go on.
Chapter 1

Introduction
1.1 Arabidopsis is a powerful model system for molecular and genetic studies

*Arabidopsis thaliana* is a member of the *Brassicaceae* or mustard family of flowering plants (O'Kane, 1997). It has a relatively small size and can easily grow in the greenhouse or small indoor growth chambers. Arabidopsis also has a short life cycle and high fecundity. A single plant produces thousands of seeds in less than three months. The relatively small and completely sequenced genome, easy transformability and large collections of genetic markers for different ecotypes, make both forward and reverse genetic studies convenient in Arabidopsis (Meinke et al., 1998; Arabidopsis Genome Initiative, 2000). The rapidly accumulating DNA, RNA, protein databases and technology, and large mutant collections provide additional resources and have promoted rapid research progress in Arabidopsis. All these features make Arabidopsis a convenient and powerful model system for molecular and genetic studies in plants.

1.2 Arabidopsis flower development and the ABC model

The Arabidopsis flower consists of four rings or whorls of floral organs with the reproductive organs surrounded by external sterile organs. From outside to inside, there are four leaf-like sepals, four white petals, six stamens (male reproductive organs, with four long stamens and two short stamens), and a gynoecium with two fused carpels (female reproductive organs), respectively. The process of flower development follows floral induction, the transition from vegetative to reproductive growth, and begins with the floral meristem (Kanno et al., 2007).

Arabidopsis flower development has been divided into 12 stages from floral meristem initiation to anthesis, by using landmark events characterized by the shape, size, and surface features of developing floral organs (Smyth et al., 1990). The floral meristem is initiated at stage 1 on the flank of the apical meristem, and develops into a nearly spherical dome at stage 2.
Stages 3 through 8 involve the initiation of the floral organ primordia and organogenesis. Specifically, the sepal primordia arise at stage 3 and overlie the flower meristem at stage 4. The petal and stamen primordia form at the beginning of stage 5. At stage 6, the floral bud is completely covered by sepals, the stamen primordia form a domed shape, while the petal primordia are still very small, and the gynoecium is formed on the center dome of the floral meristem. At stage 7, the long stamen primordia are stalked at the base, and the petal primordia form a domed shape, while the gynoecium forms a tube. At stage 8, the stamens increase dramatically in size comparing to the petals, and locules are visible in the long stamens. From stages 9 to 12, floral organs expand dramatically in size and cell differentiation occurs. Remarkably, at stage 9, the petal primordia become stalked at the base and the growth of the petal accelerates. Petals reach the height of the short stamens at stage 10. Soon afterwards, stigmatic papillae form on the top surface of the gynoecium at stage 11. Stage 12 begins when petals reach the height of the long stamens and is the last stage before flower anthesis. After stage 12, flowers open to allow for pollination, followed by fruit development (Smyth et al., 1990; Robles and Pelaz, 2005).

Genetic studies of several homeotic mutants in both Arabidopsis and Antirrhinum have led to the proposal of the ABC model in controlling flower development (Coen and Meyerowitz, 1991; Weigel and Meyerowitz, 1994). Particularly, in Arabidopsis, APETALA1 (AP1) and APETALA2 (AP2) are A function genes; APETALA3 (AP3) and PISTILLATA (PI) are B function genes; and AGAMOUS (AG) is a C function gene (Irish and Sussex, 1990; Yanofsky et al., 1990; Jack et al., 1992; Mandel et al., 1992; Goto and Meyerowitz, 1994; Jack et al., 1994; Jofuku et al., 1994). The activity of A function is limited to the first two whorls, while the activities of B and C functions are restricted to the middle two whorls and the inner two whorls, respectively (Irish

All the ABC function genes, except AP2, belong to the family of MADS-box genes encoding transcription factors with a conserved domain called MADS-domain (Coen and Meyerowitz, 1991; Riechmann and Meyerowitz, 1997; Ma and dePamphilis, 2000; Ng and Yanofsky, 2001; Parenicova et al., 2003). Recently, another four MADS-box genes, SEPALLATA1, SEPALLATA2, SEPALLATA3 and SEPALLATA4 (SEP1, SEP2, SEP3, and SEP4), were found to be redundantly involved in specifying the identities of all the floral organs, and were named as E function genes (Pelaz et al., 2000; Goto et al., 2001; Honma and Goto, 2001; Ng and Yanofsky, 2001; Parenicova et al., 2003; Ditta et al., 2004). The loss of E function causes floral organs to be converted into leaf-like organs (Ditta et al., 2004). Accordingly, A and E functions specify sepal identity; A, B and E functions specify petal identity; B, C and E functions specify stamen identity; C and E functions specify carpel identity (Jack, 2001; Ma, 2005; Kanno et al., 2007).
1.3 Anther development

In flowering plants, the stamen is composed of the vascular filament and the pollen-yielding anther. The filament provides physical support for the anther and transports the nutrients necessary for anther development. The anther provides the space for the male reproductive process and contains both highly specialized reproductive cells and non-reproductive cells that are required for male gametophyte formation, differentiation, and development. Therefore, anther development is one of the most important processes during plant development (McCormick, 1991; Goldberg et al., 1993; Ma, 2005).

In Arabidopsis, anther development can be divided into two phases (Goldberg et al., 1993; Sanders et al., 1999; Ma, 2005). The first phase involves cell and tissue differentiation, and male meiosis. The second phase involves anther growth, pollen grain differentiation, tissue degeneration, anther dehiscence and stamen senescence. On the basis of morphological landmarks and cellular features of cross sections of anthers observed under the light microscope, Arabidopsis anther development can be divided into 14 stages (Sanders et al., 1999). Stages 1 through 8 belong to the first phase, and stages 9 through 14 belong to the second phase. At the beginning of the first phase, the anther primordium has three cell layers at stage 1, called L1, L2 and L3, from outside to inside. L1 layer cells give rise to the epidermis. L3 layer cells form connective and vascular bundle tissues. L2 layer cells give rise to archesporial cells at stage 2. At stage 3, the archesporial cells divide into primary parietal cells and primary sporogenous cells, which further give rise to two secondary parietal cell layers and sporogenous cells, respectively. At stage 4 (Figure 1.1A), the outer secondary parietal cells divide and differentiate into the endothecium layer and the middle layer, while the inner parietal cells differentiate to form the tapetum layer. At this point, a bilaterally symmetric four-lobed structure is formed. The
sporogenous cells divide and differentiate to form microsporocytes at the center of each lobe during stage 5 (Figure 1.1B). The cross section of the stage 5 anther shows that the anther has a butterfly shape with two outer larger lobes and two inner smaller lobes. Each lobe is a reproductive unit, and is composed of five cell layers, called epidermis, endothecium, middle layer, tapetum, and microsporocytes, from outside to inside. The microsporocytes are then detached from the tapetum and from each other at stage 6 (Figure 1.1C), and complete meiosis at stage 7 (Figure 1.1D), resulting in the formation of tetrads in anther lobes. The thick callose wall of tetrads degenerates and each tetrad releases four haploid microspores at stage 8. At stage 9, the anther continues to grow and expand, and the microspore forms an exine wall and becomes vacuolated. The tapetum degenerates during stages 10 and 11. Microspores undergo two rounds of mitotic divisions during stages 11 and 12, and develop into tricellular pollen grains at stage 12, followed by anther dehiscence at stage 13 and stamen senescence at stage 14.

Although the corresponding relationships between the stages of anther development and the stages of floral development are not well defined, some of them could be estimated. The anther stages 1 through 4 correspond approximately to the floral stages 5 through 8. The anther stages 5 through 8 are within the long floral stage 9. It is difficult to match the anther stages 9 through 12 to the floral stages 10 through 12, but anther stage 13 is at the time of floral stage 13, at which the pollen is released and anthesis occurs (Smyth et al., 1990; Sanders et al., 1999; Ma, 2005).
Figure 1.1. Early anther development in the wild type (Ler). (A), (B), (C), (D) are from sections of wild type anthers. (A) Stage 4. (B) Stage 5. (C) Stage 6. (D) Stage 7. E, epidermis; OSP, outer secondary parietal cells; ISP, inner secondary parietal cells; SS, secondary sporogenous cells; En, endothecium; ML, middle layer; T, tapetum; Ms, microsporocytes; Tds, tetrads. Scale bars: 10µm.
1.4 Molecular and genetic studies of several genes essential for early anther development

**SPOROCYTELESS and BAM1/BAM2**

The process of early anther development involves both cell division and cell differentiation. Recent studies in Arabidopsis have identified several genes that are essential for this process. One of them is *SPOROCYTELESS (SPL)* also called *NOZZLE (NZZ)*, which is required for differentiation of tissues derived from archesporial cells in Arabidopsis (Schiefthaler et al., 1999; Yang et al., 1999b). The *spl/nzz* plant is male and female sterile. In the *spl/nzz* anther, archesporial cells fail to divide into primary parietal cells and primary sporogenous cells resulting in the absence of the endothecium, middle layer, tapetum, and microsporocyte (also called pollen mother cell) (Schiefthaler et al., 1999; Yang et al., 1999b). *SPL/NZZ* encodes a novel transcription factor, and might activate some downstream target genes that are involved in early anther development. The finding that the *SPL/NZZ* gene is a direct target of *AG*, which encodes a MADS-box protein, indicates that *SPL/NZZ* might be a key regulator that links the organ identity gene *AG* and some downstream genes essential for early anther development (Ito et al., 2004).

Two other genes, *BAM1* and *BAM2*, are also essential for cell division and cell differentiation during early anther development in Arabidopsis. *BAM1* and *BAM2* are two recently duplicated homologs and encode leucine-rich repeat receptor-like protein kinases closely related to CLAVATA1 (CLV1), which controls shoot apical meristem development (Clark et al., 1997; DeYoung et al., 2006). The *bam1* and *bam2* single mutants have no obvious phenotype, but the *bam1 bam2* double mutant has smaller shoot apical meristem than normal and is male and female sterile (DeYoung et al., 2006). In the *bam1 bam2* anther, the daughter cells of archesporial cells fail to form the endothecium, middle layer, tapetum and pollen mother cell;
instead, they form pollen mother cell-like cells only (Hord et al., 2006). The idea that these cells are pollen mother cell-like cells is further supported by the fact that they expressed two meiotic genes, SDS and RAD51 (Azumi et al., 2002; Li et al., 2004). In addition, these cells also showed meiotic chromosome condensation (Hord et al., 2005).

Further expression studies of SPL in the bam1 bam2 mutant and of BAM1 in the spl mutant found that the SPL expression was expanded spatially in the bam1 bam2 mutant, whereas the BAM1 expression was reduced in the spl mutant, suggesting that SPL positively regulates BAM1 expression and that BAM1/BAM2 may repress SPL expression (Hord et al., 2006). The results of gene expression studies and mutant phenotypic analyses indicate SPL and BAM1/BAM2 might form a positive-negative feedback loop to control the sporogenous and somatic cell fates (Hord et al., 2006).

**Cell-cell signaling controlling tapetum formation**

The findings of several receptor-like kinases controlling tapetum formation suggest that cell-cell signaling might be critical for cell division and cell differentiation during anther development. Among them, *EXCESS MICROSPOROCYTES1/EXTRA SPOROGENOUS CELL* (*EMS1/EXS*) is the first gene reported to be essential for normal tapetum formation. The *ems1/exs* plant is male and female sterile. The mutant anther lacks tapetal cells and has extra microsporocytes (Zhao et al., 2002; Canale et al., 2002). *EMS1/EXS* encodes a putative leucine-rich repeat receptor-like protein kinase, which is localized to the cell plasma membrane. And the recombinant EMS1 protein kinase domain is able to auto-phosphorylate. *EMS1/EXS* is evenly expressed in the precursors of the tapetum and microsporocyte, and the expression of *EMS1/EXS*
becomes stronger in the tapetum and is reduced dramatically in the microsporocyte when the
tapetum formed.

Another gene called \textit{TAPETUM DETERMINANT1 (TPD1)} is also found to be essential
for normal tapetum formation (Yang et al., 2003a; Yang et al., 2005). Similar to the \textit{ems1/exs}
anther, the \textit{tpd1} anther also lacks the tapetum layer and has extra microsporocytes. \textit{TPD1} has
similar expression patterns as \textit{EMS1/EXS} in the precursors of tapetal cells and microsporocytes,
but its expression becomes stronger in microsporocytes instead of tapetal cells. \textit{TPD1} encodes a
putatively secreted protein and is believed to be the ligand of \textit{EMS1/EXS} (Yang et al., 2003a;
Yang et al., 2005). Further studies found that the phenotype of the \textit{tpd1 ems1} double mutant is
similar to that of the single mutants and the function of ectopic \textit{TPD1} requires a functional \textit{EMS1}
gene, indicating that \textit{TPD1} may act through \textit{EMS1} to control tapetum formation. These results
supports the idea that the signal of tapetum formation is from microsporocytes through the
interaction of \textit{TPD1} and \textit{EMS1/EXS} in \textit{Arabidopsis} (Yang et al., 2003a; Yang et al., 2005).

Two other recently duplicated protein kinases, \textit{SERK1} and \textit{SERK2} function redundantly
in normal tapetum formation in \textit{Arabidopsis} (Hecht et al., 2001; Albrecht et al., 2005; Colcombet
et al., 2005). The single mutants have no detectable phenotype, and the double mutant is male
and female sterile. Similar to the \textit{ems1/exs} and \textit{tpd1} anther, the \textit{serk1 serk2} anther fails to form
tapetal cells and has extra microsporocytes. Previous studies also found that \textit{BAK1/SERK3}, a
close homolog of \textit{SERK1} and \textit{SERK2} in \textit{Arabidopsis}, can physically interact with \textit{BRI1}, a
receptor kinase involved in brassinosteroid signal transduction (He et al., 2000; Wang et al.,
2001; Li et al., 2002; Nam and Li, 2002). The fact that \textit{EMS1/EXS} is a close homolog of \textit{BRI1}
suggests that \textit{EMS1/EXS} and \textit{SERK1/SERK2} might also interact as co-receptors to control
tapetum formation through the signaling pathway (Shiu and Bleecker, 2001; Hord et al., 2006).
Regulation of tapetum function by several transcription factors during early anther development

Among the several somatic cell types of the anther, the tapetum plays the most important role during anther development. Many proteins, lipids, polysaccharides and other molecules necessary for pollen development are generated in the tapetum. Therefore, the tapetum is characterized by active protein synthesis and secretion, a high rate of energy metabolism, and expression of many tapetum-preferential genes (Dickinson and Bell, 1976; Liu and Dickinson, 1989; Hernould et al., 1998; Rubinelli et al., 1998; Taylor et al., 1998; Zheng et al., 2003; Scott et al., 2004).

Although a large number of genes are expressed in the tapetum, only a few are found to be important for normal tapetum function during early anther development. Reverse genetic studies have uncovered that MYB33 and MYB65 function redundantly in controlling normal tapetum function after tapetum formation (Millar and Gubler, 2005). MYB33 and MYB65 are recently duplicated genes belonging to the MYB family, the largest transcription factor family (Stracke et al., 2001; Millar and Gubler, 2005). The loss of functions of MYB33 and MYB65 causes the tapetum to expand starting from anther stage 6. The mutant tapetal cells do not undergo programmed cell death during anther stages 10 and 11. The abnormal tapetum causes the degeneration of microsporocytes after meiosis, and eventually the meiotic cells degrade in the anther locule. Unlike some other early tapetum defect mutants (fat tapetum, (GUS-negative) gne1 and gne4), the phenotype of myb33 myb65 could be restored when mutant plants grow under the higher light intensities or at the lower light intensities (Chaudhury et al., 1994; Sanders et al., 1999; Sorensen et al., 2002; Millar and Gubler, 2005). Although fat tapetum, gne1 and gne4 have been reported with preliminary phenotypic analyses, all three mutants are not cloned.
yet. The mutant phenotypes are different from that of myb33 myb65, the tapetal cells appear abnormal with enlarged vacuoles at stages 5 to 6, and degenerate afterwards. These results indicate that MYB33 and MYB65 may be within a transcriptional network essential for normal tapetum function during the meiosis stage and the genes in this network may be different from the FAT TAPETUM, GNE1 and GNE4, which were defined only by mutations (Chaudhury et al., 1994; Sanders et al., 1999; Sorensen et al., 2002; Millar and Gubler, 2005).

Previous studies also reported two other genes, ABORTED MICROSPORES (AMS) and MALE STERILITY1 (MS1), which are essential for male fertility (Wilson et al., 2001; Ito and Shinozaki, 2002; Sorensen et al., 2003). AMS encodes a bHLH transcription factor, belonging to the MYC class bHLH proteins, which can form heterodimers with their protein partners to regulate the transcription of target genes (Ledent et al., 2002; Sorensen et al., 2003; Toledo-Ortiz et al., 2003). The ams anther exhibits normal callose deposition and can complete meiotic cytokinesis to form normal tetrads at stage 7 (Sorensen et al., 2003). The defects of mutant anther can be observed as early as stage 8, at which stage tapetal cells become prematurely flattened and separated from each other and initiate the degeneration process, and microspores degrade right after their release from tetrads (Sorensen et al., 2003). The continually expanded and vacuolated tapetal cells occupy the space of the anther locule, in which microspores degraded, resulting in the complete absence of pollen grains in mutant plants (Sorensen et al., 2003). MS1 encodes a PHD domain protein, which can potentially regulate transcription (Wilson et al., 2001; Ito and Shinozaki, 2002). The development of ms1 anther is normal in callose production, the meiosis process, the formation of tetrads, the release of microspores from callose wall, and the initiation of exine synthesis (Wilson et al., 2001; Ito and Shinozaki, 2002). The mutant anther defects occur immediately after the release of microspores from the callose wall.
The mutant tapetal cells become enlarged and vacuolated with the degeneration of cytoplasm, followed by abnormal pollen development (Wilson et al., 2001; Ito and Shinozaki, 2002). Both \textit{AMS} and \textit{MS1} are expressed specifically in the tapetum and can potentially regulate transcription. Therefore, \textit{AMS} and \textit{MS1} may be part of a transcriptional network essential for post-meiotic tapetal functions that support pollen development (Wilson et al., 2001; Ito and Shinozaki, 2002; Sorensen et al., 2003).

\textbf{1.5 Meiosis}

Meiosis is a highly conserved process essential for eukaryotic sexual reproduction. During meiosis, a diploid parental cell undergoes a single round of DNA replication followed by two rounds of cell division to produce four haploid gametophytes. Afterwards, the haploid male and female gametes are fused to form a diploid zygote, which further develops into the mature individual to complete the life cycle. The diploid cell contains two versions of each chromosome, one from the father and the other from the mother, which are called homologs. The two rounds of cell divisions are subsequently named as meiosis I and meiosis II. Meiosis I is a unique process, involving homologous chromosome pairing and separation to ensure the gametophyte containing only half of the chromosomes of parental cells. Meiosis II is similar to mitosis, involving the segregation of replicated sister chromatins.

Studies in yeast, animal and plant have uncovered many critical events of meiosis, including meiosis initiation, DNA replication, chromosome condensation, homologous chromosome pairing, synapsis and synaptonemal complex (SC) formation, recombination, homolog separation, chromosome cohesion and segregation, tetrad formation and the release of gametes from tetrad (Roeder, 1997; Zickler and Kleckner, 1999; Armstrong and Jones, 2003; Ma,
Although extensive studies of meiosis have been carried out in budding yeast, the information obtained from this unicellular micro-organism may not be applicable to higher eukaryotes such as animals and plants (Loidl, 2000; Jones et al., 2003). Many animal meiotic mutants are lethal, making it difficult to fully understand the mechanisms of their function in the meiotic process (Venkitaraman, 2002; Wilson and Yang, 2004). The non-lethality of most meiotic mutants in plants provides a good opportunity to study the function of meiotic genes (Gudmundsdottir and Ashworth, 2004; Siaud et al., 2004).

**Male meiosis in Arabidopsis**

In Arabidopsis, male meiosis occurs in the anther of the stamen. The microsporocyte is formed at anther stage 5 and undergoes meiosis during anther stages 5 and 6. It completes cytokinesis at stage 7 and releases four haploid microspores at stage 8 followed by pollen development to produce mature pollen grains for fertilization (Smyth et al., 1990; Goldberg et al., 1993; Armstrong and Jones, 2003). Since a single anther in Arabidopsis contains approximately 50 meiotic cells, the anther has been a good system used by researchers to study meiosis in plants by both cytological and genetic approaches.

Arabidopsis male meiosis stages have been described in details based on the major landmarks of chromosomal behavior under the light microscope (Smyth et al., 1990; Goldberg et al., 1993; Armstrong and Jones, 2003; Jones et al., 2003). Following DNA replication in interphase, male meiotic cells enter into prophase I, which can be further divided into five substages: leptotene, zygotene, pachytene, diplotene, and diakinesis. The homologous chromosomes begin to condense as thin lines in leptotene, and initiate pairing with each other from late leptotene to early zygotene. The initiation and progression of synapsis and chromosome recombination occur during zygotene. At this stage, the paired homologs can be
first clearly observed under the light microscope. The fully synapsed homologs can be observed as thick thread-like structures at the stage of pachytene, in which the typical tripartite SC forms and is visible under the transmission electron microscope. At the diplotene stage, synapsed homologs begin to desynapse and the SCs also disappear at the end of this stage. Homologs separate except at specific regions called chiasmata, the sites of the cross-over of homologs. Homologs are highly condensed at the stage of diakinesis (Ross et al., 1996).

After the diakinesis stage, meiotic cells enter into metaphase I, at which stage the bivalents become attached to the meiotic spindle at the centomeres and are aligned at the cell equator. Homologs are separated and moved to opposite poles of the spindle at anaphase I. The cytoplasmic organelles are expelled from the spindle to form a distinct organelle band separating the two daughter nuclei during metaphase I and anaphase I. At the end of meiosis I, five chromosomes of each daughter nuclei form a cluster at each pole of the meiotic cell and partially decondense at telophase I. Afterwards, each daughter nucleus enters into meiosis II.

Meiosis II is similar to mitosis in which sister chromatids are segregated. At prophase II, chromosomes in each daughter nucleus recondense. The recondensed chromosomes are then aligned at metaphase II. Sister chromatids are separated and move to the opposite poles. Finally at telophase II, four clusters of chromosomes in each meiotic cell are formed and chromosomes undergo decondensation. Following telophase II, cytokinesis occurs and produces four haploid microspores (Brown and Lemmon, 2001). A large number of meiotic genes have been identified and studied in Arabidopsis, studies on several meiotic genes will be reviewed briefly below.

**SOLO DANCERS (SDS)**

Cyclins are major regulators of the activity of cyclin-dependent kinases, which play critical roles in controlling eukaryotic cell cycle progression through the well-controlled
phosphorylation of specific substrates on serine/threonine residues (Morgan, 1997). Although a large number of cyclins are encoded in Arabidopsis, SDS is one of the very few cyclins whose mutation causes meiotic defects (Azumi et al., 2002; Wang et al., 2004b). The sds mutant was initially identified by screening for sterile mutants caused by Ds-transposon insertion in Arabidopsis thaliana ecotype Landsberg erecta (Ler) background in our lab. The mutant plant appears normal in vegetative growth and the mutant flower has normal floral organs except having short filaments of stamens and anthers with few or almost no pollen grains at the time of anthesis. Further analysis indicated that SDS plays essential roles in homolog paring, synapsis, recombination and bivalent formation (Azumi et al., 2002). Phylogenetic studies have shown that plant cyclins underwent faster gene duplication events than animal cyclins (Wang et al., 2004a). In Arabidopsis, 49 cyclins have been identified based on the conserved cyclin domain, which has approximately 250 amino acid residues. The expression study of Arabidopsis cyclins suggests possible gene redundancy among some closely related members. It was also shown that SDS is a newly evolved cyclin and its orthologs are only detected in some flowering plants (Wang et al., 2004a).

**MALE MEIOCYTE DEATH1 (MMD1)**

MMD1 encodes a novel PHD-containing nuclear protein. The PHD domain can interact with protein, DNA, and RNA, and is often found in proteins that are involved in chromatin remodeling (Aasland et al., 1995). The mmd1 mutant was initially isolated by screening for male sterile mutants generated by the Ds-transposon insertion in the Ler background in our lab (Yang et al., 2003b). Phenotypic analysis suggested that the mutant plant has normal vegetative growth, but the mutant anther does not produce any pollen grains. Further analysis indicated that MMD1 is required for normal male meiosis. Male meiosis in mutant plants appears normal up to
diakinesis. After that, meiotic cells undergo programmed cell death with chromosome fragmentation and cytoplasmic shrinkage. Eventually, all meiotic cells die and fail to undergo cytokinesis. *MMD1* is preferentially expressed in male meiocytes at the time of meiosis. The presence of the PHD domain and two putative nuclear localization domains, along with the mutant phenotype, suggests that *MMD1* plays important roles in transcriptional regulation and chromatin remodelling during male meiosis (Yang et al., 2003b).

**ARABIDOPSIS SKP1 LIKE 1 (ASK1)**

ASK1 is the first homolog of the yeast SKP1 (S-phase kinase-associated protein 1) identified in Arabidopsis. ASK1 is a major component of the SCF complex, an E3 ubiquitin ligase, involved in ubiquitin dependent protein degradation pathway (Yang et al., 1999a). The *ask1-1* mutant was first identified in our lab by screening for sterile mutants generated by *Ds*-transposon insertion in the *Ler* background. Unlike many other meiotic mutants, *ask1-1* was abnormal in both vegetative growth and reproductive growth. The mutant plants are shorter and smaller than normal plants. Mutant flowers also have mild defects in sepalas and petals with reduced number and size. Mutant anthers produce abnormal tetrads and microspores, and hardly produce any pollen grains. Further studies indicated that both male meiosis and female meiosis are abnormal in *ask1-1*. *ASK1* plays a variety of roles in several aspects of plant development and is essential for chromosome separation and segregation and for chromosome remodeling (Yang et al., 1999a; Zhao et al., 1999; Zhao et al., 2001; Zhao et al., 2003a; Wang and Yang, 2006; Yang et al., 2006; Zhao et al., 2006).
1.6 Protein degradation and the SCF complex

The vast majority of protein degradation is mediated through the ubiquitin-proteasome pathway, which plays essential roles in cell cycle progression, transcriptional regulation and signal transduction (Hershko and Ciechanover, 1998). The ubiquitylation cascade involves three ubiquitin enzymes (Koepp et al., 2001; Pickart, 2001). E1, the ubiquitin-activating enzyme activates ubiquitin in the presence of ATP, and activated ubiquitin is then transferred from E1 to the ubiquitin-conjugating enzyme (E2). The third step involves the ubiquitin-protein ligase (E3), which can interact with the protein substrate and E2, and bring the ubiquitin to the substrate. Several ubiquitins are linked to form a polyubiquitin chain attached with the protein substrate. The polyubiquitin chain tagged protein substrate can be rapidly degraded by the 26S proteasome into small peptides. Comparative genome analysis has found that both E1 and E2 are less specific and are well conserved. Generally, only a few genes encode E1s and tens of genes encode E2s. In contrast, hundreds of genes encode E3s (Bachmair et al., 2001). The specific recognition of protein targets is achieved through E3s.

During the last 20 years, a large number of cullin based E3 enzymes have been uncovered in yeast, human, and plant (Lyapina et al., 1998; Kobayashi et al., 2004; Willems et al., 2004; Hong et al., 2005). One of the largest and the best characterized class of E3s is called the SCF complex. The SCF complex is composed of SKP1 (S-phase kinase associated protein), Cullin1/Cdc53, Rbx1, and an F-box protein (Hershko and Ciechanover, 1998; Hershko, 2005). Studies of protein-protein interaction and protein crystal structures indicate that cullin1/Cdc53 interacts with SKP1 and Rbx1 through the long N-terminal stalk domain and the C-terminal globular domain, respectively (Krek, 1998; Skowyra et al., 1999; Zheng et al., 2002). Rbx1 contains a ring finger domain that interacts with E2, while SKP1 bridges Cullin1 and an F-box
protein (Kamura et al., 1999). The F-box protein has a relatively conserved F-box domain near the N-terminus interacting with SKP1 and a less conserved protein–protein interaction domain(s) at the C-terminus specifying the ubiquitylational target(s) (del Pozo and Estelle, 2000).

**SKP1 homologs**

*SKP1* is well conserved in eukaryotic organisms. In budding yeast and human, the single functional SKP1 protein forms multiple SCF complexes, playing an essential role in cell cycle progression, transcriptional regulation, and signal transduction (Kornitzer et al., 1994; Piatti et al., 1996; Henchoz et al., 1997; Skowyra et al., 1999; Peters, 2003). Multiple *SKP1* homologs have been identified in *Arabidopsis thaliana, Drosophila melanogaster, and Caenorhabditis elegans*. Specifically, 21 have been identified in *Arabidopsis thaliana* (called *ASK*), 7 in *D. Melanogaster*, and 21 in *C. elegans* (called *SKR*) (Farras et al., 2001; Nayak et al., 2002; Yamanaka et al., 2002; Risseeuw et al., 2003; Kong et al., 2004). Phylogenetic studies indicated that *SKP1* genes underwent rapid gene duplication in Arabidopsis (Kong et al., 2004).

Among the 21 Arabidopsis *ASK* genes, 19 *ASK* genes have been detected to be expressed with distinct patterns (Zhao et al., 2003b). *ASK1* and *ASK2* are the most closely related homologs. They have similar gene expression patterns, but *ASK1* is expressed at a much higher level than *ASK2*, indicating that they may share some redundant functions and *ASK1* might have a more dominant role (Zhao et al., 2003a). In contrast to the *ask1-1* mutant, the *ask2-1* mutant plant is normal without any detectable phenotypic defect. However, the *ask1 ask2* double mutant is defective in embryo development and is seedling lethal, supporting the idea that *ASK1* and *ASK2* are partially redundant (Liu et al., 2004). Slight defects in early flower development are detectable in *ASK1* RNAi plants and functions of all other *ASK* genes are still unclear (Ni et al., 2004).
The F-box protein family

The F-box motif was initially identified in the human cyclin F protein that plays important roles in cell cycle progression. The F-box motif contains approximately 60 amino acid residues, among which the first 40 residues are believed to be the core of the SKP1-binding site and the remaining 20 residues provide additional contacts to help the SKP1-binding. Unlike SKP1, Cullin1, and Rbx1 that are well conserved in organisms and have only a small number of homologs, F-box proteins are very diversified and have rapidly evolved (Gagne et al., 2002; Risseeuw et al., 2003; Kong et al., 2004). Approximately 700 F-box proteins have been detected in Arabidopsis, and previous studies have found that different F-box proteins may interact with different SKP1 homlogs in Arabidopsis, suggesting the existence of a large set of SCF complexes in Arabidopsis (Nayak et al., 2002; Yamanaka et al., 2002; Risseeuw et al., 2003). F-box proteins can be divided into many different subfamilies based on their C-terminal protein-protein interaction domains, including WD40 repeat, Leucine-rich repeat, Tub, Lectin, Kelch repeat and some other motifs (Gagne et al., 2002; Risseeuw et al., 2003). In contrast to Arabidopsis, there are only 68 F-box proteins encoded in the human genome. How and why Arabidopsis or plants have gained such a large number of F-box proteins is still unknown (Gagne et al., 2002; Kuroda et al., 2002; Jin et al., 2004).

1.7 The microarray technique

The microarray technique has been shown to be a powerful tool for studying global gene expression profiles, since the expression patterns of a large number of genes can be obtained at the same time. This technique has been well employed in various research fields, including the analysis of global gene expression, identification of organ specific genes, and studies of some
functional pathways (Birnbaum et al., 2003; Brinker et al., 2004; Ohgishi et al., 2004; Wellmer et al., 2004; Zhang et al., 2005). The GeneChip® Arabidopsis ATH1 Genome Array was designed by Affymetrix and The Institute for Genome Research (TIGR) (Affymetrix, Santa Clara, CA). On the ATH1 chip, each probe is designed to have 11 pairs of oligonucleotides corresponding to a specific gene. The ATH1 genome array contains 22,810 probe sets representing approximately 80% of Arabidopsis nuclear genes, and is one of the most powerful tools to analyze the global gene expression patterns.

1.8 Thesis overview

Previous studies have reported that more than ten thousand genes are expressed in young inflorescences, stage 12 flowers, and some other tissues (Zhang et al., 2005). It is likely that a large number of genes are expressed in the anther. As in inflorescences, hundreds of them might be differentially expressed in the anther (Zhang et al., 2005). To address these hypotheses and uncover new genes important for early anther development, in this thesis, I studied activities of anther genes by using several different approaches. In Chapter 2, I used the microarray technology to identify anther preferential genes and genes regulated by SDS and/or MMD1. Further analyses of these genes have enabled us to gain insights into early anther development and male meiosis. In Chapter 3, I chose the bioinformatic approach to study the molecular evolution of Kelch repeat-containing F-box protein (KFB) family. KFBs can interact with the SKP1 homologs, which are essential for male meiosis. The results of my study suggest that plant KFBs underwent rapid gene duplications and the origin of KFBs occurred before the divergence of animals and plants. The several newly evolved anther-specific KFBs may play important roles in early anther development or male meiosis. In Chapter 4, I performed functional studies of
DYT1 that encodes a putative bHLH transcription factor. The phenotypic analysis, gene expression studies, and genetic studies indicated that DYT1 is a major component in the genetic network that controls normal tapetum development and function during early anther development in Arabidopsis.
References:


KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. Plant Physiol 127, 803-816.


Chapter 2

Exploration of Global Gene Expression Profiles in the Arabidopsis Anther

Yujin Sun¹, Wei Zhang¹, Xiaohong Zhang, Asela Wijeratne, Wenlei Liu,

Naomi Altman and Hong Ma

¹: These authors contributed equally to this paper

Manuscript is in preparation
2.1 Abstract

Anther development and male meiosis are two critical processes in plant development. However, information on molecular mechanisms of anther development and male meiosis is still very limited. Gene regulation at the transcriptional level is believed to be critical for both development and male meiosis. In this chapter, the Affymetrix microarray technology was employed to compare the global expression profiles between wild type anthers and young inflorescences, and between wild type and mutant anthers. Arabidopsis anthers at approximately stages 4 to 6 from wild type plants and two meiotic mutants, solo dancers (sds) and male meiocyte death1 (mmd1), were collected and total RNA was extracted for microarray experiments. A total of 707 genes were identified as A/I (anther vs inflorescence) differential genes by a comparison of microarray data for wild type anthers and wild type young inflorescences (floral stages 1 to 9). These genes include previously identified anther preferential and/or meiosis genes. From the comparisons of anther gene expression profiles between wild type and sds, and between wild type and mmd1, we found that the expression levels of 240 genes and 46 genes were changed significantly in sds anthers and mmd1 anthers, respectively. Among the 240 genes with altered expression levels in the sds anther, 231 genes showed reduced expression, and 9 genes showed increased expression. It was also found that MMD1 itself and almost all MMD1-dependent genes are altered in the sds mutant, indicates that MMD1, encoding a putative PHD chromatin remodeling factor, may play a role in the SDS-dependent pathway.
2.2 Introduction

In flowering plants, male reproduction occurs in the stamen, which is composed of the vascular filament and the pollen-yielding anther. Anther development is an excellent model for organogenesis study in plants (Goldberg et al., 1993; Ma, 2005). Anther development can be divided into 14 stages, which can be grouped into two phases with male meiosis as the temporal divider. Phase one includes anther stages 1 to 8, and phase two is composed of anther stages 9 to 14 (Goldberg et al., 1993; Sanders et al., 1999).

In Arabidopsis, the anther primordium forms three cell layers at stage 1, called L1, L2 and L3, from the outside to the inside. These cell layers further develop into a structure that is a butterfly shaped in anther cross sections, with two larger lobes and two smaller lobes connected by central vascular and connective tissues. L1 layer cells differentiate into the epidermis, L3 layer cells differentiate into the center vascular and connective tissues, and L2 layer cells give rise to both somatic cells and sporogenous cells. At stage 2, the L2 layer cells differentiate into archesporial cells. The archesporial cells then divide and become primary parietal cells and primary sporogeneous cells at stage 3. The primary parietal cells further divide and differentiate into two layers of secondary parietal cells at stage 4. The outer secondary parietal cells give rise to the endothecium layer and the middle layer, while the inner parietal cells form the tapetum layer at stage 5. The primary sporogenous cells first differentiate into sporogenous cells at stage 4 and further develop into microsporocytes, also called pollen mother cells (PMCs), at stage 5. PMCs undergo meiosis during stages 5 to 6, and complete meiosis at stage 7, resulting in the formation of tetrads. Four haploid microspores are released from each tetrad at stage 8. Microspores then proceed into anther development phase two and develop into male gametes, pollen grains (Goldberg et al., 1993; Sanders et al., 1999; Ma, 2005).
It is logical to think that these specific developmental events are regulated by specific programs regulating gene expression, like the situations occurring in other developmental processes in plants, such as embryogenesis (Torres-Ruiz et al., 1996; Hardtke and Berleth, 1998), transition from vegetative to reproductive growth (Irish and Sussex, 1990; Weigel et al., 1992), floral organ identity determination (Yanofsky et al., 1990; Bowman and Meyerowitz, 1991; Coen and Meyerowitz, 1991; Weigel and Meyerowitz, 1994), shoot apical meristem pattern formation (Long et al., 1996; Lenhard et al., 2002), and root meristem pattern formation (Birnbaum et al., 2003). Indeed, to date it has been found that many critical meiotic genes are temporally and spatially meiosis-specific, such as \textit{AtSPO11-1} (Grelon et al., 2001), \textit{AtDMCI} (Klimyuk and Jones, 1997; Doutriaux et al., 1998), \textit{ASYNAPTIC1} (ASY1) (Caryl et al., 2000; Armstrong et al., 2002), and \textit{SOLO DANCERS} (\textit{SDS}) (Azumi et al., 2002). In Arabidopsis, \textit{SPOROCYTELESS/NOZZLE} (\textit{SPL/NZZ}) (Schiefthaler et al., 1999; Yang et al., 1999b; Azumi et al., 2002) and \textit{EXCESS MICROSPORCYTES1/EXTRA SPOROGENOUS CELL} (\textit{EMS1/EXS}) (Canales et al., 2002; Zhao et al., 2002) are two key genes essential for cell fate determination during early anther development. Undoubtedly, information on anther global expression profiles and anther preferential genes will facilitate the identification of some genes essential for anther development and/or male meiosis.

Meiosis is a key landmark in reproductive development, and usually is regarded as the final stage of the early anther development phase. It is comprised of a single round of DNA replication followed by two rounds of cell division. A coordinated series of events during meiotic prophase I, including homolog pairing, synapsis, result in the formation of bivalents. Meiotic recombination occurs between non-sister chromatids of the paired homologs. The homologous chromosomes are separated during anaphase I resulting in the reduction of the
number of chromosomes in the daughter nuclei. The second meiotic division separates the sister chromatids of each chromosome to give rise to haploid microspores (Armstrong and Jones, 2003; Ma, 2005).

In Arabidopsis, SDS is a unique meiosis-specific cyclin gene essential for male meiosis (Azumi et al., 2002). In the sds mutant, bivalents can not be formed during meiotic prophase I. The recombination rate in sds mutant is significantly reduced. In addition, some sds PMCs undergo apoptosis during meiosis. So SDS is not only required for homolog pairing and bivalent formation, but also is essential for normal PMC development. SDS is PMC-specific; its expression accumulates in meiotic prophase I and diminishes after meiosis. Although the genetic consequence of loss-of-function of SDS is clear, little is known about the molecular mechanism of the SDS function. Studies in yeast and human found that cyclins are the major regulators of cyclin dependent kinases (CDKs) that can phosphorylate downstream components, which further regulate cell cycle progression (Morgan, 1997; Murray, 2004).

In fission yeast, during meiotic recombination, chromatin remodeling process changes the conformation of chromatin, enabling meiotic DNA metabolism enzymes to be loaded onto the meiotic hotspots more easily (Mizuno et al., 1997; Mizuno et al., 2001; Hirota et al., 2003). The activities of some DNA-binding proteins such as transcription factors, activators and histone covalent modification proteins play important roles in chromatin remodeling (Mizuno et al., 1997; Mizuno et al., 2001; Hirota et al., 2003). In Arabidopsis, MALE MEIOCYTE DEATHI (MMD1) encodes a PHD domain nuclear protein and is preferentially expressed in meiotic cells (Yang et al., 2003). The PHD domain can potentially interact with proteins, DNA and RNA and is often found in proteins involved in chromatin remodeling (Aasland et al., 1995). In the mmd1 anthers, male meiosis is abnormal after diakinesis, and meiotic cells often die with chromosome
fragmentation and cytoplasmic shrinkage (Yang et al., 2003). The mmd1 mutant phenotype and the MMD1 protein sequence suggest that MMD1 may play important roles in transcriptional regulation and chromatin remodelling during male meiosis (Yang et al., 2003).

To characterize genes expressed during anther development and meiosis, researchers have tried many methods including conventional differential display and subtractive hybridization. For example, Rubinelli et al. (1998) isolated 13 anther-specific genes by subtractive hybridization between wild type and agamous (ag) young inflorescences. However, due to the limitations of previous methods, usually only a small number of genes with high expression levels could be detected, many other important genes with relatively low expression levels are very hard to be identified. The microarray technology provides a very good opportunity to study global gene expression patterns and has been well used to identify organ specific genes and to study functional pathways (Mussig et al., 2002; Menges et al., 2003; Ohgishi et al., 2004). The GeneChip® Arabidopsis ATH1 Genome Array is one of the best Arabidopsis genome arrays, representing about 80 percent of Arabidopsis nuclear genes (Affymetrix, Santa Clara, CA). In this study, microarray experiments were performed by using total RNA extracted from wild type young inflorescences (floral stages 1 to 9) and anthers (anther stages 4 to 6) from wild type, sds and mmd1. After statistical analyses, we identified 707 genes whose expression levels in wild type anthers are at least 2 fold of those in wild type young inflorescences. These genes were defined as A/I differential genes, and were classified into ten functional categories according to their putative functional domains and Gene Ontology (GO) annotations from the TAIR database (www.arabidopsis.org). We also identified 240 genes potentially regulated by SDS and 46 genes affected by MMD1. Interestingly, we found that some putative DNA metabolism genes are included in the SDS-regulating gene category and most of
the MMD1-regulating genes are also included in the SDS-regulating category, suggesting that certain chromosome remodeling processes may be part of the SDS-dependent pathway in male meiosis and MMD1 may act downstream of SDS.

2.3 Materials and Methods

Plant material collection

Wild type tissue samples were collected from the Arabidopsis thaliana ecotype Landsberg erecta (Ler). Both sds and mmd1 are under the Ler background, and were previously generated by Ds-transposon insertion in our lab (Azumi et al., 2002; Yang et al., 2003). All seeds were chilled at 4°C for 4 days, then were planted and grown on soil at 23°C with the long-day light cycle (16 hours light and 8 hours dark) in a greenhouse. All the tissues were collected approximately one week after plant bolting (approximately 28 to 35 days after planting seeds). Wild type young inflorescences were collected with shoot apical meristem and flower buds at floral stages 1 to 9. Anthers at stages 4 to 6 from wild type, sds and mmd1 were dissected out from small young buds under a dissection microscope. All the tissues were immediately frozen in liquid nitrogen for storage after collection. During collection, anthers were also examined frequently by using the 0.1% Toluidine blue O staining method under a microscope to check anther stages. Two independently grown populations of wild type young inflorescences and anthers from each genotype were harvested and used to extract total RNA individually.
Microarray experiments

The microarray experiments were performed as described in Affymetrix GeneChip® Expression Analysis Overview (Affymetrix, Santa Clara, CA). Total RNA was extracted and purified by using the RNeasy Plant Kit according to the manufacturer’s instruction (Qiagen, Valencia, CA). Double-stranded cDNA was synthesized by using SuperScript Choice System (Invitrogen, Carlsbad, CA) with 10 µg of total RNA and 100 pmol oligo (dT)$_{24}$ primer containing a 5' T7 RNA polymerase promoter, and was purified by the GeneChip® Sample Cleanup Module (Affymetrix, Santa Clara, CA). Biotin-labeled target cRNA was prepared from cDNA by in vitro transcription using the BioArray High-Yield RNA Transcript Labeling Kit (Enzo Biochem, New York) in the presence of biotinylated UTP and CTP. After purification by the GeneChip® Sample Cleanup Module (Affymetrix, Santa Clara, CA), cRNA was fragmented in the Fragmentation Buffer (Affymetrix, Santa Clara, CA) to produce RNA fragments with sizes of approximately 35 to 200 bases. Fifteen µg of fragmented cRNA was used to hybridize to the GeneChip Arabidopsis ATH1 Chips (Affymetrix, Santa Clara, CA). The Affymetrix Arabidopsis genome ATH1 array contains 22,810 probe sets. Each probe set is composed a set of 11 pairs of oligonucleotides with either perfect match (PM) or mismatch (MM) for a particular gene. Among the 22,810 probe sets, 22,746 probe sets are designed to be complementary to Arabidopsis sequences, and the other 64 probe sets are used as internal controls. The hybridization, washing, staining, scanning, and raw data collection were carried out at the Pennsylvania State University DNA Microarray Facility.
**Microarray data analysis**

The microarray data analysis was performed in the Department of Health Evaluation Sciences, Hershey Medical School at the Pennsylvania State University. The raw data were converted into expression data by using the R-affymetrix package in Bioconductor (version 1.3.25, Irizarry, Gautier, and Bolstad). Normalization was done by the “RMA” method (Irizarry et al., 2003), which assumes the distribution of probe intensities for each array in the dataset should be the same. For pair-wise tissue comparisons, genes with significantly different expression were identified using the Significance Analysis of Microarrays (SAM) method (Tusher et al., 2001). Gene clustering was carried out using the GENESIS program (Release: 1.6.0 Beta 1) (Sturn et al., 2002). Complete linkage hierarchical clustering method was used to perform the cluster analysis for A/I differential genes and genes regulated by SDS during early anther development (Han and Kamber, 2001; Sturn et al., 2002).

**RT-PCR**

Two µg of total RNA from each sample for microarray experiments was used for reverse transcription to synthesize cDNA (Invitrogen, Carlsbad, CA). The cDNA was then used as the template for PCR using gene specific primers. UBQ1 was used as the internal control. The PCR reaction was carried out as described previously (Zhang et al., 2005).
2.4 Results and Discussion

Microarray data reproducibility and statistical analysis

To assess the reproducibility of the microarray data, two independent populations from each genotype were used to collect tissues for microarray experiments. The hybridization signals of microarray experiments were pre-scaled to a same target intensity of 150. All the expression data sets were analyzed and normalized to make the different data sets comparable by the “RMA” method (Irizarry et al., 2003). Pearson correlation coefficients ($r$) were calculated for the duplicated data sets for the same genotype. All the $r$-values were larger than 0.94, indicating a small difference between two biological replicates (Table 2.1). Therefore, average expression values were calculated for each genotype and used for further analyses unless otherwise specified.

Global gene expression profiles in anthers and young inflorescences

We used both intensity values and expression call values as the criteria to identify whether or not a gene is expressed. As discussed by Zhang et al. (2005), signal intensity below 50 is regarded as background or an unreliable signal for gene expression. So we choose 50 as the cut off for expression. In the microarray data output, each probe is also assigned an expression call: present (P), marginal (M) or absent (A) (Affymetrix, Santa Clara, CA). A gene was considered to be expressed if the two replicated probe sets assigned are PM or PP calls. With that standard, the total numbers of genes expressed in each tissue are as follows: 13028 (57.3%) in wild type young inflorescences, 12862 (56.5%) in wild type anthers, 12692 (55.8%) in $sds$ anthers, and 12475 (54.8%) in $mmd1$ anthers (Table 2.1). The numbers of commonly expressed
genes in wild type young inflorescences, *sds* anther and *mmd1* anther versus wild type anther, respectively, are as follows: 12085 (53.1%), 12250 (53.8%), and 12155 (53.4%).
Table 2.1. Pearson correlation coefficient (r-value) of two replicates for each genotype.

A.

<table>
<thead>
<tr>
<th></th>
<th>In</th>
<th>WT-An</th>
<th>sds-An</th>
<th>mmd1-An</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation coefficient</td>
<td>0.966</td>
<td>0.971</td>
<td>0.986</td>
<td>0.943</td>
</tr>
<tr>
<td>Number of expressed genes</td>
<td>13028</td>
<td>12862</td>
<td>12692</td>
<td>12475</td>
</tr>
</tbody>
</table>

B.

<table>
<thead>
<tr>
<th></th>
<th>S12</th>
<th>Si</th>
<th>St</th>
<th>Lf</th>
<th>Rt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation coefficient</td>
<td>0.983</td>
<td>0.991</td>
<td>0.995</td>
<td>0.985</td>
<td>0.99</td>
</tr>
</tbody>
</table>

All the tissues collected for microarray experiments are in the same ecotype background. All the microarray data was normalized using the “RMA” method. The r-values were calculated for each genotype in Excel. A. The r-values and numbers of expressed genes for In (wild type young inflorescences), WT-An (wild type anthers), sds-An (sds anthers), and mmd1-An (mmd1 anthers). If a gene has an average expression level higher than 50 and has both “Present” calls in two replicates or has one “Present” call and one “Marginal” call, it was considered as an expressed gene. B. The r-values for S12 (wild type stage-12 flowers), Si (wild type siliques), St (wild type stems), Lf (wild type leaves), Rt (wild type roots).
Identification of A/I (Anther vs Inflorescence) differential genes

The transcripts of anther preferentially expressed genes should be enriched in anthers when compared with those in young inflorescences. Therefore, these genes can be identified by comparing gene expression profiles in wild type anthers with those in wild type young inflorescences. Since male meiosis occurs at anther stages 5 and 6, the meiosis-preferential genes would be also included in this category.

The SAM method was applied to identify the differentially expressed genes in wild type anthers or young inflorescences (Tusher et al., 2001). When the false discovery rate (FDR) was set as 0.05, we found 1,507 genes with significantly increased expression and 1,117 genes with significantly decreased expression in wild type anthers compared to wild type young inflorescences. Further analysis was focused on these 1507 genes. Although each probe set usually represents a single gene, some probe sets can hybridize with more than one gene and they are defined as ambiguous probe sets. All the data related with ambiguous probe sets were removed. As discussed above, we also focused on genes with intensity value at or above 50 in the wild type anther. To get a list of more dramatically preferentially expressed genes, only the gene whose expression level increased more than 2 fold in wild type anthers compared to that in young inflorescences were included for further analysis. A total of 707 genes were identified and were defined as A/I differential genes.

Among A/I differential genes, we found many previously reported genes essential for early anther development and/or male meiosis (Table 2.2). One example is SDS (At1g14750), a meiotic-specific gene essential for homologous chromosome pairing and bivalents formation in meiotic prophase I. SDS expression in anthers is 2.3 fold of that in young inflorescences (Azumi et al., 2002). Some other genes that are crucial for male meiosis are AtSPO11-1 (At3g13170,
5.52 fold), \textit{AtDMC1} (At3g22880, 6.2 fold), \textit{MMD1} (At1g66170, 7.5 fold), \textit{ASY1} (At1g67370, 4.5 fold) and \textit{MS5} (At4g20900, 4.7 fold) (Klimyuk and Jones, 1997; Glover et al., 1998; Caryl et al., 2000; Grelon et al., 2001; Yang et al., 2003). A number of known genes essential for early anther development were also found in the list, such as \textit{SPL/NZZ} (At4g27330, 5.8 fold), \textit{MYB33} (At5g06100, 2.5 fold), \textit{MYB65} (At3g11440, 4.1 fold) and \textit{ABORTED MICROSPORES (AMS)} (AT2G16910, 38.8 fold) (Schiefthaler et al., 1999; Yang et al., 1999b; Sorensen et al., 2003; Millar and Gubler, 2005). In contrast, the floral meristem-specific genes, like \textit{LEAFY (LFY)} (At5g61850) and \textit{APETALAI (API)} (At1g69120) that have obviously lower expression levels in anthers compared with young inflorescences, are in the list of genes with significantly decreased expression in young inflorescences (Irish and Sussex, 1990; Weigel et al., 1992).
Table 2.2. Expression patterns of known genes in the list of A/I differential genes.

<table>
<thead>
<tr>
<th>Locus Identifier</th>
<th>Gene name</th>
<th>An/In</th>
<th>An</th>
<th>In</th>
<th>S12</th>
<th>Si</th>
<th>St</th>
<th>Lf</th>
<th>Rt</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT4G14080</td>
<td>A6</td>
<td>49.78</td>
<td>3692.1</td>
<td>74.2</td>
<td>57.5</td>
<td>1.2</td>
<td>1.6</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>AT5G07230</td>
<td>A9</td>
<td>56.30</td>
<td>5099.8</td>
<td>90.6</td>
<td>197.5</td>
<td>8.2</td>
<td>2.8</td>
<td>7.8</td>
<td>3.3</td>
</tr>
<tr>
<td>AT3G42960</td>
<td>ATA1</td>
<td>43.25</td>
<td>2090.4</td>
<td>48.3</td>
<td>71.6</td>
<td>11.3</td>
<td>15.1</td>
<td>12.0</td>
<td>5.8</td>
</tr>
<tr>
<td>AT4G28395</td>
<td>ATA7</td>
<td>19.66</td>
<td>283.2</td>
<td>14.4</td>
<td>498.8</td>
<td>9.5</td>
<td>4.4</td>
<td>11.0</td>
<td>8.2</td>
</tr>
<tr>
<td>AT3G15400</td>
<td>ATA20</td>
<td>10.38</td>
<td>2167.7</td>
<td>208.8</td>
<td>7099.5</td>
<td>7.6</td>
<td>10.7</td>
<td>14.4</td>
<td>23.6</td>
</tr>
<tr>
<td>AT4G27330</td>
<td>SPL</td>
<td>5.76</td>
<td>402.5</td>
<td>69.9</td>
<td>92.5</td>
<td>2.5</td>
<td>3.0</td>
<td>3.3</td>
<td>1.8</td>
</tr>
<tr>
<td>AT5G06100</td>
<td>MYB33</td>
<td>2.51</td>
<td>238.8</td>
<td>95.1</td>
<td>99.0</td>
<td>48.3</td>
<td>63.7</td>
<td>73.8</td>
<td>62.9</td>
</tr>
<tr>
<td>AT3G11440</td>
<td>MYB65</td>
<td>4.15</td>
<td>335.5</td>
<td>80.9</td>
<td>121.1</td>
<td>39.0</td>
<td>43.8</td>
<td>61.5</td>
<td>87.7</td>
</tr>
<tr>
<td>AT2G16910</td>
<td>AMS</td>
<td>38.76</td>
<td>348.9</td>
<td>9.0</td>
<td>39.5</td>
<td>1.2</td>
<td>7.6</td>
<td>2.8</td>
<td>4.7</td>
</tr>
<tr>
<td>AT3G11980</td>
<td>MS2</td>
<td>76.75</td>
<td>690.8</td>
<td>9.0</td>
<td>66.5</td>
<td>1.5</td>
<td>2.0</td>
<td>1.4</td>
<td>4.8</td>
</tr>
<tr>
<td>AT4G20900</td>
<td>MS5</td>
<td>4.72</td>
<td>75.5</td>
<td>16.0</td>
<td>16.0</td>
<td>24.0</td>
<td>24.0</td>
<td>33.3</td>
<td>9.1</td>
</tr>
<tr>
<td>AT3G13170</td>
<td>SPO11</td>
<td>5.52</td>
<td>59.7</td>
<td>10.8</td>
<td>8.3</td>
<td>10.7</td>
<td>22.2</td>
<td>15.0</td>
<td>21.8</td>
</tr>
<tr>
<td>AT1G67370</td>
<td>ASY1</td>
<td>4.53</td>
<td>220.6</td>
<td>48.7</td>
<td>23.5</td>
<td>28.6</td>
<td>16.5</td>
<td>28.7</td>
<td>25.9</td>
</tr>
<tr>
<td>AT3G22880</td>
<td>DMC1</td>
<td>6.19</td>
<td>2199.5</td>
<td>355.3</td>
<td>150.8</td>
<td>58.4</td>
<td>49.0</td>
<td>58.8</td>
<td>42.4</td>
</tr>
<tr>
<td>AT1G14750</td>
<td>SDS</td>
<td>2.27</td>
<td>80.5</td>
<td>35.4</td>
<td>19.6</td>
<td>17.9</td>
<td>21.4</td>
<td>21.8</td>
<td>14.8</td>
</tr>
<tr>
<td>AT1G66170</td>
<td>MMD1</td>
<td>7.51</td>
<td>83.0</td>
<td>11.1</td>
<td>6.8</td>
<td>4.7</td>
<td>10.6</td>
<td>18.2</td>
<td>6.7</td>
</tr>
<tr>
<td>AT3G27730</td>
<td>MER3/RCK1</td>
<td>2.60</td>
<td>81.7</td>
<td>31.4</td>
<td>17.4</td>
<td>5.0</td>
<td>9.7</td>
<td>3.5</td>
<td>12.1</td>
</tr>
<tr>
<td>AT4G17380</td>
<td>MSH4</td>
<td>2.41</td>
<td>120.6</td>
<td>50.2</td>
<td>2.9</td>
<td>13.6</td>
<td>15.1</td>
<td>14.0</td>
<td>15.1</td>
</tr>
<tr>
<td>AT1G12790</td>
<td>PTD</td>
<td>3.18</td>
<td>175.1</td>
<td>55.1</td>
<td>57.1</td>
<td>69.5</td>
<td>57.9</td>
<td>105.0</td>
<td>88.0</td>
</tr>
<tr>
<td>AT5G10140</td>
<td>FLF</td>
<td>23.75</td>
<td>251.5</td>
<td>10.6</td>
<td>37.8</td>
<td>13.8</td>
<td>14.8</td>
<td>22.1</td>
<td>3.1</td>
</tr>
<tr>
<td>AT5G07280</td>
<td>EMS1/EXS</td>
<td>1.75</td>
<td>1062.9</td>
<td>608.9</td>
<td>282.0</td>
<td>342.8</td>
<td>39.5</td>
<td>70.2</td>
<td>29.0</td>
</tr>
<tr>
<td>AT5G61850</td>
<td>LFY</td>
<td>0.13</td>
<td>46.7</td>
<td>350.8</td>
<td>36.8</td>
<td>19.1</td>
<td>19.1</td>
<td>28.4</td>
<td>18.7</td>
</tr>
<tr>
<td>AT1G69120</td>
<td>API</td>
<td>0.09</td>
<td>137.0</td>
<td>1524.0</td>
<td>948.8</td>
<td>27.0</td>
<td>21.5</td>
<td>23.6</td>
<td>18.6</td>
</tr>
</tbody>
</table>

An/In: the fold change of gene expression levels in wild type anthers (anther stages 4 to 6) compared to wild type young inflorescences (floral stages 1 to 9); An: wild type anthers; In: wild type young inflorescences; S12: wild type stage-12 flowers; Si: wild type siliques; St: wild type stems; Lf: wild type leaves; Rt: wild type roots.
Expression patterns of A/I differential genes

To gain further insights into the expression patterns of the A/I differential genes, we used the previously reported microarray data from our lab on several other wild type tissues, including stage-12 flowers (S12), siliques (Si), stems (St), leaves (Lf), and roots (Rt). First, the microarray data for wild type anthers (An), young inflorescences (In) and on all the other five tissues were normalized together as described above. The expression patterns of 707 A/I differential genes were obtained and the Log2 values were calculated. Log2 value was defined as Log2 (ratio of a gene’s intensity in tissue A to wild type anthers). For example, if the value is 1, the gene expression in tissue A is 2 fold of that in wild type anthers; if the value is –2, the gene expression in tissue A is a quarter of that in wild type anthers. The Log2 values were used for cluster analysis to identify clusters of co-expressed genes by using the GENESIS software (Sturn et al., 2002). The expression data of A/I differential genes were clustered by hierarchical complete linkage clustering model with Pearson squared distance and five clusters were estimated and were named as C1 to C5 (Figure 2.1). C1 had 74 genes (10.5%), and most of them were highly expressed in Rt and An. C2 included 107 genes (15.1%), and most of them were lowly expressed in An and In. In C3, there were 110 genes (15.6%), and most of them were highly expressed in S12 and An. C4 had 390 genes (55.2%), and most of them were expressed more highly in wild type anthers than any other tissues we compared. All the known function genes identified in the list of A/I differential genes were belonged to the C4 cluster.
Figure 2.1. Expression patterns of 707 A/I differential genes. Log$_2$ values, defined as Log$_2$ (ratio of intensity in tissue A to that in wild type anthers), were used for cluster analysis. Gene cluster analysis was performed with hierarchical complete linkage clustering model and Pearson squared distance by the Genesis software. A/I differential genes were classed into five clusters, as C1 to C5. Genes more highly expressed in anthers are highlighted in green. Genes with relatively low expression levels in anthers are highlighted in red. An, wild type anthers at stages 4 to 6; In, wild type young inflorescences at stages 1 to 9; S12, wild type stage-12 flowers; Si, wild type siliques; St, wild type stems; Lf, wild type leaves; Rt, wild type roots. In/An, S12/An, Si/An, St/An, Lf/An, Rt/An refer to the corresponding Log$_2$ values.
Identification of anther preferential genes and anther specific genes

Since some known genes that are essential for male meiosis and/or early anther development are anther preferential or anther specific. I would like to identify these genes. To achieve this goal, the normalized microarray data were used for analysis. I started with the list of A/I differential genes identified above and chose genes whose expression in wild type anthers is at least two fold higher than that in any other wild type tissues tested in this study. By this standard, 250 genes were identified and were defined as anther preferential genes. Although all these anther preferential genes are only highly expressed in the wild type anther, some of them are still expressed in some other tissues. I further checked the expression of these genes in all the seven tissues to define genes that are only expressed in anthers. I started with anther preferential genes and only chose genes whose expression level in wild type anthers is above 50 and that in all the other tissues is below 50. By this standard, 106 genes were identified and were defined as anther specific genes (Table 2.3), including SDS, ROCK ROLLER1 (RCK1), MS5, SPO11-1, MMD1, flowering locus F (FLF), ASY1 (Glover et al., 1998; Sheldon et al., 1999; Grelon et al., 2001; Armstrong et al., 2002; Yang et al., 2003; Chen et al., 2005).
<table>
<thead>
<tr>
<th>Locus</th>
<th>An</th>
<th>Infl</th>
<th>S12</th>
<th>Sili</th>
<th>St</th>
<th>Lf</th>
<th>Rt</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1G01280</td>
<td>431</td>
<td>2</td>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>cytochrome P450 family protein</td>
</tr>
<tr>
<td>At1G03390</td>
<td>317</td>
<td>7</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>transferase family protein; contains Pfam transferase family domain PF002458</td>
</tr>
<tr>
<td>At1G06320</td>
<td>168</td>
<td>42</td>
<td>21</td>
<td>32</td>
<td>31</td>
<td>43</td>
<td>13</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>At1G07860</td>
<td>139</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>9</td>
<td>expressed protein</td>
</tr>
<tr>
<td>At1G10710</td>
<td>123</td>
<td>38</td>
<td>17</td>
<td>21</td>
<td>19</td>
<td>28</td>
<td>9</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>At1G14750</td>
<td>80</td>
<td>35</td>
<td>20</td>
<td>18</td>
<td>21</td>
<td>22</td>
<td>15</td>
<td>cyclin, putative (SDS) identical to cyclin-like protein</td>
</tr>
<tr>
<td>At1G15190</td>
<td>253</td>
<td>7</td>
<td>4</td>
<td>10</td>
<td>5</td>
<td>9</td>
<td>5</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>At1G15320</td>
<td>71</td>
<td>10</td>
<td>13</td>
<td>12</td>
<td>16</td>
<td>14</td>
<td>10</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>At1G15320</td>
<td>68</td>
<td>23</td>
<td>10</td>
<td>6</td>
<td>4</td>
<td>11</td>
<td>15</td>
<td>expressed protein</td>
</tr>
<tr>
<td>At1G23700</td>
<td>201</td>
<td>33</td>
<td>30</td>
<td>36</td>
<td>32</td>
<td>38</td>
<td>26</td>
<td>protein kinase family protein contains protein kinase domain</td>
</tr>
<tr>
<td>At1G30800</td>
<td>123</td>
<td>17</td>
<td>13</td>
<td>18</td>
<td>19</td>
<td>19</td>
<td>9</td>
<td>expressed protein ; expression supported by MPSS</td>
</tr>
<tr>
<td>At1G33500</td>
<td>53</td>
<td>12</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>At1G43900</td>
<td>80</td>
<td>24</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>pentatricopeptide (PPR) repeat-containing protein</td>
</tr>
<tr>
<td>At1G49000</td>
<td>62</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>expressed protein</td>
</tr>
<tr>
<td>At1G50470</td>
<td>88</td>
<td>29</td>
<td>22</td>
<td>18</td>
<td>26</td>
<td>27</td>
<td>16</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>At1G54230</td>
<td>104</td>
<td>23</td>
<td>18</td>
<td>21</td>
<td>16</td>
<td>14</td>
<td>14</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>At1G661070</td>
<td>1229</td>
<td>30</td>
<td>20</td>
<td>11</td>
<td>12</td>
<td>14</td>
<td>5</td>
<td>plant defensin-fusion protein, putative (PDF2.4) plant defensin family member</td>
</tr>
<tr>
<td>At1G62130</td>
<td>161</td>
<td>43</td>
<td>49</td>
<td>42</td>
<td>50</td>
<td>50</td>
<td>42</td>
<td>AAA-type ATPase family protein</td>
</tr>
<tr>
<td>At1G63950</td>
<td>130</td>
<td>13</td>
<td>12</td>
<td>27</td>
<td>23</td>
<td>9</td>
<td>6</td>
<td>heavy-metal-associated domain-containing protein</td>
</tr>
<tr>
<td>At1G66110</td>
<td>79</td>
<td>16</td>
<td>10</td>
<td>20</td>
<td>19</td>
<td>20</td>
<td>12</td>
<td>hypothetical protein contains Pfam profile PF04510; Family of unknown function (DUF577)</td>
</tr>
<tr>
<td>At1G66170</td>
<td>83</td>
<td>7</td>
<td>5</td>
<td>11</td>
<td>18</td>
<td>7</td>
<td>5</td>
<td>PHD finger family protein (MMD1) contains Pfam profile: PF00628: PHD-finger</td>
</tr>
<tr>
<td>At1G67370</td>
<td>221</td>
<td>49</td>
<td>24</td>
<td>29</td>
<td>17</td>
<td>29</td>
<td>26</td>
<td>meiotic asynaptic mutant 1 (ASY1)</td>
</tr>
<tr>
<td>At1G69500</td>
<td>1636</td>
<td>33</td>
<td>40</td>
<td>7</td>
<td>11</td>
<td>7</td>
<td>5</td>
<td>cytochrome P450 family protein</td>
</tr>
<tr>
<td>At1G73050</td>
<td>488</td>
<td>6</td>
<td>3</td>
<td>15</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>(R)-mandelonitrile lyase, putative / (R)-oxynitrilase</td>
</tr>
<tr>
<td>At2G01810</td>
<td>177</td>
<td>26</td>
<td>20</td>
<td>31</td>
<td>23</td>
<td>21</td>
<td>19</td>
<td>PHD finger family protein</td>
</tr>
<tr>
<td>At2G14800</td>
<td>91</td>
<td>38</td>
<td>27</td>
<td>25</td>
<td>36</td>
<td>23</td>
<td>23</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>At2G16910</td>
<td>349</td>
<td>9</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>basic helix-loop-helix (bHLH) family protein</td>
</tr>
<tr>
<td>At2G17170</td>
<td>81</td>
<td>14</td>
<td>6</td>
<td>10</td>
<td>13</td>
<td>15</td>
<td>1</td>
<td>protein kinase family protein contains protein kinase domain</td>
</tr>
<tr>
<td>At2G24800</td>
<td>486</td>
<td>20</td>
<td>16</td>
<td>26</td>
<td>19</td>
<td>21</td>
<td>15</td>
<td>peroxidase, putative similar to peroxidase</td>
</tr>
<tr>
<td>At2G28090</td>
<td>146</td>
<td>49</td>
<td>7</td>
<td>4</td>
<td>14</td>
<td>14</td>
<td>12</td>
<td>heavy-metal-associated domain-containing protein</td>
</tr>
<tr>
<td>At2G29600</td>
<td>183</td>
<td>28</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>33</td>
<td>25</td>
<td>kelch repeat-containing F-box family protein</td>
</tr>
<tr>
<td>At2G29800</td>
<td>75</td>
<td>15</td>
<td>6</td>
<td>18</td>
<td>17</td>
<td>30</td>
<td>7</td>
<td>kelch repeat-containing F-box family protein</td>
</tr>
<tr>
<td>At2G31210</td>
<td>176</td>
<td>17</td>
<td>20</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>basic helix-loop-helix (bHLH) family protein</td>
</tr>
<tr>
<td>At2G32350</td>
<td>589</td>
<td>26</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>ubiquitin family protein</td>
</tr>
<tr>
<td>At2G32360</td>
<td>53</td>
<td>1</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>ubiquitin family protein</td>
</tr>
<tr>
<td>At2G33000</td>
<td>118</td>
<td>36</td>
<td>23</td>
<td>30</td>
<td>34</td>
<td>31</td>
<td>26</td>
<td>ubiquitin-associated (UBA)/NS-N domain-containing protein-related</td>
</tr>
<tr>
<td>At2G34230</td>
<td>57</td>
<td>4</td>
<td>13</td>
<td>10</td>
<td>1</td>
<td>6</td>
<td>11</td>
<td>hypothetical protein contains Pfam profiles PF04780: Protein of unknown function (DUF629)</td>
</tr>
<tr>
<td>At2G42940</td>
<td>407</td>
<td>15</td>
<td>45</td>
<td>18</td>
<td>11</td>
<td>14</td>
<td>12</td>
<td>DNA-binding family protein contains a At hook motif</td>
</tr>
<tr>
<td>At2G44170</td>
<td>146</td>
<td>46</td>
<td>24</td>
<td>30</td>
<td>23</td>
<td>28</td>
<td>19</td>
<td>myristoyl-CoA:protein N-myristoyltransferase 1 (NMT1)</td>
</tr>
<tr>
<td>At2G44590</td>
<td>117</td>
<td>17</td>
<td>19</td>
<td>29</td>
<td>20</td>
<td>25</td>
<td>19</td>
<td>dynamin-like protein D (DL1D)</td>
</tr>
<tr>
<td>At2G46480</td>
<td>66</td>
<td>22</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>13</td>
<td>glycosyl transferase family 8 protein</td>
</tr>
<tr>
<td>At3G05720</td>
<td>70</td>
<td>26</td>
<td>35</td>
<td>16</td>
<td>11</td>
<td>15</td>
<td>7</td>
<td>importin alpha-1 subunit</td>
</tr>
<tr>
<td>At3G12420</td>
<td>110</td>
<td>21</td>
<td>28</td>
<td>13</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>At3G12850</td>
<td>89</td>
<td>21</td>
<td>11</td>
<td>28</td>
<td>19</td>
<td>33</td>
<td>19</td>
<td>COP9 signalosome complex-related / CSN complex-related protein</td>
</tr>
<tr>
<td>At3G13170</td>
<td>60</td>
<td>11</td>
<td>8</td>
<td>11</td>
<td>22</td>
<td>15</td>
<td>22</td>
<td>DNA topoisomerase VIA (SPO11-1) identical to AtSPO11-1</td>
</tr>
<tr>
<td>At3G13220</td>
<td>795</td>
<td>26</td>
<td>26</td>
<td>13</td>
<td>15</td>
<td>11</td>
<td>9</td>
<td>ABC transporter family protein</td>
</tr>
<tr>
<td>At3G13420</td>
<td>96</td>
<td>25</td>
<td>14</td>
<td>22</td>
<td>17</td>
<td>14</td>
<td>13</td>
<td>expressed protein</td>
</tr>
<tr>
<td>At3G16380</td>
<td>189</td>
<td>41</td>
<td>36</td>
<td>24</td>
<td>46</td>
<td>43</td>
<td>25</td>
<td>polyadenylate-binding protein</td>
</tr>
<tr>
<td>At3G21000</td>
<td>235</td>
<td>13</td>
<td>40</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>13</td>
<td>expressed protein ; expression supported by MPSS</td>
</tr>
<tr>
<td>At3G21620</td>
<td>142</td>
<td>6</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>24</td>
<td>early-responsive to dehydration protein-related / ERD protein-related</td>
</tr>
<tr>
<td>At3G23770</td>
<td>498</td>
<td>25</td>
<td>25</td>
<td>9</td>
<td>22</td>
<td>21</td>
<td>13</td>
<td>glycosyl hydrolase family 17 protein similar to A6 anther-specific protein</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td>Value 5</td>
<td>Value 6</td>
<td>Value 7</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>At3G24230</td>
<td>340 23</td>
<td>32 39</td>
<td>21 31</td>
<td>18</td>
<td>pectate lyase family protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G255990</td>
<td>69 25</td>
<td>34 34</td>
<td>34 20</td>
<td>20</td>
<td>DNA-binding protein Gl-1-related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G27730</td>
<td>82 31</td>
<td>17 5 10</td>
<td>4 12</td>
<td>ATP-dependent DNA helicase, AtMER3/RCK (Rock Roller)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G28470</td>
<td>759 40</td>
<td>7 11 16</td>
<td>15 7</td>
<td>myb family transcription factor (MYB35)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G45000</td>
<td>61 12</td>
<td>1 1 1</td>
<td>1 1</td>
<td>SNF7 family protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G49830</td>
<td>72 17</td>
<td>7 10 8</td>
<td>26</td>
<td>DNA helicase-related similar to DNA helicase GI:4521249 from [Mus musculus]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G55580</td>
<td>99 32</td>
<td>32 16</td>
<td>18 20</td>
<td>regulator of chromosome condensation (RCC1) family protein UVB-resistance protein UVR8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G55780</td>
<td>252 39</td>
<td>21 31</td>
<td>34 29</td>
<td>glycoyl hydrolase family 17 protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G57370</td>
<td>170 13</td>
<td>14 17</td>
<td>5 7</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G57440</td>
<td>128 47</td>
<td>23 33</td>
<td>40 24</td>
<td>myb family transcription factor (MYB35)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G57620</td>
<td>135 5</td>
<td>39 3 5</td>
<td>6 2</td>
<td>glyoxal oxidase-related contains</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G57960</td>
<td>77 13</td>
<td>16 22</td>
<td>27 18</td>
<td>emsy N terminus domain-containing protein / ENt domain-containing protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G58280</td>
<td>100 12</td>
<td>19 22</td>
<td>29 19</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G60010</td>
<td>125 36</td>
<td>47 38</td>
<td>33 32</td>
<td>E3 ubiquitin ligase SCF complex subunit SKP1/ASK1 (At13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G60020</td>
<td>77 6</td>
<td>6 5 5</td>
<td>5 3</td>
<td>E3 ubiquitin ligase SCF complex subunit SKP1/ASK1 (At15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G60280</td>
<td>207 21</td>
<td>11 14</td>
<td>23 32</td>
<td>uclacyanin 3 (UCC3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G60580</td>
<td>245 45</td>
<td>46 37</td>
<td>40 17</td>
<td>zinc finger (C2H2 type) family protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G61340</td>
<td>124 20</td>
<td>35 45</td>
<td>19 21</td>
<td>F-box family protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3G61810</td>
<td>105 2</td>
<td>5 3 3</td>
<td>2 3</td>
<td>glycoyl hydrolase family 17 protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G04690</td>
<td>105 29</td>
<td>22 29</td>
<td>25 32</td>
<td>F-box family protein (FBX15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G09080</td>
<td>97 13</td>
<td>4 13</td>
<td>6 10</td>
<td>chloroplast outer membrane protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G18490</td>
<td>68 21</td>
<td>14 6</td>
<td>16 25</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G19130</td>
<td>105 46</td>
<td>13 38</td>
<td>19 15</td>
<td>replication protein-related protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G19620</td>
<td>54 13</td>
<td>3 1 3</td>
<td>2 1</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G19980</td>
<td>100 25</td>
<td>28 20</td>
<td>9 20</td>
<td>expressed protein ; expression supported by MPSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G20900</td>
<td>75 16</td>
<td>16 24</td>
<td>24 33</td>
<td>male sterility MS5 / pollenless 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G21760</td>
<td>90 40</td>
<td>38 39</td>
<td>34 27</td>
<td>glycoyl hydrolase family 1 protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G24860</td>
<td>220 44</td>
<td>31 20</td>
<td>32 47</td>
<td>AAA-type ATPase family protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G26020</td>
<td>52 10</td>
<td>5 4 7</td>
<td>4 3</td>
<td>expressed protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G28130</td>
<td>61 9</td>
<td>5 3 6</td>
<td>3 5</td>
<td>diacylglycerol kinase accessory domain-containing protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G30040</td>
<td>53 5</td>
<td>8 23</td>
<td>13 20</td>
<td>aspartyl protease family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G33870</td>
<td>140 28</td>
<td>28 41</td>
<td>34 41</td>
<td>17 peroxidase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4G37780</td>
<td>51 2</td>
<td>5 2 1</td>
<td>1 5</td>
<td>myb family transcription factor (MYB87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G10140</td>
<td>252 11</td>
<td>38 14</td>
<td>15 22</td>
<td>MADS-box protein flowering locus F (FLF) identical to FLOWERING LOCUS C protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G14980</td>
<td>68 8</td>
<td>32 27</td>
<td>16 14</td>
<td>9 esterase/lipase/thioesterase family protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G15480</td>
<td>53 19</td>
<td>12 18</td>
<td>17 18</td>
<td>zinc finger (C2H2 type) family protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G17830</td>
<td>87 14</td>
<td>12 15</td>
<td>7 16</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G24820</td>
<td>1009 27</td>
<td>11 15</td>
<td>6 16</td>
<td>aspartyl protease family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G25950</td>
<td>898 11</td>
<td>1 25 1</td>
<td>1 1</td>
<td>hypothetical protein various predicted proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G35600</td>
<td>86 7</td>
<td>2 4 8</td>
<td>5 6</td>
<td>histone deacetylase, putative (HDA7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G37420</td>
<td>113 21</td>
<td>4 7 12</td>
<td>15 11</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G40940</td>
<td>264 5</td>
<td>4 2 10</td>
<td>3 1</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G44330</td>
<td>90 1</td>
<td>1 2 1</td>
<td>1 1</td>
<td>male sterility MS5 family protein similar to male sterility MS5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G48390</td>
<td>69 9</td>
<td>3 15 9</td>
<td>12 4</td>
<td>tetratricopeptide repeat (tPR)-containing protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G48420</td>
<td>97 34</td>
<td>23 32</td>
<td>35 45</td>
<td>27 hypothetical protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G52160</td>
<td>1108 28</td>
<td>23 8 10</td>
<td>4 4</td>
<td>protease inhibitor/seed storage/lipid transfer protein (LIP) family protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G55590</td>
<td>96 17</td>
<td>21 11</td>
<td>6 15</td>
<td>pectinesterase family protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G56110</td>
<td>93 20</td>
<td>24 32</td>
<td>20 19</td>
<td>myb family transcription factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G60080</td>
<td>214 9</td>
<td>9 7 20</td>
<td>14 8</td>
<td>protein kinase family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G60090</td>
<td>260 14</td>
<td>7 12 1</td>
<td>18 7</td>
<td>protein kinase family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G61110</td>
<td>224 13</td>
<td>6 13 4</td>
<td>13 12</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G63390</td>
<td>61 4</td>
<td>1 2 2</td>
<td>3 3</td>
<td>expressed protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5G66710</td>
<td>92 39</td>
<td>23 44</td>
<td>40 38</td>
<td>30 protein kinase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Numbers in columns 2 to 8 represent the normalized gene expression levels. An, wild type anthers; In, wild type young inflorescences; S12, wild type stage-12 flowers; Si, wild type siliques; St, wild type stems; Lf, wild type leaves; Rt, wild type roots. Gene descriptions were based on the TAIR database.
Identification of *SDS*-dependent genes and *MMD1*-dependent genes

Using the same strategy as described above, when FDR is 0.05, I identified 240 genes and 39 genes, whose expression levels exhibited more than 2 fold changes in *sds*, *mmd1* anthers compared to wild type anthers, respectively. Among 240 *SDS*-regulating genes, the expression of 231 genes is down regulated in *sds* anthers, indicating that *SDS* possibly positively regulates the transcription of many meiotic genes. And these 231 genes are defined as *SDS*-dependent genes. The expression of all the *MMD1*-regulating genes is down regulated in *mmd1* anthers and these genes are defined as *MMD1*-dependent genes. It is worth noting that most of the *MMD1*-dependent genes (33 of 39) are included in the list of *SDS*-dependent genes, indicating *MMD1* is possible a downstream component in the *SDS*-dependent pathway.

Expression patterns of *SDS*-dependent genes

To gain further insights into the expression patterns of genes regulated by *SDS*, all microarray data were normalized together as described above, including *sds* anthers, *mmd1* anthers, and wild type anthers, stage-12 flowers (S12), siliques (Si), stems (St), leaves (Lf), and roots (Rt). The expression patterns of 231 *SDS*-dependent genes were obtained. The Log2 (ratio of intensity in tissue A to that of wild type anthers) value were calculated and used to identify the co-expressed gene clusters by hierarchical complete linkage clustering model with Pearson squared distance. For 231 down-regulated genes, seven clusters were identified and were named as S1 to S7 (Figure 2.2). The S6 cluster includes 99 genes (42.9%); almost all of them are expressed higher in wild type anthers than any other tissues. All the known function genes regulated by *SDS* are in this cluster, indicating that these 99 genes may specify the features of male meiosis in Arabidopsis.
Figure 2.2. Expression patterns of 231 SDS-dependent genes. The Log₂ (ratio of a gene’s intensity in tissue A to wild type anthers) value was used to perform the gene cluster analysis with hierarchical complete linkage clustering model and Pearson squared distance by the Genesis software. The SDS-dependent genes were classed into 7 clusters, named as S1 to S7. The genes more highly expressed in the wild type anthers are highlighted in green. The genes with relatively low expression levels in the wild type anthers are highlighted in red. An, wild type anthers at stages 4 to 6; In, wild type young inflorescences at stages 1 to 9; S12, wild type stage-12 flowers; Si, wild type siliques; St, wild type stems; Lf, wild type leaves; Rt, wild type roots; sds, sds anthers. In/An, S12/An, Si/An, St/An, Lf/An, Rt/An, sds/An refer to the corresponding Log₂ values.
Assessing the fidelity of microarray data by RT-PCR

To assess the fidelity of the microarray data, we chose twelve genes and performed RT-PCR experiments to check whether the RT-PCR results are consistent with the microarray results (Figure 2.3). Among these genes, five show constitutive expression patterns, four are anther preferential genes, and the remaining three genes are inflorescence preferential genes. As shown in Figure 2.3, the RT-PCR results are generally consistent with the microarray data, indicating that the microarray data are reliable.
Figure 2.3. Assessing the fidelity of microarray data by RT-PCR. Numbers under the RT-PCR bands represent the gene expression values from the normalized microarray data. An, wild type anthers at stages 4 to 6; In, young inflorescences at floral stages 1 to 9.
Figure 2.4. Functional categorization of 707 A/I differential genes

- Putative meiosis-related genes (22 genes, 3.1%)
- Cell cycle regulators (12 genes, 1.7%)
- DNA-binding proteins, transcription factors, and zinc finger proteins (82 genes, 11.6%)
- Cellular Signaling (73 genes 10.3%)
- Protein synthesis, transport and degradation (57 genes, 8.1%)
- RNA-binding, splicing and modification (9 genes, 1.3%)
- Cellular structure genes (44 genes, 6.2%)
- Genes related with environment, stress and disease (49 genes, 6.9%)
- Metabolism genes (204 genes, 28.9%)
- Function unknown or hypothetical genes (155 genes, 21.9%)
Functional categorization of A/I differential genes

To further understand these A/I differential genes, we performed GO functional category analysis through the TAIR website. We also predicted their putative protein domains through the PFAM website to gain more information. Accordingly, these A/I differential genes were classified into 10 functional categories, as described as below (Figure 2.4).

1. Putative meiosis-related genes (22 genes, 3.1%)

This category includes genes that potentially participate in DNA replication, DNA repair and homologous chromosome recombination. Putative chromosome/chromatin structure related proteins and synaptonemal complex (SC) related proteins were also classified into this category. Some known meiotic genes are in this category, such as \textit{AtSPO11-1} (At3g13170), \textit{AtDMC1} (At3g22880), \textit{MS5} (At4g20900) and \textit{ASY1} (At1g67370) (Klimyuk and Jones, 1997; Glover et al., 1998; Bai et al., 1999; Grelon et al., 2001; Armstrong et al., 2002; Cai et al., 2003; Yang et al., 2003). Previous studies have found that the AtSPO11-1 related proteins induce DNA double strand breaks (DSBs), and DSBs initiate DNA damage repair processes and homologous chromosome recombination simultaneously (Grelon et al., 2001). Among the A/I differential genes, we also found a DNA mismatch repair gene (At4g17380, \textit{MSH4}), a DNA helicase gene (At3g27730), and a bifunctional nuclease gene (At1g68290) (Higgins et al., 2004). The covalent bond modification of histones is required for the activities of the chromatin and is important for normal meiosis (Prymakowska-Bosak et al., 1999). SET domain proteins have been reported to play important roles in histone methylation (Springer et al., 2003). Three genes encoding SET domain proteins (At2g17900, At4g27910 and At2g35160) were found in the list of A/I differential genes. One gene encoding a histone protein (At1g54240) and one histone deacetylase gene (At5g35600) are also highly expressed in wild type anthers, indicating that the components
and conformational changes of histones might be crucial to control chromatin activity during male meiosis in Arabidopsis (Pandey et al., 2002).

Among the SDS-dependent genes, we found a histone-like protein (At2g18050), and a histone deacetylase (At5g35600), indicating that they might play some roles in the SDS-dependant bivalent formation and meiotic recombination in Arabidopsis (Azumi et al., 2002). In the list there are also two DNA helicases (At3g49830, At3g58510), a gene encoding the putative chromosome scaffold protein (At5g61260) and a gene encoding the mismatch repair protein (At4g17380, MSH4) (Higgins et al., 2004). All these genes may participate in DNA replication, recombination and repair. Two genes encode the MS5 family proteins, which share some homology to rat synaptonemal complex (SC) subunit protein SCP1 (Paredes et al., 2005).

Indeed, by exploring genes in this category, we have successfully uncovered three meiotic genes, PARTING DANCERS (PTD, At1g12790, 3.2 fold), RCK1/MER3 (At3g27730, 2.6 fold), and MSH4 (At4g17380, 2.4 fold), which are all essential for normal male meiosis (Higgins et al., 2004; Chen et al., 2005; Wijeratne et al., 2006). Future analysis of other genes in this category may reveal additional gene functions in regulating the process of male meiosis in Arabidopsis.

2. Cell cycle regulators (12 genes, 1.7%)

Among A/I differential genes, it is remarkable that we found some putative cell cycle regulating factors, including several homologs of animal tumor-related and apoptosis-related genes. Cell cycle regulation has been extensively studied in yeast and animals. In plants, the studies of cell cycle regulation are mainly focused on mitosis. Currently, SDS is one of few cyclin genes found to be essential for meiosis in plants (Azumi et al., 2002; Wang et al., 2004a;
Wang et al., 2004b). According to the microarray data, the expression level of SDS in the wild type anther is 2.9 fold of that in the young inflorescence. The earliest defect in sds can be detected in pachytene stage. Some sds PMCs gradually die during meiosis, and the other PMCs can finish meiosis abnormally. To date, there are no definite checkpoints reported in plants as have been reported in yeast and animals (Caryl et al., 2003).

The main function of cyclins is to bind and activate cyclin-dependent kinases, which regulate cell cycle progression (Morgan, 1997). Studies in the animal system have found that cell cycle regulators are critical for tumor development and apoptosis (Funk and Kind, 1997; Huang et al., 2004; Zeng et al., 2004). We classified putative cyclins, putative cyclin-dependent kinases and their inhibitors, together with the tumor related and apoptosis related homologs into this category. Further functional exploration of genes in this category could provide new insights into the SDS-dependent pathway in regulating cell growth, cell division and cell death in early anther development and/or male meiosis.

3. DNA-binding proteins, transcription factors, and zinc finger proteins (82 genes, 11.6%)

This category likely plays a central role in regulating early anther development or meiosis, because transcription factors have been shown to be important for several other developmental processes. Previous studies have reported that SPL/NZZ (At4g27330), encoding a transcription factor, is an essential master gene in specifying anther cell type formation during early anther development (Schiefthaler et al., 1999; Yang et al., 1999b). The spl/nzz anther lacks both the tapetum and PMCs. As in other structures such as the flower, embryo, shoot apical meristem and root meristem, gene regulation at the transcriptional level in anthers would be essential in regulating anther development or male meiosis (Irish and Sussex, 1990; Yanofsky et al., 1990; Bowman and Meyerowitz, 1991; Coen and Meyerowitz, 1991; Weigel et al., 1992;
Weigel and Meyerowitz, 1994; Long et al., 1996; Lenhard et al., 2002; Birnbaum et al., 2003). According to the putative function domains, we further classified this category into six subcategories as discussed below.

1) **MYB family.** MYB genes encode the largest family of plant transcription factors (Chen et al., 2006). This subcategory includes 16 MYB genes. In yeast and mice, some MYB genes are involved in processes of meiosis and spermatogenesis (Toscani et al., 1997; Nimmo et al., 1998). However, few MYB genes have been reported to be involved in male meiosis in plants. Among these anther preferential MYB genes, *MYB33* and *MYB65* have been found to be essential for normal tapetum development (Millar and Gubler, 2005). The functions of all the others are still unknown.

2) **MADS box genes.** The ABC model of flowering development is one of the most exciting findings of the functions of MADS box genes in plants. Almost all ABC function genes belong to the MADS box family (Coen and Meyerowitz, 1991; Riechmann and Meyerowitz, 1997; Ma and dePamphilis, 2000; Ng and Yanofsky, 2001; Parenicova et al., 2003). However, the floral organ identity genes such as *AG* (At4g18960) and *PI* (At5g20240) are not in the list of A/I differential genes, because the expression changes are below the cut off, 1.92 fold for *AG* and 1.8 fold for *PI*. Recently, studies have found that *AG* can directly regulate *SPL* activity, which is essential for early anther development (Ito et al., 2004). Three MADS box genes, *AGL18*, *AGL40*, and *FLF* were found to be A/I differential genes. Previous studies have found that *AGL18* and *FLF* are involved in flowering time control (Sheldon et al., 1999; Adamczyk et al., 2005). The function of *AGL40* is still not clear. Therefore, these three anther preferential MADS box genes may also play some roles in early anther development or male meiosis (Coen and Meyerowitz, 1991; Weigel and Meyerowitz, 1994; Ma and dePamphilis, 2000).
3) **bHLH family.** About 150 bHLH proteins have been identified in Arabidopsis (Buck and Atchley, 2003; Heim et al., 2003). However, the functions of most bHLH genes are still unknown. Three genes have been identified in the list, including At2g31210, At1g06170 and *AMS* (At2g16910). *AMS* is required for normal tapetum function after meiosis (Sorensen et al., 2003). The first two genes and another gene At2g31220, the probe for which is not present on the chip, are recently duplicated genes. Our microarray data suggested they might redundantly play some important roles in early anther development.

4) **Genes encoding homeodomain proteins.** Previous studies have shown that homeodomain proteins play crucial roles in animal and plant development, such as SHOOTMERISTEMLESS (STM), KNAT1, WUSCHEL (WUS) (Chuck et al., 1996; Laux et al., 1996; Long et al., 1996). A homeodomain gene, *WUSCHEL* (At2g17950, 2.2 fold), and a homeobox-leucine zipper gene (*HAP9*, 3.2 fold) are in the list (Schena and Davis, 1994; Laux et al., 1996). WUS is involved in shoot apical meristem development and the function of HAP9 is still unknown. The anther preferential expression suggested that *HAP9* might play a role in anther development.

5) **Zinc-finger proteins.** This subcategory includes several kinds of zinc-finger containing proteins, such as WRKY superfamily protein (Cai et al., 2003; Dong et al., 2003), bZIP domain protein (Jakoby et al., 2002), TAZ zinc finger protein, RING-H2 protein and PHD proteins, which may participate in DNA/protein interaction (Aasland et al., 1995; Tague and Goodman, 1995; Jakoby et al., 2002; Dong et al., 2003). In Arabidopsis, the functions of most zinc-finger proteins in this subcategory are still not clear. Among these zinc finger proteins, some of them are putative transcription factors, which may directly regulate the expression of genes involved in anther development and/or meiosis. *MMD1* (7.5 fold) is the only gene with
known function in this subcategory. MMD1 is a PHD domain containing protein and has been reported to play essential roles in male meiosis progression and chromatin modeling (Yang et al., 2003). Among this subcategory there are two other genes encoding PHD proteins, At2g01810 (6.9 fold) and At5g58610 (3.5 fold), and it is necessary to investigate their functions during the early anther development and/or male meiosis.

6) Others. There are also a number of other putative transcription factors, including three AP2 domain proteins, two B3 domain proteins, two NAM proteins and some others. One member of AP2 family, APETALA2 (AP2) is an A function gene controlling floral organ identity (Drews et al., 1991; Mizukami and Ma, 1992). It would be interesting to investigate the functions of these three AP2 genes in anther development and/or meiosis. B3 family proteins are only found in plants to date, and some of them are preferentially expressed in reproductive meristems (Stone et al., 2001; Franco-Zorrilla et al., 2002). The functions of the two anther preferential B3 genes are still not reported. Their expression patterns suggest that they may play some roles in anther development and/or male meiosis.

4. Cellular Signaling (73 genes, 10.3%)

Previous studies have shown that three kinases, EMS1/EXS (At5g07280, 1.75 fold), SERK1, and SERK2, are required for tapetum identity, indicating that the cell-cell signaling pathway plays essential roles during early anther development (Hecht et al., 2001; Canales et al., 2002; Zhao et al., 2002; Albrecht et al., 2005; Colcombet et al., 2005). The finding of 17 kinases in this category provides new opportunities to uncover some new genes involved in anther development and/or male meiosis through cell-cell signaling pathways.
5. Protein synthesis, transport and degradation (57 genes, 8.1%)

There have been some studies on SCF complexes, the E3 ubiquitin ligases, which have been reported to play multiple and essential roles in development (Kuroda et al., 2002; Zheng et al., 2002; Zhao et al., 2003; Yasuhara et al., 2004; Hershko, 2005; Yang et al., 2006). The fact that many putative F-box genes, several SKP1 like genes, and some cullin family genes are A/I differential genes suggests that SCF complexes may play important roles in anther development and male meiosis through the protein degradation pathway. Previous studies have shown that Arabidopsis SKP1 like 1 (ASK1) plays essential roles during male meiosis in Arabidopsis (Yang et al., 1999a). However, no F-box protein has been reported to be involved in plant meiosis. The identification of several anther preferential F-box genes provides opportunities to uncover their possible roles in male meiosis.

6. RNA-binding, splicing and modification (9 genes, 1.3%)

Recently, researchers began to realize that the splicing and modification of RNA play essential roles in animal and higher plant development. In addition, RNA interference is believed to be a universal phenomenon in eukaryotes. Great progresses have been made towards understanding the functions of small RNAs and micro-RNAs in animals and plants (Llave et al., 2002; Rhoades et al., 2002). The fact that some putative RNA-binding proteins are preferentially expressed in the wild type anther implies that regulation by RNA splicing and modification may also exist in anther development. For example, the members of ARGONAUTE family are involved in diverse post-transcriptional RNA-mediated gene silencing systems. They play important roles in gene silencing in Drosophila and fission yeast and are also involved in the programmed DNA elimination in Tetrahymena (Zilberman et al., 2003). Two ARGONAUTE like genes, At5g21150 (AGO4) and At2g27880, are preferentially expressed in wild type anthers,
suggesting the post-transcriptional gene-silencing pathway may also be important for early anther development and/or meiosis.

7. **Cellular structure genes (44 genes, 6.2%)**

   Among this category, there are proteins related with cellular structures, including plasma membranes, cytoskeletons, cellular vesicle transporters and some other organelle-related proteins.

8. **Genes responding to environment, stress and disease (49 genes, 6.9%)**

   The adaptability of plants to environment is mainly determined by stress and disease resistance properties. Most genes in this category may be involved in multiple biological processes of anther development and/or meiosis. P450 genes encode a very large cytochrome protein family (Lewis et al., 1998; Danielson, 2002). They play important roles in hypersensitivity to environment (Godiard et al., 1998), leaf development (Kim et al., 1998), gibberellin biosynthesis (Helliwell et al., 2002), and auxin homeostasis (Barlier et al., 2000). Sixteen putative P450 genes are preferentially expressed in the wild type anther, indicating that anther development might be affected by diverse external and internal factors, and these genes may play important roles in maintaining anther development homeostasis.

9. **Metabolism genes (204, 28.9%)**

   This category is a large portion of A/I differential genes. Most members are enzymes, which participate in certain metabolic processes. Although we have classified some phytohormone synthesis and metabolism genes into this category, some of them might also take part in the signaling of anther development. In Arabidopsis several metabolism related genes have been found to be essential for pollen development. One example is MS2 (At3g11980) (Aarts et al., 1997).
10. Genes with unknown functions or hypothetical genes (155, 21.9%)

This category also represents a large portion of A/I differential genes, and their functions are still unknown. Exploring their function may provide the next frontier of research in anther development and meiosis.

It should be noted that the boundaries of these categories are not absolutely definitive due to several reasons. First of all, category classification is mostly dependent on the predicted gene function, but the functions of most genes are not known yet. Secondly, some genes may have multiple functions. For example, some DNA-binding genes may possibly participate in DNA/protein interaction during meiosis, and they could also be involved in environment signaling pathways. Here, we used information on GO annotation and predicted function domains to propose and emphasize the main functions of these genes and classified them into 10 non-overlapped categories.

2.5 Conclusion

Microarray is a very powerful tool for rapidly obtaining global gene expression profiles. We used it to identify anther preferential genes by comparing the microarray data generated for wild type anthers and young inflorescences. Previously, our lab had obtained and published the global gene expression profiles for six major tissues, but not for the anther. Previous study identified hundreds of genes preferentially expressed in the young inflorescence (floral stages 1 to 9) (Zhang et al., 2005). However, many genes involved in early anther development and/or male meiosis were not uncovered, since the expression intensities of anther preferential genes would be greatly diluted if young inflorescences were used for RNA extraction. To gain more information on anther expressed genes, microarray experiments were performed using the total
RNA extracted from anthers. Although Arabidopsis anthers are very tiny and the length of
anthers at stages 4 to 6 is less than 300 um, two other labmates and I collected enough anther
materials from wild type and several mutants defective in early anther development or male
meiosis, including \textit{sds} and \textit{mmd1}.

Theoretically, if a gene is preferentially expressed in anthers compared to young
inflorescences, its expression level is normally higher in microarray data for anthers than for
young inflorescences, since anther is only a portion of the young inflorescence. Therefore, we
compared the microarray data for anthers with the data for young inflorescences to identify A/I
differential genes. A total of 707 A/I differential genes were identified and many known genes
essential for early anther development or meiosis were found in the list of A/I differential genes.
The comparison of gene expression profiles of wild type anthers versus \textit{sds} anthers and \textit{mmd1}
anthers also enabled us to identify 231 SDS-dependent genes and 43 MMD1-dependent genes.

All the tissues used for the microarray experiments are in the same ecotype background,
making all the microarray data eligible for comparison after normalization. These global gene
expression profiles provide a great resource to rapidly obtain the expression pattern of any gene
represented on the ATH1 chip, which includes about 80\% Arabidopsis nuclear genes. These
gene expression patterns can guide future experimental design to investigate gene functions. This
database can also be used to identify other tissue specific genes.

The identification of A/I differential genes, especially anther preferential genes, has
greatly narrowed down the candidate genes that might be essential for early anther development
and/or meiosis, from more than 20 thousand genes to hundreds of genes. Future functional
studies of these candidate genes through reverse genetic approaches will uncover many novel
genes that are essential for early anther development and/or meiosis. Initial exploration of these
A/I differential genes has enabled us to identify three genes (*RCK1/MER3, PTD, MSH4*) that are essential for male meiosis (Higgins et al., 2004; Chen et al., 2005; Wijeratne et al., 2006).

Despite the rapid progress in plant meiosis studies, many questions of meiosis are still not clear. For example, how can homologous chromosomes recognize each other efficiently? What are the roles of telomeres and centromeres in meiotic recombination? What is the relationship between the synaptonemal complex and meiotic recombination? Are there any checkpoint in plant meiosis? The identification of anther preferential genes, *SDS* dependent genes and *MMD1* dependent genes, as well as the future analyses of their functions will provide great opportunities to characterize *SDS* and *MMD1* functional pathways and uncover many novel genes essential for male meiosis and early anther development.
Acknowledgements

I collaborated with Dr. Wei Zhang on this study. We worked together on anther collection, microarray experiments, and data analysis. We thank Dr. Asela Wijeratne for his help in anther collection and Dr. Xiaohong Zhang for offering her microarray data for wild type tissues other than anthers. We also thank Dr. Wenlei Liu for the SAM analysis and Dr. Naomi Altman for the microarray data normalization.
References


Chapter 3

Genome-wide Analysis of Kelch Repeat-containing F-box Protein Family

Yujin Sun, Xiaofan Zhou and Hong Ma

As the first author, I did most of the work as described in the paper. Xiaofan Zhou helped with gene annotation for several species. Dr. Hong Ma supervised all the work.
3.1 Abstract

The ubiquitin-dependent protein degradation pathway plays diverse roles in eukaryotes. Previous studies indicate that both F-box and Kelch motifs are common in a variety of organisms. F-box proteins are subunits of E3 ubiquitin ligase complexes called SCFs (SKP1, Cullin1, F-box protein, and Rbx1); they have an N-terminal F-box motif that binds to SKP1 (S-phase kinase associated protein), and often have C-terminal protein–protein interaction domain(s), which specifies the protein substrate(s) for degradation via the ubiquitin pathway. One of the most frequently found protein interaction domains in F-box proteins is the Kelch domain. Although both the F-box and Kelch repeats are ancient motifs, Kelch repeat-containing F-box proteins (KFB) have only been reported for human and Arabidopsis previously. The recent sequencing of the rice genome and other plant genomes provides an opportunity to examine the possible evolutionary history of KFB. I carried out extensive BLAST (Basic Local Alignment Search Tool) searches to identify putative KFBs in selected organisms, and analyzed their relationships phylogenetically. I also carried out the analyses of both gene duplication and gene expression of the KFBs in rice and Arabidopsis. This study indicates that the origin of KFBs occurred before the divergence of animals and plants, and plant KFBs underwent rapid gene duplications.

Key words: Arabidopsis thaliana; F-box motif; Kelch repeats; Kelch repeat-containing F-box proteins; SCF; SKP1 like.
3.2 Introduction:

Selective protein degradation through the ubiquitin-dependent pathway plays essential roles in cell cycle progression, transcriptional regulation, and signal transduction (Hershko and Ciechanover 1998). The ubiquitin-activating enzyme (E1) and the ubiquitin-conjugating enzyme (E2) function with the ubiquitin ligase (E3), which specifies the protein target(s), to facilitate the degradation of the ubiquitinated substrates by the 26S proteasome (Koepp et al. 2001; Pickart 2001). Previous studies in human and yeast have uncovered a class of cullin based ubiquitin ligases (E3) (Lyapina et al. 1998; Kobayashi et al. 2004; Willems et al. 2004; Hong et al. 2005). One of the largest and best-characterized families of cullin-based ubiquitin ligases is the SCF complex, which consists of SKP1 (S-phase kinase associated protein), Cullin1/Cdc53, Rbx1, and an F-box protein. Several studies have found that Cullin1/Cdc53 interacts with SKP1 and Rbx1 through its long N-terminal stalk domain and the C-terminal globular domain, respectively (Krek 1998; Skowyra et al. 1999; Zheng et al. 2002). Rbx1 contains a ring finger domain that interacts with the E2 enzyme, while SKP1 bridges Cullin1 and an F-box protein (Kamura et al. 1999). F-box proteins have a relatively conserved F-box domain near the N-terminus interacting with SKP1 and a less conserved protein–protein interaction domain(s) at the C-terminus specifying the ubiquitylational target(s) (del Pozo and Estelle 2000).

Among the SCF subunits, Cullin1 and Rbx1 are highly conserved in diverse organisms and are present at low copy numbers, whereas the numbers of SKP1 homologs range from one in fungi and vertebrates to more than 20 in plants and invertebrates (Gagne et al. 2002; Risseeuw et al. 2003; Kong et al. 2004). The numbers of putative F-box proteins are even greater, particularly in plants; Arabidopsis and human have approximately 700 and 68 predicted F-box proteins, respectively (Gagne et al. 2002; Kuroda et al. 2002; Jin et al. 2004). Genetic studies have
uncovered the functions of several F-box proteins in Arabidopsis, including TIR1, COI1, SLY1, and EBF1/EBF2 in hormone signaling (Ruegger et al. 1998; Xie et al. 1998; McGinnis et al. 2003; Parry and Estelle 2006), SON1 in defense response (Kim and Delaney 2002), ORE9/MAX2 in controlling shoot branching and leaf senescence (Woo et al. 2001; Stirnberg et al. 2002), EID1 in photomorphogenesis (Dieterle et al. 2001), ZTL, FKF1, and LKP2 in flowering time and the circadian clock (Nelson et al. 2000; Somers et al. 2000; Yasuhara et al. 2004), and UFO in floral organ development (Ingram et al. 1995; Samach et al. 1999; Zhao et al. 2001). In addition, molecular analyses suggest that the same Arabidopsis F-box protein may bind multiple SKP1 homologs, suggesting the combinatorial potential for formation of a very large set of SCF complexes (Takahashi et al. 2004).

The large plant F-box protein family can be divided into subfamilies according to the presence of additional protein–protein interaction domains near the C-terminus. These domains include the WD40 repeat, the Leucine-rich repeat, Tub, Lectin, the Kelch repeat and other motifs. The Kelch motif contains 44–56 amino acid residues and was initially identified in the Drosophila melanogaster KELCH protein that is related to galactose oxidase (Xue and Cooley 1993; Bork and Doolittle 1994). Previous studies have uncovered the consensus of Kelch motif that is characterized by four highly conserved residues: two adjacent glycines (G), and a pair of tyrosine (Y) and tryptophan (W) separated by about six residues (Adams et al. 2000; Prag and Adams 2003). A single Kelch motif forms four beta sheets, and multiple Kelch motifs can associate together, forming a bladed beta-propeller that interacts with other proteins (Ito et al. 1991). For example, the well-studied human Keap1 protein contains seven Kelch motifs and can form an E3 ligase together with Cul3 and Rbx1 to ubiquitinate the Srf2 protein (Li et al. 2004).
Although the Kelch motif is commonly found in many organisms, including viruses, bacteria, fungi, plants and animals, only a few Kelch repeat-containing F-box proteins (KFBs) have been characterized (Xue and Cooley 1993; Bork and Doolittle 1994; Adams et al. 2000). The only well-studied KFBs are the three highly similar Arabidopsis proteins (ZTL, FKF, LKP2), which are involved in the flowering time and circadian control (Nelson et al. 2000; Han et al. 2004; Somers et al. 2004; Yasuhara et al. 2004; Imaizumi et al. 2005). Furthermore, little is known about the evolutionary history of the KFBs. The recent determination of the genomic sequences of several plants has allowed a thorough analysis of KFBs in plants. In this study, I carried out extensive BALST searches for all putative KFBs in several organisms and carried out phylogenetic analyses of both animal and plant KFBs. The gene expression profiles of Arabidopsis KFBs were also provided by microarray data analyses. In addition, the information on the chromosome distribution and possible gene duplication events in both rice (OsKFBs) and Arabidopsis (AtKFBs) was presented. The existence of KFBs in both plants and animals suggests an origin(s) that predates the divergence of animals and plants, although none was detected in fungi and other kingdoms. Comparative analysis of the plant KFBs from angiosperms, a gymnosperm and a moss indicated that the KFBs form a number of subfamilies that are well conserved in plants. Moreover, one subfamily has experienced rapid gene birth primarily through tandem duplication events that occurred before the split of Arabidopsis and Brassica. Most of these recently duplicated genes are expressed at very low levels in seven Arabidopsis organs/structures that we analyzed. My results indicate that the KFB family has expanded in plants, and contains both members that are highly stable and conserved, as well as members that are very dynamic and rapidly evolving.
3.3 Material and Methods

Sequence retrieval and protein domain analysis

The protein sequences of known KFBs from previous studies in Arabidopsis were downloaded from the Arabidopsis database (www.arabidopsis.org) (Andrade et al. 2001). Both genomic sequences and protein sequences of Arabidopsis thaliana (TIGR [The Institute for Genome Research] release version 5.0) and Oryza sativa (TIGR release version 4.0) were downloaded for local searches. The genomic sequences of Brassica rapa (www.arabidopsis.org), Populus trichocarpa (www.jgi.doe.gov, release version), Physcomitrella patens (www.jgi.doe.gov; access kindly granted by R. Quatrano; Quatrano et al. 2007), and the EST sequences of Pinus taeda (TIGR) were also downloaded for local BLAST searches. To search for the plant KFBs, we used the protein sequences of all known KFBs (Andrade et al. 2001) as queries to carry out both the TBLASTN and BLASTP against the Arabidopsis genome with a cut off of E-value at 1e-5. All new sequences were then used as queries to carry out another round of BLAST searches. The process was repeated until no new sequences were obtained. The protein sequences that lack either the F-box domain or the Kelch motif based on the Pfam domain analysis (http://www.sanger.ac.uk/Software/Pfam/search.shtml) were eliminated. By choosing the cut off of E-value at 0.5 for both F-box domain and Kelch motif, we identified 97 KFBs in the Arabidopsis genome, five of which, At3g24610, At4g34170, At4g39560, At2g29860 and At2g20380, were modified from the prediction according to the multiple sequence alignment (see below). To search for the KFBs from several other plant species, the protein sequences of all Arabidopsis 97 KFBs were used as queries to carry out TBLASTN searches against the downloaded plant databases as mentioned above. I also carried out BLAST searches against the
Zea Mays genome on the website www.plantgdb.org. The genomic sequences of the BLAST hits were then retrieved, and protein sequences were predicted based on sequence similarities.

Among the 68 human F-box proteins, only a single KFB called F-box 42 was detected previously (Jin et al. 2004). It was used as a query to carry out BLASTP, TBLASTN, and PSI-BLAST searches of both animal and fungi KFBs in the NCBI database.

All predicted KFB protein sequences collected in this study were examined using the Pfam domain analysis with the default cut off (http://www.sanger.ac.uk/Software/Pfam/). Sequences with only one kind of domain, either F-box domain or Kelch motif(s) were then analyzed individually with a cut off of E-value at 1.0. Finally, all the domain information was collected and the protein sequences with the E-values of F-box domain or the Kelch motifs more than 0.5 were eliminated from the further analysis.

**Multiple sequence alignment**

Multiple sequence alignments of all protein sequences were carried out by using Clustal X 1.83 with BLOSUM 30 as the protein weight matrix, and different values of Gap opening and Gap extension were tried. Finally, the Gap opening value of 4.0 and the default value for Gap extension were chosen since they produced the best alignment results (Jeanmougin et al. 1998). The MUSCLE (version 3.52) software was also used to carry out the multiple sequence alignment to compare with the Clustal results (Edgar 2004). All sequences were then grouped into subgroups based on the preliminary NJ tree generated by MEGA 3.0 (Kumar et al. 1994). The protein sequences of each subgroup were aligned, and realigned between subgroups using the profile alignment in Clustal X. Alignments of all the protein sequences were finally adjusted.
manually using both alignments generated by MUSCLE and the results of Pfam domain analysis as the references. The amino acid sequences and alignment are available upon request.

**Phylogenetic analysis**

Phylogenetic analyses were conducted by using both NJ and maximum likelihood (ML) methods. The NJ trees were generated by MEGA (3.0) with the “parewise deletion” option, “Poisson correction” model, and bootstrap of 1000 replicates (Kumar et al. 1994; Guindon and Gascuel 2003). The ML trees were constructed using PHYML (version 2.4.4) with a bootstrap of 100 replicates, JTT (Jones, Taylor and Thornton) substitution model, and gamma distributed rates (determined by PHYML) (Kumar et al. 1994; Guindon and Gascuel 2003). ML tree files were then viewed and modified in MEGA. Only NJ trees were presented in this study, with the bootstrap values from the analyses of both NJ and ML methods. Although plant KFBs had a range of numbers of the Kelch motif and some plant KFBs had only one Kelch motif, if a cut off of E-value as 0.5 was used, in most cases I saw at least another degenerated Kelch motif in the alignment. Finally, I used the sequences of the F-box domain and the first-two Kelch motifs for the phylogenetic analysis. For the subfamily of plant KFBs, additional regions may be used depending on the conservation of the protein sequence in a specific group.

**Chromosome distribution and gene duplication types of AtKFBs and OsKFBs**

To understand the mechanism of the gene duplication events of plant KFBs, I analyzed the chromosome distribution of the KFBs from Arabidopsis and rice and investigated possible duplication types of these genes. Three main types of the gene duplication have been reported previously, including tandem duplication, segment duplication, and gene duplication caused by retrotransposition (Vision et al. 2000; Baumbusch et al. 2001; Cannon et al. 2004). If closely
related genes are arrayed in tandem on the same chromosome, the duplication type is called tandem duplication. Large chromosomal blocks with syntenic distribution of similar genes provide evidence for segment duplication. For the retrotransposition type, duplicated genes (also called retrogenes) normally lack intron, may have the stretches of poly(A) at the 3’ end and short direct repeats at both ends, and are located on different chromosome positions.

**Gene expression analysis of KFBs**

The anthers (at anther stages 4–6) from Arabidopsis Landsberg erecta were collected under a dissection scope, and total RNA was then extracted from two biological anther samples using an RNeasy Plant Kit (Qiagen, Valencia, CA, USA). The kit was then used to carry out the microarray experiment as described previously (Zhang et al. 2005). The public microarray data of other six tissues, including roots, stems, leaves, young inflorescences (stages 1–9), stage-12 flowers, and siliques, were provided by Zhang et al. (2006) in our lab. All the microarray data were analyzed and normalized to make the data comparable as described previously (Zhang et al. 2005). The Pearson’s correlation coefficients for the two biological replicates of each of the seven tissues were all greater than 95%, indicating a very small variation between two biological replicates. For simplicity, the average signal intensity values were used and presented for the gene expression here. The signal intensity value of 50 was used as a conservative cut off for reliable detection of gene expression as discussed in Zhang et al. (2006). To search for OsKFB ESTs, the genomic sequences of rice KFB were downloaded from TIGR and used as query sequences to search for the highly similar ESTs (at least 95% identity) in the TIGR database.
3.4 Results

Plant genomes encode a large number of KFBs with different numbers of Kelch motif

We used the protein sequences of known Arabidopsis KFBs as queries to carry out BLAST (Basic Local Alignment Search Tool) searches in the Arabidopsis genome. Ninety-seven KFBs were detected in the Arabidopsis genome, with the Pfam E-value cut off at 0.5 for both the F-box domain and the Kelch motif. We then used these Arabidopsis KFBs as queries to carry out BLAST searches against sequences of other plant genomes, including *Brassica rapa*, *Populus trichocarpa*, maize, rice, *Physcomitrella patens*, and pine ESTs (Expressed Sequence Tags). We also searched against the budding yeast genome, but no KFB was detected. The single human KFB called F-box 42 was also used as query to search for animal, fungi, protist and prokaryote KFBs through the NCBI (National Center for Biotechnology Information) website. In summary, no KFB was found in single-cell organisms, and only a single copy of KFB was detected in human, zebrafish (Identity to human = 471/683 [68%], Similarity = 509/683 [74%]), *Drosophila melanogaster* (Identity = 215/706 [30%], Similarity = 321/706 [45%]), and other insects. All animal KFBs are close homologs of the human F-box 42, and each contains three Kelch motifs. In contrast, a large number of KFBs were identified in plant genomes. For example, at least 43 were detected in *Brassica rapa* (36 partial sequences with more than 60% identity to Arabidopsis homologs are not included in this study), 41 in *Populus*, 28 in rice, 34 in maize, 10 in pine and 20 in *Physcomitrella*. Furthermore, plant KFBs contain different numbers of Kelch motif, from one to five. Among the 273 plant KFBs included in this study, 37 KFBs contain a single Kelch motif, 161 have two Kelch motifs, 53 have three Kelch motifs, 10 have four Kelch motifs, and the remaining 12 have five Kelch motifs.
Plant KFB family has expanded dramatically via multiple duplication events

To investigate the evolutionary relationships of KFBs, multiple protein sequence alignment of KFBs was carried out as described in the Materials and Methods. Both neighbor joint (NJ) and maximum likelihood (ML) methods generated trees with similar topology (Figure 3.1; a larger phylogenetic tree with 284 sequences is available upon request). Interestingly, all plant KFBs form a separate clade from animal KFBs, with 100% bootstrap support, suggesting that both plant and animal KFBs could be derived from as few as a single gene in the common ancestor of animals and plants. Furthermore, the well-studied Arabidopsis ZTL subfamily (G6, see below) occupies the basal position within the plant KFB lineage, suggesting that the ZTL members might resemble the ancestral KFB in plants, consistent with the fact that, among plant KFBs, the ZTL proteins are most closely related to the human F-box 42 and its vertebrate orthologs (25% identity and 40% similarity to the human KFB).

While the KFBs in animals remain single copy, plant KFBs have increased dramatically in number and could be grouped into 18 highly supported subfamilies, named G1 to G18, for small to moderately-sized clades that have good bootstrap support (at least 65/82 for NJ/ML) from the phylogenetic analysis (Figure 3.1). The subfamilies are further supported by the presence of additional conserved motifs that are shared by members of a subfamily. One large clade with 85/95% bootstrap values was not considered as a single subfamily because it was too large and complex. Eleven subfamilies were found to have members from at least one angiosperm species analyzed here, and from pine and/or Physcomitrella, six subfamilies were only detected in the angiosperm, one subfamily was only detected in Physcomitrella, indicating that the majority of the subfamilies were generated by duplications that occurred before the split between gymnosperms and angiosperms. Because I have examined only a few species and the complete
The phylogeny of the KFBs also provides evidence for more recent duplications within the specific lineages of flowering plants. Among well-supported clades with both rice and maize sequences, 17 have one from each species, five have one rice KFB and two or three maize KFBs, one has one maize KFB and three rice KFBs, and four have only rice or maize genes. Maize is a recent tetraploid and has a much larger genome than rice; it is possible that additional maize KFBs will be identified as more maize genome sequences become available. In clades with Arabidopsis and poplar members, nine have one Arabidopsis KFB and two close poplar paralogs, five have one from each species, three have a pair of paralogs from each species, and one has one Arabidopsis gene and three close paralogs from poplar. The frequent detection of two poplar paralogs corresponding to one Arabidopsis gene is consistent with the fact that poplar is a recent tetraploid.
Figure 3.1. Phylogenetic tree of 113 representative Kelch repeat-containing F-box proteins (KFBs). The KFBs were selected based on the phylogenetic analysis of all 284 KFBs identified in this study (available upon request). The tree was constructed by the neighbor-joining method with Poisson correction, pairwise deletion and bootstrap of 1000 replicates. The bootstrap values of both neighbor-joining (NJ) tree (first number; 1000 replicates) and maximum likelihood (ML) tree (second number; 100 replicates) higher than 50 are shown for each clade. I divided the plant KFBs into 18 subfamilies named as G1 to G18. Animal KFBs form a single clade. The KFB name in the tree combines subfamily, species name, and Kelch motif information. For example, G2 ZmKFB05 2 means “Zea Mays KFB05 with two Kelch motifs, belonging to the G2 subfamily”. Ag, Anopheles gambiae; Am, Apis mellifera; At, Arabidopsis thaliana; Dp, Drosophila pseudoobscura; Dr, Danio rerio; Hs, Homo sapiens; Os, Oryza sativa; Pl, Pinus taeda; Pp, Physcomitrella patens; Pt, Populus trichocarpa; Xl, Xenopus laevis; Zm, Zea mays.
G5 KFBs underwent multiple recent gene duplications in Arabidopsis and Brassica

The initial analysis (not shown) indicated that the G5 subfamily of KFBs has a large number of members in Arabidopsis, but only a single copy in rice, poplar, and pine (Figure 3.1), suggesting that there have been recent gene duplication events in the lineage leading to Arabidopsis since the divergence from poplar. To further study the evolutionary history of the G5 subfamily of KFBs, I carried out analysis with the addition of KFBs from Brassica rapa, which is in the same family (Brassicaceae) as Arabidopsis and also has the most abundant genomic sequence information available among the Brassica species. As shown in Figure 3.2, multiple KFBs closely related to Arabidopsis G5 members were identified in Brassica rapa, suggesting that many of the duplication events predated the split of Brassica and Arabidopsis. Members of the G5 subfamily form nine highly supported clades, named as G5-0 to G5-8. Using the G11 KFBs (closest to the G5 subfamily) as the outgroup, G5-0 is the basal clade and contains the single copy sequences from pine, rice, and maize, suggesting that they present the ancestral state of the G5 subfamily in the early seed plants. G5-1 includes AtKFB25, BrKFB42 and the single-copy PtKFB20 (poplar) forming a sister clade to all remaining G5 members; it is possible that G5-1 is the closest to the eudicot origin of the G5 KFB. Each of the remaining seven clades contains sequences from both Arabidopsis and Brassica, suggesting that they originated between the time of separation from poplar and the time of split of Arabidopsis and Brassica due to several rounds of duplications. Furthermore, within these clades, small clades of only BrKFBs or only AtKFBs provide evidence for gene duplication events after the divergence of the Arabidopsis and Brassica.
**Figure 3.2.** Phylogenetic tree of 102 Kelch repeat-containing F-box proteins (KFBs) in the plant G5 subfamily. The tree was constructed by the neighbor-joining (NJ) method with Poisson correction, pairwise deletion and bootstrap of 1000 replicates. The bootstrap values of both NJ (1000 replicates) and maximum likelihood (ML) trees (100 replicates) higher than 50 are shown for each clade with the first number from the NJ tree and second number from the ML tree. The G11 subfamily was used as the out-group. A large number of G5 KFBs were identified in Arabidopsis and *Brassica*, and only one in rice, maize, *Populus* and pine. The G5 subfamily are further divided into nine clades and named as G5-0 to G5-8. Ag, *Anopheles gambiae*; Am, *Apis mellifera*; At, *Arabidopsis thaliana*; Dp, *Drosophila pseudoobscura*; Dr, *Danio rerio*; Hs, *Homo sapiens*; Os, *Oryza sativa*; Pl, *Pinus taeda*; Pp, *Physcomitrella patens*; Pt, *Populus trichocarpa*; Xl, *Xenopus laevis*; Zm, *Zea mays*. 
Most of G5 AtKFBs are present as tandem repeats in the Arabidopsis genome

To investigate the gene duplication events of plant KFBs, I carried out the analysis of chromosome distributions of plant KFBs in both Arabidopsis and rice. Since G5-KFBs have expanded greatly in Arabidopsis and Brassica, but not in other plants that I analyzed, I investigated them separately (Figure 3.3). As shown in Figure 3.3, the 66 AtKFBs in the G5 subfamily distribute unevenly on the chromosomes with high densities on the lower arm of chromosomes II and IV. Among them, multiple groups of closely related AtKFBs form tandem arrays on the same chromosomes, strongly suggesting that they were generated by tandem duplications. Specifically, AtKFB29 to AtKFB36 form a clade of tandemly arrayed genes on chromosome II. Similarly, AtKFB03 and AtKFB04, AtKFB56 and AtKFB57, AtKFB60 and AtKFB61, AtKFB77 to AtKFB79, AtKFB85 and AtKFB86, AtKFB94 and AtKFB95 are members of the same clades, respectively, and are adjacently located on the same chromosomes. On the other hand, although AtKFB71 to AtKFB76 also form a tandem array on chromosome IV, they form a large clade with other AtKFBs, suggesting that other mechanism(s) of gene duplication might also be involved. Similar situations are found with AtKFB26 and AtKFB27, AtKFB80 to AtKFB83, AtKFB87 and AtKFB88, AtKFB94 (AtKFB95) and AtKFB96. In summary, at least 38 of 66 G5 AtKFBs seem to have been generated by tandem duplication events.

The other groups of AtKFBs and OsKFBs also seem to distribute unevenly in both Arabidopsis and rice (Figures 3.4, 3.5), with high density on chromosome I in Arabidopsis and on chromosome II in rice. Only AtKFB91 and AtKFB92 are adjacent and possibly generated by tandem duplication, no tandem duplication is obvious in rice KFBs.
Figure 3.3. Chromosome distribution of 66 G5-Kelch repeat-containing F-box proteins (KFBs) in Arabidopsis. Thirty-eight of 66 G5-KFBs are tandem duplicates. KFBs within the same subgroup were labeled with a number on the right. The tandem duplicated KFBs were marked with line to the right of the gene names.
Figure 3.4. Chromosome distribution of 31 Arabidopsis non-G5 Kelch repeat-containing F-box proteins (KFBs). All these AtKFBs except AtKFB91 and AtKFB92 were likely to have been generated by duplications other than tandem duplication.
**Figure 3.5.** Chromosome distribution of 28 Kelch repeat-containing F-box proteins (KFBs) in rice. There is no evidence for tandem duplication event detected for rice $KFBs$. 
Gene expression profiles of \textit{AtKFBs} and \textit{OsKFBs}

Because gene expression patterns often provide important clues for gene functions, I examined microarray data to learn about expression profiles of \textit{AtKFBs}. Among 97 \textit{AtKFBs}, 67 genes were included in the Affymetrix chip (Figure 3.6); 41 of them belong to the G5 subfamily and 26 to the other subfamilies. As shown in Figure 3.6, 15 of the G5 \textit{AtKFBs} were expressed in one or more organs/structures, whereas the expression intensity of the remaining 26 G5 genes were below 50 in all of the seven organs/tissues, indicating that they are not expressed at reliably detectable levels in these tissues. These genes might be expressed at higher levels in some other tissues or under conditions different from our growth conditions; also some of them could be pseudogenes. It is worth noting that half of these 26 G5 genes with little or no expression are found in tandem arrays, as described above, such as \textit{AtKFB30}, \textit{AtKFB34}, and \textit{AtKFB36}, also \textit{AtKFB80}, \textit{AtKFB81}, \textit{AtKFB82} and \textit{AtKFB83}.

Among the remaining G5 members, seven were expressed ubiquitously. For example, \textit{AtKFB42} and \textit{AtKFB63} are close paralogs with similar gene expression patterns, suggesting that they may share some redundant function. On the other hand, these genes might still have different functions, either because the slight sequence divergence between these two genes might be sufficient to cause functional differences, there might be expression differences that were not detected by the microarray analysis, or these genes may have different expression in other organs or conditions that were not tested. \textit{AtKFB25}, \textit{AtKFB69}, \textit{AtKFB73} and \textit{AtKFB96} are expressed at relatively high levels compared with other \textit{AtKFBs} in the G5 subfamily. They may play important roles in Arabidopsis. Four other genes, \textit{AtKFB04}, \textit{AtKFB29}, \textit{AtKFB32}, and \textit{AtKFB75}, are expressed specifically in the immature anther, suggesting functions in the developing anther. Although \textit{AtKFB29} and \textit{AtKFB32} are highly similar in sequence, \textit{AtKFB29} is expressed at much
higher levels than AtKFB32, suggesting that both genes may be involved in the same pathway or share some redundant function, and AtKFB29 may play a major role, compared with that of AtKFB32. Other possibilities also exist similar to what were suggested above for AtKFB42 and AtKFB63. The AtKFB14 gene is specifically expressed in stage 12 flowers, and AtKFB31 is preferentially expressed in both anther and leaf tissues. In contrast to the G5 AtKFBs, the other KFBs are mostly expressed ubiquitously at high levels, suggesting that they play important general roles. Interestingly, the gene expression pattern of AtKFB17 is also different from the other members of ZTLs, further supporting the idea that its function may be different from its homologs in the ZTL family.

To obtain information about the expression of the rice KFB genes, the publicly available EST data was examined. And 24 of 28 rice KFBs were found to have corresponding EST information (not shown), indicating that they are expressed, but OsKFB08, OsKFB13, OsKFB21, and OsKFB28 did not have an EST. Either these four genes are not expressed, are expressed at low levels, or are expressed under certain conditions that were different from those used to grow the plants for the EST analysis. Of three closely related genes, OsKFB13, OsKFB25 and OsKFB28, only OsKFB25 has EST information, suggesting that it may play an important role in plants grown in common conditions.
Figure 3.6. Expression patterns of Kelch repeat-containing F-box proteins (KFBs) in Arabidopsis. (A) A phylogenetic tree of 97 KFBs in Arabidopsis. The tree was constructed by the neighbor-joining (NJ) method with Poisson correction, pairwise deletion and bootstrap of 1000 replicates. Only bootstrap values higher than 50 are shown. (B) Gene expression profiles of KFBs in Arabidopsis. 67 AtKFBs were analyzed previously by microarray and labeled with the intensity value; the other 30 shown as blank were not included in the Affymetrix microarray chip. The intensity value of 50 is regarded as the cut off for reliable detection of gene expression. The expression data for AtKFB03/04 were from the same probe set, and the same as those for AtKFB85/86 and AtKFB91/92 gene pairs. All tissues were from wild type Arabidopsis Landsberg erecta plants. An, anther; In, young inflorescence; Lf, leaf; Rt, root; S12, flower at flower stage 12; Si, silique; St, stem.
3.5 Discussion

Rapid gene birth evolution of plant KFBs

Protein degradation through the ubiquitin-mediated pathway is a key process in regulating cell cycle progression, transcription and signal transduction in eukaryotic organisms. Previous studies have found that both F-box proteins and Kelch repeat-containing proteins are ancient and widely distributed, and that both can interact with other proteins that participate in protein degradation processes (Xue and Cooley 1993; Bork and Doolittle 1994; Adams et al. 2000; del Pozo and Estelle 2000; Li et al. 2004; Lechner et al. 2006). But the Kelch repeat-containing F-box proteins were only reported in animals and plants (Andrade et al. 2001; Jin et al. 2004). My results suggest that the F-box and the Kelch motifs are present together in the same proteins only in eukaryotes. The fact that KFBs are only detected in multi-cellular organisms suggests that the combination of the F-box and Kelch motifs might have contributed to the evolutionary success of multi-cellular organisms, which probably needed more complicated mechanisms of protein degradation to regulate complex biological processes.

Although only a single copy of KFB is highly conserved in human and other animals, dozens of KFBs were found in plants, suggesting that rapid gene birth events have occurred in plants. Among the 18 subfamilies of plant KFBs, 11 subfamilies were well conserved in both angiosperms and a gymnosperm (pine) or a moss, suggesting that their functions had diversified in early land plants and have been conserved during seed plant evolution. Further analysis of these KFBs found that members of most of the subfamilies were expressed ubiquitously at relatively high levels, supporting the idea that they may play important roles in plants. Since the divergence of gymnosperms and angiosperms, while most subfamilies have been relatively stable
or only expanded slightly, a dramatic example of rapid gene birth is the G5 subfamily, which experienced numerous gene duplication events in the lineage leading to Arabidopsis and *Brassica*. Although the G5 members form a large subfamily, the functions of most members are not clear, since they are expressed at very low levels and might be pseudogenes. Nevertheless, I found that some of the G5 members are expressed more specifically, suggesting that they may have evolved more specialized functions following gene duplications.

**Mechanism for controlling flower timing and circadian oscillator may be conserved in flowering plants**

The well-supported G6 subfamily contains four Arabidopsis members (AtKFB12, AtKFB17, AtKFB22, AtKFB98), with three of them (LKP1/ZTL/ADO1/AtKFB98, LKP2/ADO2/FKL2/AtKFB22, FKF1/ADO3/AtKFB17) having been shown genetically to be important for the timing of normal flowering and the circadian clock. LKP1 and LKP2 are recent duplicates in Arabidopsis and share some redundant function (Nelson et al. 2000; Somers et al. 2004; Yasuhara et al. 2004). The phylogenetic results indicate that the duplication of LKP1/LKP2 and FKF1 is likely to have occurred before the divergence of eudicots and monocots. The existence of the LKP1/LKP2 and FKF1 (co-)orthologs in both eudicots and monocots strongly suggests that the function of these genes in controlling the circadian clock and flowering time is highly conserved in angiosperms. Unlike the three well-characterized members, AtKFB12 (At1g51550) lacks the light-absorbing LOV domain and has a distinct gene expression pattern, suggesting that its function might have diverged from the other members in the G6/ZTL subfamily. In addition, the absence of orthologs of AtKFB12 in rice and other angiosperm species suggests further evolutionary and possible functional differences between AtKFB12 and other G6 members.
Contribution of tandem duplication to Arabidopsis KFBs

The results showed that multiple AtKFBs, particularly G5 members, are tandemly located on the same chromosome, suggesting their generation by tandem duplication. In contrast, no tandem arrays of KFBs were found in rice. Furthermore, most of the tandem arrayed G5 members are expressed either at very low levels or below reliable detection levels in the organs/structures tested in this study. In addition, many of them have degenerated Kelch motifs, suggesting that they might be pseudogenes or their functions may be divergent. Indeed, some of the G5 members exhibit preferential expression in some organs, supporting the idea of recently evolved functions for these members. It is possible that some of the novel functions provide selective advantages, allowing the duplicated copies to persist in the genomes of Arabidopsis and Brassica. Although rice and poplar do not have similar rapid gene births to those seen in the G5 subfamily in Brassicaceae, it is possible that other plant genomic efforts may reveal additional expansions of KFBs in the near future.

F-box proteins are known or thought to interact with SKP1 homologs as subunits of the SCF complexes (del Pozo and Estelle, 2000; Zheng et al. 2002; Risseeuw et al. 2003). The evolution of the SKP1 gene family has a similar pattern to that of the KFBs (Kong et al. 2004). In vertebrate animals, there is only one copy of SKP1 in each genome, whereas rice and Arabidopsis each have more than 20 SKP1 homologs, indicating that rapid gene birth events also happened in the SKP1 family in plants. Furthermore, the Arabidopsis SKP1 homologs (ASKs) also form several tandem repeats. It has been shown that different ASKs could interact with different F-box proteins, including KFBs (Yamanaka et al. 2002; Risseeuw et al. 2003). The similarity in the patterns of evolution of SKP1s and KFBs suggests possible co-evolution between these two gene families. Furthermore, the analysis of protein–protein interaction
between the KFBs and SKP1 homologs may provide insights into this possible co-evolution of these key regulators of protein degradation.

In this study, it was shown that the plant KFBs experienced numerous gene duplication events since the divergence of animals and plants, including many that resulted in many subfamilies shared by angiosperms and gymnosperms, and even more in the Brassicaceae, forming the large G5 subfamily. In addition, during the evolution of angiosperms, most of the subfamilies have remained very stable, preserving (co-)orthologous relationships for many genes between eudicots and monocots. It is possible that the first expansion of KFBs had contributed to the evolution and success of land plants, with general conserved functions of most subfamilies in many cells and tissues of the angiosperms. It is also possible that the more recent expansion of G5 in Brassicaceae has created many opportunities for further divergence and specialization of gene functions. Therefore, the KFB family exhibits both rapid expansion and stable maintenance of gene numbers, in different periods of evolution and in different subfamilies. This is a fascinating example of gene family evolution that should continue to yield insights into the evolution of gene families, gene functions, and organisms.
Acknowledgments

We thank Ralph Quatrano for sharing unpublished information on *Physcomitrella patens* genomic sequences, Hongzhi Kong for helpful discussion, and Guanfang Wang for comments on the manuscript.
References


Chapter 4

Regulation of Arabidopsis tapetum development and function by \textit{DYSFUNCTIONAL TAPETUM1 (DYT1)} encoding a putative bHLH transcription factor

Zhang W$^1$, Sun Y$^1$, Timofejeva L, Chen C, Grossniklaus U, Ma H


$^1$: These authors contributed equally to this paper
I collaborated with Dr. Wei Zhang on this paper. I mainly did the phenotypic analysis of the *dyt1* mutant (semi-thin sections, pollen staining), gene expression studies of *DYTI* and some other anther genes in wild type and *dyt1* (RT-PCR, real-time PCR), the phylogenetic analysis of *DYTI* homologs. I also made constructs with the *DYTI* transgene (rescue, over-expression) and did the analysis of transgenic plants and *ems1/dyt1* double mutant plants. Dr. Wei Zhang mainly did the preliminary phenotypic analysis of the *dyt1* mutant plant. He also did the *DYTI* mapping and cloning, DAPI staining of chromosome spreads in meiotic cells. Dr. Zhang and I worked together on the characterization of the *dyt1* mutant insertion and the *DYTI* RNA *in situ* hybridization experiments for wild type, *dyt1, spl, and ems1*. Dr. Ljudmilla Timofejeva helped with the preliminary phenotypic analysis. Dr. Changbin Chen helped to design the mapping primers. Dr. Ueli Grossniklaus provided us *dyt1* seeds. Dr. Hong Ma supervised the entire project and provided the funding.
4.1 Abstract

In flowering plants, male fertility depends on proper cell differentiation in the anther. However, relatively little is known about the genes that regulate anther cell differentiation and function. Here, we report the analysis of a new Arabidopsis male sterile mutant, *dysfunctional tapetum1 (dyt1)*. The *dyt1* mutant exhibits abnormal anther morphology beginning at anther stage 4, with tapetal cells that have excess and/or enlarged vacuoles and lack the densely stained cytoplasm typical of normal tapetal cells. The mutant meiocytes are able to complete meiosis I, but they do not have a thick callose wall; they often fail to complete meiotic cytokinesis and eventually collapse. *DYT1* encodes a putative bHLH transcription factor and is strongly expressed in the tapetum from late anther stage 5 to early stage 6, and at a lower level in meiocytes. In addition, the level of *DYT1* mRNA is reduced in the *sporocyteless/nozzle (spl/nzz)* and excess *microsporocytes1/extra sporogenous cell (ems1/exs)* mutants; together with the mutant phenotypes, this suggests that *DYT1* acts downstream of *SPL/NZZ* and *EMS1/EXS*. RT-PCR results showed that the expression levels of many tapetum-preferential genes are reduced significantly in the *dyt1* mutant, indicating that *DYT1* is important for the expression of tapetum genes. Our results support the hypothesis that *DYT1* is a crucial component of a genetic network that controls anther development and function.

**Key words:** Pollen development, *DYT1*, *SPL/NZZ*, *EMS1/EXS*, Tapetum, bHLH Transcription factor
4.2 Introduction

In flowering plants, pollen grains are formed within the anther portion of the male reproductive organ, the stamen (Ma, 2005). Molecular genetic studies have revealed that the B and C functions of the well-known ABC model together determine the stamen identity, whereas the C function alone specifies the carpel identity (Coen and Meyerowitz, 1991; Ma, 2005). In Arabidopsis, *APETALA3 (AP3)* and *PISTILLATA (PI)* are two essential B function genes, and *AGAMOUS (AG)* is required for C function (Coen and Meyerowitz, 1991; Ma, 2005). The Arabidopsis anther is a bilaterally symmetrical structure with four lobes. Each lobe is comprised of four distinct somatic cell layers from outer to inner they are the epidermis, endothecium, middle layer and tapetum (Goldberg et al., 1993; Sanders et al., 1999). The tapetum surrounds meiocytes and generates many proteins, lipids, polysaccharides and other molecules necessary for pollen development (Goldberg et al., 1993). Accordingly, the tapetum is characterized by active protein synthesis and secretion, a high rate of energy metabolism, and expression of many tapetum-preferential genes (Dickinson and Bell, 1976; Liu and Dickinson, 1989; Hernould et al., 1998; Rubinelli et al., 1998; Taylor et al., 1998; Zheng et al., 2003; Scott et al., 2004).

Several genes have been identified that are required for normal anther development (Ma, 2005). For example, the *SPOROCYTELESS/NOZZLE (SPL/NZZ)* gene is required for cell type specification in both male and female reproductive organs (Schietthaler et al., 1999; Yang et al., 1999). The *spl/nzz* mutant anthers lack the endothecium, middle layer, tapetum, and meiocytes. Recently, Ito et al. (2004) found that *SPL/NZZ* is a direct target of AG, and that ectopic expression of *SPL/NZZ*, independent of AG, induces the formation of anther locule with pollen grains (Ito et al., 2004). *SPL/NZZ* encodes a putative transcription factor, suggesting that
SPL/NZZ regulates anther cell differentiation by activating downstream genes (Schieffthaler et al., 1999; Yang et al., 1999).

Furthermore, *EXCESS MICROSPOROCYTES1/EXTRA SPOROGENOUS CELLS* (*EMS1/EXS*) and *TAPETUM DETERMINANT1* (*TPD1*) are early (pre-meiosis) genes that encode a receptor-like protein kinase and a small protein, respectively, and act in the same genetic pathway to control the tapetal cell identity (Canales et al., 2002; Zhao et al., 2002; Sorensen et al., 2003; Yang et al., 2003). Recently, the *SERK1* and *SERK2* genes encoding closely related receptor-like protein kinases were shown to have redundant functions in controlling tapetum formation (Albrecht et al., 2005; Colcombet et al., 2005). *MALE STERILITY1* (*MS1*) and *ABORTED MICROSPORES* (*AMS*) act after meiosis and encode a PHD domain-containing protein and a bHLH transcription factor, respectively; they are essential for late stage functions of the tapetum (Wilson et al., 2001; Ito and Shinozaki, 2002; Sorensen et al., 2003). Remarkably, the “early” gene *EMS1/EXS* is expressed in both stamen and gynoecium, whereas the “late” genes *MS1* and *AMS* are anther-specific. However, it is not known how the “male-specific” *MS1* and *AMS* expression is regulated.

Here we report the isolation of a new Arabidopsis mutant, which is male-sterile and defective in tapetum differentiation and function. Because the mutant has an abnormal tapetum, we named the gene *DYSFUNCTIONAL TAPETUM1* (*DYT1*). *DYT1* encodes a putative bHLH transcription factor and is preferentially expressed in tapetal cells as early as anther stage 5, spatially similar to, but temporally earlier than, *MS1* and *AMS*. Furthermore, our results suggest that *DYT1* likely acts downstream of *SPL/NZZ* and *EMS1/EXS*, and is required for normal expression of *AMS*, *MS1* and other tapetum-preferential genes. We propose that *DYT1* is a component of a genetic network for tapetum differentiation and function.
4.3 Materials and methods

Plant materials and growth

*Arabidopsis thaliana* plants in this study are in the Landsberg *erecta* (Ler) background for all experiments with the exception of the mapping of the *DYT1* gene, for which we crossed *dyt1* with Columbia. The *spl, ems1,* and *tpd1* mutants were kindly provided by Drs. W. Yang, D. Zhao, and D. Ye, respectively (Yang et al., 1999; Zhao et al., 2002; Yang et al., 2003). The *dyt1* mutant was isolated from the progeny of a *Ds* insertional line (ET4262). The *ems1 dyt1* double mutant was obtained by pollinating the *dyt1* pistil with pollen from an *ems1/+* plant. Plants were grown under long-day conditions (16 hours light/8 hours dark) in a ~22°C growth room.

**Characterization of mutant phenotypes**

Plants were photographed with a Sony digital camera, DSC-F707 (Sony Corp., Tokyo, Japan). Flower pictures were taken using a Nikon dissecting microscope (Nikon Corp., Tokyo, Japan) with an Optronics digital camera (Optronics Inc., Goleta, CA, USA). To determine pollen viability, mature anthers were stained with the Alexander solution (Alexander, 1969) and photographed under an Olympus BX-51 microscope (Olympus Corp., Tokyo, Japan) with a SPOT II RT camera (Diagnostic Instruments, Inc., Sterling Heights, MI, USA). Chromosome spreading and 4′-6-Diamidino-2-phenylindole (DAPI) staining were performed as described (Ross et al., 1996). Wild type and mutant inflorescences were collected and fixed as described (Zhao et al., 2002). Floral buds were embedded in Spurr’s resin; semi-thin (0.5µm) sections were made by using an Ultracut E ultramicrotome (Leica Microsystems, Nussloch, Germany), stained with 0.05% of Toluidine blue O for 40 to 60 seconds, and photographed under the Olympus BX-51 microscope with the SPOT II RT camera.
Mapping and functional complementation

The dyt1 mutation was found to be distinct from the Ds insertional locus (not shown). The dyt1 mutant (in the Ler background) was crossed with Columbia to obtain F1 and F2 seeds. About 500 F2 plants with mutant phenotypes were genotyped by using the SSLP and dCAPS markers (Li et al., 2001). Several predicted genes were found in the mapped region flanked by recombination events; these genes were amplified from wild type and dyt1 plants and sequenced. TAIL PCR was used to determine the sequences of a retrotransposon insertion in the dyt1 mutant (Liu and Whittier, 1995). To verify the DYT1 coding region, we used the primers oMC1872/oMC1873 (see Table1 for all primer sequences), designed according to the annotated At4g21330 locus, to amplify wild type Ler cDNA.

To generate a complementation construct, a 2.7 kb wild type genomic fragment was amplified by PCR using primers oMC2241 and oMC2242, and then cloned into a modified pCAMBIA1300, yielding pMC2969. The Agrobacterium strain GV1301 was transformed with the plasmid pMC2969, then used to transform dyt1/+ plants. The seeds of transformed plants were screened for hygromycin resistant seedlings, which were transferred into the soil. The T1 plants were genotyped to identify homozygous dyt1 plants with the T-DNA, using a primer (oMC1945) for the retrotransposon insertion at the DYT1 locus and DYT1-gene-specific primer oMC1834. The oMC1823/oMC1873 and oMC2194/oMC1834 primers were used to identify the dyt1 allele and the transgene, respectively.
Table 4.1. PCR primers were used in this study.

<table>
<thead>
<tr>
<th>oMC number</th>
<th>sequence (5’ to 3’)</th>
<th>oMC number</th>
<th>sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>271</td>
<td>CGATGAGGAATCTCAAGCGC</td>
<td>2083</td>
<td>CATACTCTGTTCGTTCGTTGC</td>
</tr>
<tr>
<td>272</td>
<td>CGTCTCCTCACATTAAACC</td>
<td>2084</td>
<td>GTACTCTGAAAAGGGCTCACTCC</td>
</tr>
<tr>
<td>490</td>
<td>CGTCTCGTCTTTTGTTTTTTAC</td>
<td>2085</td>
<td>CATTCCCATAGGCTCAAAACC</td>
</tr>
<tr>
<td>499</td>
<td>AACAAACCCGTCATTTA</td>
<td>2098</td>
<td>GTTCAAGGCGACAGTACATG</td>
</tr>
<tr>
<td>500</td>
<td>ACCGGGAGAATGGGTGCAC</td>
<td>2099</td>
<td>GGAATGGTCCGTACGATGAG</td>
</tr>
<tr>
<td>509</td>
<td>GTGATGATCCAGACCATCGTATTGTTTAC</td>
<td>2100</td>
<td>ATGGTACATCTCAAAGTCCCTTT</td>
</tr>
<tr>
<td>510</td>
<td>CGCTCGGATTTAAGGTTATAGGCTCAT</td>
<td>2101</td>
<td>CTAGGCACCTTAGCCACAATC</td>
</tr>
<tr>
<td>1533</td>
<td>GGTAAATGGTCTCACGTTAGG</td>
<td>2102</td>
<td>GCGACTTTGGGAAATCTGAGTG</td>
</tr>
<tr>
<td>1534</td>
<td>AAGGACCTTTAATCTTCATGCTGC</td>
<td>2103</td>
<td>GGAATAGTTGAAGGCGAGTCC</td>
</tr>
<tr>
<td>1823</td>
<td>ACTCTAACTGGAATTCACATTTCAC</td>
<td>2104</td>
<td>GCACCCAGCATTCAATATGAATC</td>
</tr>
<tr>
<td>1834</td>
<td>CAGAGCCATAAAGGCGACATA</td>
<td>2105</td>
<td>ATAGGGTGCAAGGCGAGTAC</td>
</tr>
<tr>
<td>1872</td>
<td>ATGGGTGGAGGAAGCAGATTC</td>
<td>2127</td>
<td>CGGTCACTCGGTAAGCAACCTA</td>
</tr>
<tr>
<td>1873</td>
<td>TTATGGATTGTCTTCATATAACCT</td>
<td>2128</td>
<td>CCTCGCCCTTACGCTAAGAACC</td>
</tr>
<tr>
<td>1945</td>
<td>TCAAGAAGTTCCTCTCCTACATCTC</td>
<td>2129</td>
<td>TTCACAGTATGAGCCGCTCTTTC</td>
</tr>
<tr>
<td>1963</td>
<td>TAGACGCGTTTCCGCCCTTGACG</td>
<td>2130</td>
<td>ATGGGTCTATTTACATCCGCTCC</td>
</tr>
<tr>
<td>1964</td>
<td>ACGCGCAATAATGGTTCCTGA</td>
<td>2194</td>
<td>GCCGGAGCTAAGATTGCTACGATA</td>
</tr>
<tr>
<td>2044</td>
<td>CGTTGAGAGGGAATGGGTGTAGC</td>
<td>2238</td>
<td>TCGAGCTCTCTAGAATGGTGGAGAG</td>
</tr>
<tr>
<td>2045</td>
<td>AAACCTGAACGCTCGTCTAGCGATTT</td>
<td>2239</td>
<td>TCGGGGCGCGGAGATCTTTATGGATTGCTTTCATACCT</td>
</tr>
<tr>
<td>2046</td>
<td>ATTTGTCGCTAGGGCATGTCCT</td>
<td>2241</td>
<td>CTCGCGCTGTCGAAGGATAC</td>
</tr>
<tr>
<td>2047</td>
<td>GTTAAAGGAAGGCTCGTCGCT</td>
<td>2242</td>
<td>CGTGGAGTAGGAAGACTCGAG</td>
</tr>
<tr>
<td>2048</td>
<td>TCGCTCTGTCCTCCAGATAACC</td>
<td>2280</td>
<td>TCATCCCTCAGCTCAGCTCC</td>
</tr>
<tr>
<td>2049</td>
<td>TTCCTGACAGGATGTCCTTAC</td>
<td>2281</td>
<td>GTGGCCAGATGAAAGCTCGCA</td>
</tr>
<tr>
<td>2050</td>
<td>TCGTGCTCCACCTGGAGTTTAC</td>
<td>2362</td>
<td>TCCCAAGGGTGGTTGATTTC</td>
</tr>
<tr>
<td>2051</td>
<td>TCCCTGCTTCTCTCGCTGATAG</td>
<td>2363</td>
<td>TGTCGCCGATAGTTGTTTGTC</td>
</tr>
<tr>
<td>2052</td>
<td>GAGGTTATACAGGATGTCCTGAG</td>
<td>2364</td>
<td>CTCATCTCCTCAAGCGCTCA</td>
</tr>
<tr>
<td>2053</td>
<td>GAGCCTCCTCTTTCCTTATCC</td>
<td>2365</td>
<td>AGACACGTAACCGCAAACCA</td>
</tr>
<tr>
<td>2054</td>
<td>ACTTAAACCGGCTCTACCAGGCTAG</td>
<td>2366</td>
<td>GGTATCGGGCAGAGCTCTCT</td>
</tr>
<tr>
<td>2055</td>
<td>CGATATCCATCCTCCAGGACTC</td>
<td>2367</td>
<td>CCAATTCTCTCGCGCTCTACA</td>
</tr>
<tr>
<td>2056</td>
<td>GGCAGACTCATCGACCTCAACA</td>
<td>2368</td>
<td>TTCCCCATTTCCAGCTTCCA</td>
</tr>
<tr>
<td>2057</td>
<td>GACCCCAAGAGTCAATAACCTAGC</td>
<td>2369</td>
<td>ACCAGGTCTCTGATCTCCAGCG</td>
</tr>
<tr>
<td>2058</td>
<td>ACAGGAGAGATGGTGGCTTCAC</td>
<td>2370</td>
<td>AGGCTGGAAGCTGGTGAG</td>
</tr>
<tr>
<td>2059</td>
<td>TGCTGCTACCTCCAGACTACAT</td>
<td>2371</td>
<td>GGTCTGCAGTTGGCTCTTCA</td>
</tr>
<tr>
<td>2060</td>
<td>AAGGGTTCCGGTTCCTCAATGG</td>
<td>2372</td>
<td>GTCGCCGTTTCCTCTGGAAGTCT</td>
</tr>
<tr>
<td>2061</td>
<td>ATGGTCGGGCAGCTCAGAGTAGT</td>
<td>2373</td>
<td>AAGGGCGGCTACAGCAGGCTT</td>
</tr>
<tr>
<td>2062</td>
<td>ATGGGGTTTACTCCGAAAGATC</td>
<td>2374</td>
<td>AACCGACGAAACACGTCTCT</td>
</tr>
<tr>
<td>2063</td>
<td>TCACACCGGAGTCTGATACTAG</td>
<td>2375</td>
<td>CCCAGTCAAGATTAAGGGCG</td>
</tr>
<tr>
<td>2064</td>
<td>GATATCTCTCTCCTGTCGTCGCT</td>
<td>2376</td>
<td>CTTGCTGCTATCTCGGTGTC</td>
</tr>
<tr>
<td>2065</td>
<td>TGCCAGTATAGGGACACACTAGC</td>
<td>2377</td>
<td>GGCTGCTGCTGAGAGCGTTAC</td>
</tr>
<tr>
<td>2066</td>
<td>AGCTATATAAACGGAGGAGACGAG</td>
<td>2378</td>
<td>TGGTAATAACGACTGGTGAGG</td>
</tr>
<tr>
<td>2067</td>
<td>GAGGAGAGAGATGCTATCACCAGG</td>
<td>2379</td>
<td>TGGGCAAGCCATACGTGC</td>
</tr>
<tr>
<td>2068</td>
<td>TCGTGATCAGGAGCGACTAGTATTG</td>
<td>2380</td>
<td>TGGATGACCTTACTGAGGCG</td>
</tr>
<tr>
<td></td>
<td>Sequence 1</td>
<td>Length</td>
<td>Sequence 2</td>
</tr>
<tr>
<td>----</td>
<td>------------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>2069</td>
<td>AGACTTCTTCTGTGCCCCCTACCAAC</td>
<td>2381</td>
<td>GCGCCAGTTCTAGCATGTT</td>
</tr>
<tr>
<td>2070</td>
<td>CAAGAAAGGTCTCTGGGACCCAG</td>
<td>2382</td>
<td>ATTTGCCTCAAGGATCCAG</td>
</tr>
<tr>
<td>2071</td>
<td>GAAGATGTGTTCCATTGAGCAG</td>
<td>2383</td>
<td>CCTGACTCCCATCGAATGAAG</td>
</tr>
<tr>
<td>2072</td>
<td>CAGGAAGCAATGGCATAAGTTCAC</td>
<td>2384</td>
<td>CAGCACTCGTGTTCCTTCCT</td>
</tr>
<tr>
<td>2073</td>
<td>TCCTTGCTCATGCTACTTGGAAG</td>
<td>2385</td>
<td>ACGCTGAGATTCACACCGCT</td>
</tr>
<tr>
<td>2074</td>
<td>AGTAGTGTTCCAATGGGCCATTCAC</td>
<td>2386</td>
<td>GCGGAATCATCGAGTTGAGCA</td>
</tr>
<tr>
<td>2075</td>
<td>AAGCAATGCTCAGAGGCCATAG</td>
<td>2463</td>
<td>TGCGTAACCTCGTATGTAATCG</td>
</tr>
<tr>
<td>2076</td>
<td>CATAAAGCAGGCAGCTCAAG</td>
<td>2465</td>
<td>GATCACGGTCATCGGTAGCA</td>
</tr>
<tr>
<td>2077</td>
<td>ACCGAATCGAGCATCTGGTCTAG</td>
<td>2466</td>
<td>TGCGCATACGATCGGTAGCA</td>
</tr>
<tr>
<td>2078</td>
<td>TGTGATAAGTGACCGCCCCACTTGTTC</td>
<td>2467</td>
<td>CTGCTGGCAACTTCAATGG</td>
</tr>
<tr>
<td>2079</td>
<td>TAGTGTTGAAACTCGTCAACACGC</td>
<td>2468</td>
<td>TTGAGCACAGACAACCAATGG</td>
</tr>
<tr>
<td>2080</td>
<td>TCAAGCGTCAGCGCACGTC</td>
<td>2469</td>
<td>U1 AccGGcAAGGACACTCCTCT</td>
</tr>
<tr>
<td>2081</td>
<td>TCTCAGCTCTGCTCCCACATG</td>
<td>2467</td>
<td>U2 ACGCCTCAACTTGGTGTGT</td>
</tr>
<tr>
<td>2082</td>
<td>AAAGGCGAGTCAACTCAGGTTAG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The primers whose sequences have been shown in the text are not listed in this table. U1 and U2 are the RT-PCR primers for *UBQ1*. 
RT-PCR and real-time PCR

A set of genes were selected for expression analysis by RT-PCR because they are known to be important for male reproduction: SPL/NZZ (Schiefthaler et al., 1999; Yang et al., 1999), EMS1/EXS (Canales et al., 2002; Zhao et al., 2002), TPD1 (Yang et al., 2003; Yang et al., 2005), AMS (Sorensen et al., 2003), MS1 (Wilson et al., 2001; Ito and Shinozaki, 2002), MALE STERILITY2 (Aarts et al., 1997), MALE STERILITY5 (Glover et al., 1998), AtMYB32 (Preston et al., 2004), AtMYB103 (Higginson et al., 2003), A6 and A9 (Sorensen et al., 2003), AtMYB33 and AtMYB65 (Millar and Gubler, 2005), SOLO DANCERS (SDS) (Azumi et al., 2002) and ROCK-N-ROLLERS/AtMER3 (RCK/AtMER3) (Chen et al., 2005; Mercier et al., 2005). A second group of sixteen genes was identified using microarray data of wild type and ems1 mutant anthers we had obtained in our laboratory (W.Z., Y.S., A. Wijeratne, H.M., unpublished data). The genes that were expressed in the ems1 anther at levels that were one half or less of those in the wild type anthers were regarded as candidate tapetum-preferential genes because the ems1 anthers lack tapetum. In addition, UBQ1 and AtMYB4 were used as constitutive expression controls (Vannini et al., 2004). The primers for RT-PCR and relevant microarray data are provided in the Tables 1 and 2, respectively. The primers for real-time PCR are listed in Tables 1 and 3. The PCR and data treatment were done as described previously (Ni et al., 2004). Plant tissue was collected and quickly frozen in liquid nitrogen. The anthers at approximately anther stages 4 to 7 were collected under a dissection microscope. Total RNA was extracted using the RNeasy Plant Kit (Qiagen, Valencia, CA) from young inflorescences (approximately floral stages 1-10). 1-2 µg total RNA was used for reverse transcription according to the manufacturer’s instruction to synthesize cDNA, which was used directly as PCR templates (Invitrogen, Carlsbad, CA).
RNA in situ hybridization

Non-radioactive RNA in situ hybridization was performed as described (Li et al., 2004). A 624-bp DYT1 cDNA fragment was amplified using DYT1-specific primers with XbaI and BglII sites at the 5’ end: oMC2238 and oMC2239, respectively. The PCR product was confirmed by sequencing and cloned into the pGEM-T vector (Promega, Madison, WI) to yield plasmid pMC2949. Plasmid DNA was completely digested by XbaI or BglII and used as template for transcription with SP6 or T7 RNA polymerases, respectively (Roche, Mannheim, Germany). Images were obtained using the Olympus BX-51 microscope with the SPOT II RT camera and edited using PHOTOSHOP 5.0 (Adobe system INC, San Jose, CA, USA).

Over-expression of DYT1

A DYT1 cDNA fragment was amplifies using gene-specific primers oMC1872 and oMC1873 (Table 4.1), then cloned into pGEM-T vector to yield plasmid pMC2941. After verification by sequencing, the DYT1 cDNA fragment was sub-cloned into pCAMBIA1300 downstream of the CaMV 35S promoter to produce the plasmid pMC2942. An Agrobacterium strain C3581 carrying pMC2942 was used to transform wild type, ems1/+ and spl/+ plants. The transgenic plants were selected by hygromycin resistance and verified using PCR with oMC1872 and oMC1873. EMS1 gene-specific primers oMC499 and oMC500, SPL/NZZ gene-specific primers oMC2044 and oMC2045, plus Ds5-specific primer oMC490 were used to identify ems1 and spl heterozygous and homozygous plants, respectively. Paraffin sections were prepared as described for in situ experiments (above) and photographed as described for semi-thin sections.
Phylogenetic analysis of the *DYT1* subfamily

The protein sequences of the nine genes from group 9 of the Arabidopsis bHLH family, including *DYT1*, were used to search for the closest homologs in both the rice genome (http://tigrblast.tigr.org/euk-blast/index.cgi?project=osa1) and the *Populus* genome (http://www.floralgenome.org/cgi-bin/tribedb/tribe.cgi) using both BLAST and TBLASTN programs with a cutoff of 1E-15 (Heim et al., 2003; Toledo-Ortiz et al., 2003). The multiple sequence alignment of full-length protein sequences was performed using ClustalX (Plate-Forme de Bio-Informatique, Illkirch Cedex, France) with a combination of GOP = 4.0 and GEP = 0.1. The bHLH domain region and additional conserved regions were aligned and used to perform neighbor joining (NJ) analyses with the “pairwise deletion” option, “P-distance” model and 1000 bootstrap replicates test using MEGA version 3.0 (http://www.megasoftware.net/index.html) (Kumar et al., 2004).
4.4 Results

The isolation of a new male-sterile mutant

To identify new Arabidopsis genes important for anther development, we screened for male sterile plants among Ds transposon insertional lines. One line was found to segregate normal and sterile plants with an approximate 3:1 ratio. Pollination of the mutant pistil with wild type pollen indicated that the mutant is female fertile. The mutant was named *dysfunctional tapetum1* (*dyt1*) because of its anther defects (see below). The vegetative growth of the *dyt1* mutant appeared normal (Figure 4.1A and B) and most mutant floral organs were also normal with the exception of shorter stamen filaments and smaller anthers (Figure 4.1C and D). There were no pollen grains on the anther surface of opened flowers (Figure 4.1G). Occasionally mutant siliques contained a few seeds (not shown), possibly due to residual gene function (see below).
Figure 4.1. Phenotypes of the wild type (Ler), dyt1 mutant, and transgenic plants for complementation. (A) A Ler plant. (B) A dyt1 plant, with very small siliques (arrows). (C) A Ler flower. (D) A dyt1 flower. (E) A flower of the dyt1 plant with the DYT1 transgene. (F) A wild type anther, with viable pollen grains (stained). (G) A dyt1 anther, no viable pollen. (H) An anther from a dyt1 plant with the DYT1 transgene, with a large number of viable pollen grains and some microspores (arrowheads). Scale bars: 10 mm in A, B; 500 µm in C-E; 20 µm in F-H.
The *dyt1* mutant is defective in tapetum development

Detailed analyses were performed to understand the mutant developmental defects. Chromosome spread experiments were performed and revealed that normal meiotic features could be observed in the mutant meiocytes, demonstrating that meiotic nuclear divisions can proceed normally (Figure 4.2). We also generated semi-thin anther sections to investigate the mutant anther development (Figure 4.3). From anther stage 1 to 3, the *dyt1* anthers appeared normal (data not shown). At stage 4, the mutant anther was similar to the wild type anther in cell layer and cell number, but had a slightly different shape from the wild type anther and was vacuolated in more cells than normal. In addition, the mutant sporogenous cells appeared more deeply stained than wild type cells (Figure 4.3A,B). At early stage 5, the *dyt1* anther lobe had four cell types interior to the epidermis, similar to the wild type. The wild type anther at late stage 5 (Figure 4.3E) was also vacuolated in more cells than earlier and had deeply stained meiocytes. Therefore, mutant anthers at stage 4 and early stage 5 exhibited these morphological features precociously (Figure 4.3A-D).
Figure 4.2. Meiosis in *Ler* and *dytl* anthers. (A) and (B) Pachytene images of the *Ler* and *dytl* meiocytes, respectively, with condensed chromosomes. (C) and (D) Diakinesis images of the *Ler* and *dytl* meiocytes, each with 5 bivalents of attached homologous chromosomes. (E) and (F) Telophase II images of the *Ler* and *dytl* meiocytes, respectively. Both show 4 decondensed chromosome clusters. Scale bar: 5 µm.
Figure 4.3. Anther development from stage 4 to stage 8 in the wild type (Ler) and dyt1 mutant.

Locules from anther sections:

(A), (C), (E), (G), (I) and (K) are from sections of wild type anthers; (B), (D), (F), (H), (J) and (L) were from sections of dyt1 mutant anthers. (A) and (B) Stage 4 anthers. (C) and (D) Late stage 4 or very early stage 5. Vacuolization in the dyt1 mutant occurred in more cells and the vacuoles were larger than those in the wild type. (E) and (F) Stage 5 anthers, with more and larger vacuoles in the tapetum and middle layer of the dyt1 anther (F) than the wild type (E). (G) and (H) Stage 6 anthers. The vacuolization in cells of the mutant tapetum and middle layer became more extensive. The mutant meiocytes had a much thinner callose layer around them (arrowheads). (I) and (J) Stage 7 anthers. Wild type meiocytes undergo cytokinesis to form tetrads. In dyt1 anther, the tapetum and middle layer cells were swollen and had excess vacuolization, and meiocytes generally collapsed before cytokinesis (arrowheads). (K) and (L) At stage 8, in the wild type anther locules, microspores were released from the tetrad; in the dyt1 anther, almost all meiocytes degenerated. E, epidermis; En, endothecium; ISP, inner secondary parietal cells; SS, secondary sporogenous cells; ML, middle layer; T, tapetum; Ms, meiocytes; Tds, tetrads; Msp, Microspores; D-Ms, degenerated meiocytes. Scale bars: 10µm.
In the wild type anther at late stage 5 (Figure 4.3E), the tapetum had significantly larger cells than earlier. In the mutant (Figure 4.3F), additional vacuoles were observed in tapetal cells, with a reduction of the cytoplasm. The vacuoles in the wild type tapetum at this time are fewer and smaller than those in the mutant. Also, the mutant middle layer maintained its thickness with vacuolation, unlike the reduced thickness of the wild type middle layer (Figure 4.3E and F). At stage 6, a thick callose wall forms around the meiocytes (Figure 4.3G). In contrast, the mutant meiocytes had very thin callose cell walls (Figure 4.3G and H). At this stage, most of the mutant meiocytes were undergoing meiosis (Figure 4.2), but some of them had collapsed. At stage 7 and stage 8, the completion of wild type meiosis results in the formation of tetrads and then microspores, but mutant tapetal and middle layer cells swelled with expanded vacuoles and filled the center of the locules where the meiocytes had collapsed and degraded (Figure 4.3I-L).
Table 4.2. Microarray data of the genes selected in RT-PCR.

<table>
<thead>
<tr>
<th>gene name</th>
<th>locus</th>
<th>oMC</th>
<th>Av-WT</th>
<th>Av-ems1</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPL</td>
<td>At4g27330</td>
<td>2044/2045</td>
<td>402.50</td>
<td>283.52</td>
<td>SPOROCYTELESS/NOZZLE</td>
</tr>
<tr>
<td>EMS1</td>
<td>At5g07280</td>
<td>509/511</td>
<td>1062.95</td>
<td>146.89</td>
<td>EXCESS MICROSPOROCYTES1/EXTRA SPOROCYGENOUS CELLS</td>
</tr>
<tr>
<td>TPD1</td>
<td>At4g24972</td>
<td>2046/2047</td>
<td></td>
<td></td>
<td>TAPETUM DETERMINANT1</td>
</tr>
<tr>
<td>EMS2</td>
<td>At1g11980</td>
<td>2060/2061</td>
<td>690.78</td>
<td>12.25</td>
<td>lip transfer protein</td>
</tr>
<tr>
<td>LPT12</td>
<td>At3g51590</td>
<td>2062/2063</td>
<td>446.85</td>
<td>40.33</td>
<td>LIPID TRANSFER PROTEIN</td>
</tr>
<tr>
<td>ATSTP2</td>
<td>A1g07340</td>
<td>2064/2065</td>
<td>174.90</td>
<td>60.65</td>
<td>hexas transportor</td>
</tr>
<tr>
<td>ATA1</td>
<td>At1g02960</td>
<td>2066/2067</td>
<td>2090.38</td>
<td>29.80</td>
<td>alcohol dehydrogenase (ATA1)</td>
</tr>
<tr>
<td>PAB5</td>
<td>A1g17710</td>
<td>2068/2069</td>
<td>823.25</td>
<td>106.96</td>
<td>polyadenylate-binding protein 5</td>
</tr>
<tr>
<td>At3g28470</td>
<td>2070/2071</td>
<td>758.75</td>
<td>22.98</td>
<td></td>
<td>myb transcription factor</td>
</tr>
<tr>
<td>At3g13220</td>
<td>2072/2073</td>
<td>794.97</td>
<td>14.42</td>
<td></td>
<td>ABC transporter</td>
</tr>
<tr>
<td>A1g69500</td>
<td>2074/2075</td>
<td>1636.47</td>
<td>16.85</td>
<td></td>
<td>cytochrome P450</td>
</tr>
<tr>
<td>At3g23770</td>
<td>2076/2077</td>
<td>497.53</td>
<td>17.17</td>
<td></td>
<td>beta-1,3-glucanase</td>
</tr>
<tr>
<td>A1g01280</td>
<td>2078/2079</td>
<td>430.88</td>
<td>1.96</td>
<td></td>
<td>cytochrome P450</td>
</tr>
<tr>
<td>FLF</td>
<td>At5g10140</td>
<td>2080/2081</td>
<td>251.54</td>
<td>20.77</td>
<td>MADS box protein FLOWERING LOCUS F (FLF)</td>
</tr>
<tr>
<td>ATMYB103</td>
<td>At5g56110</td>
<td>2084/2085</td>
<td>93.20</td>
<td>36.35</td>
<td>Regulate the tapetum and trichome development, anther specific gene.</td>
</tr>
<tr>
<td>A6</td>
<td>At4g14080</td>
<td>2098/2099</td>
<td>3602.11</td>
<td>41.2</td>
<td>Glycosyl hydrolase family 17 protein / anther-specific protein (A6)</td>
</tr>
<tr>
<td>A9</td>
<td>At5g07230</td>
<td>2100/2101</td>
<td>5099.79</td>
<td>81.5</td>
<td>lipid transfer protein, tapetum-specific protein A9</td>
</tr>
<tr>
<td>MYB33</td>
<td>At5g06100</td>
<td>2102/2103</td>
<td>238.76</td>
<td>176.9</td>
<td>myb family transcription factor (MYB33)</td>
</tr>
<tr>
<td>MYB65</td>
<td>At3g14400</td>
<td>2104/2105</td>
<td>335.48</td>
<td>203.4</td>
<td>myb family transcription factor (MYB65)</td>
</tr>
<tr>
<td>SDS</td>
<td>At1g06170</td>
<td>2127/2128</td>
<td>938.35</td>
<td>28.1</td>
<td>basic helix-loop-helix (bHLH) family protein</td>
</tr>
<tr>
<td>RCK/AtMER3</td>
<td>At3g27730</td>
<td>1963/1964</td>
<td>81.68</td>
<td>37.16</td>
<td>ROCK-N-ROLLERS/AtMER3, a ATP-dependent DNA helicase</td>
</tr>
<tr>
<td>UBQ1</td>
<td>At3g52590</td>
<td>U1/U2</td>
<td>4509.45</td>
<td>4007.97</td>
<td>UBIQUITIN EXTENSION PROTEIN 1 (UBQ1)</td>
</tr>
</tbody>
</table>

The genes were selected as described. The microarray intensity value of the gene for the ems1/exs anther is at most half of that from the wild type anther. Genes known to be important for early anther development are regulatory genes SPL/NZZ, TPD1, AtMYB33 and AtMYB65. Constitutive controls are AtMYB4 and the UBQ1.
Table 4.3. Primer used in real-time PCR.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>locuse</th>
<th>oMC number</th>
<th>product length</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPL</td>
<td>At4g27330</td>
<td>2362/2363</td>
<td>200</td>
</tr>
<tr>
<td>EMS1</td>
<td>At5g07280</td>
<td>2364/2365</td>
<td>101</td>
</tr>
<tr>
<td>TPD1</td>
<td>At4g24972</td>
<td>2366/2367</td>
<td>201</td>
</tr>
<tr>
<td>AMS</td>
<td>At2g16910</td>
<td>2368/2369</td>
<td>201</td>
</tr>
<tr>
<td>MS1</td>
<td>At5g22260</td>
<td>2370/2371</td>
<td>201</td>
</tr>
<tr>
<td>MYB33</td>
<td>At5g06100</td>
<td>2378/2379</td>
<td>201</td>
</tr>
<tr>
<td>MYB65</td>
<td>At3g11440</td>
<td>2380/2381</td>
<td>201</td>
</tr>
<tr>
<td>bHLH</td>
<td>At1g10610</td>
<td>2384/2385</td>
<td>201</td>
</tr>
<tr>
<td>DYT1</td>
<td>At4g21330</td>
<td>1872/1834</td>
<td>153</td>
</tr>
<tr>
<td>bHLH</td>
<td>At2g31210</td>
<td>2463/2464</td>
<td>201</td>
</tr>
<tr>
<td>bHLH</td>
<td>At1g06170</td>
<td>2465/2466</td>
<td>201</td>
</tr>
<tr>
<td>WD-40</td>
<td>At1g15850</td>
<td>2467/2468</td>
<td>201</td>
</tr>
<tr>
<td>ACT2</td>
<td>AT3G18780</td>
<td>1533/1534</td>
<td>108</td>
</tr>
</tbody>
</table>
**DYT1 encodes a putative bHLH transcription factor**

To gain further insights into its function, we cloned the *DYT1* gene. The *dyt1* mutant was from a *Ds* insertional line, but the *dyt1* mutation was genetically separable from the *Ds* element (data not shown). To clone the gene, we mapped the *dyt1* locus by analyzing approximately 500 mutant F2 progenies from a cross between the *dyt1* mutant (*Ler*) and Columbia wild type plant. The mapping results indicated that the *DYT1* gene was on chromosome 4, between At4g21220 and At4g21360 (data not shown). We sequenced candidate genes in this region from both wild type and the *dyt1* mutant, and found that only one locus, At4g21330, had an insertion mutation at 109bp upstream of the predicted translation initiation codon (Figure 4.4A). We performed TAIL PCR to obtain the DNA at the 5’ and 3’ ends of the insertion and found that they each partially matched the same region of a putative retro-transposon At5g33382 in the Columbia genomic sequence. Further PCR and sequence analysis indicated that the insertion at the *DYT1* locus matched exactly and completely to a seemingly intact retro-transposon in the *Ler* genome (W. Z., Y. S. and H. M., unpublished). To verify that At4g21330 is *DYT1*, we cloned an At4g21330 genomic fragment into a modified pCAMBIA1300 plasmid and use it to transform *dyt1/+* plants. Fifty independent transgenic plants were analyzed; twelve lines were found to be homozygous for the *dyt1* insertion. All twelve lines were fertile, including nine lines with normal fertility (Figure 4.1E and H), confirming that At4g21330 is the *DYT1* gene.

To determine the *DYT1* cDNA sequence, we performed an RT-PCR experiment with floral mRNA, and obtained a 624-bp product with an identical sequence to that indicated by the annotation. Additional RT-PCR experiments using primers that match sequences just beyond the annotated region yielded no product, confirming the annotated *DYT1* coding region. The *DYT1* gene encodes a putative transcription factor of 207 amino acid residues with a basic helix-loop-
helix (bHLH) domain. According to the annotation (http://www.arabidopsis.org), the bHLH domain spans the region from Phe$_{29}$ to Gln$_{78}$ (Figure 4.4B) (Toledo-Ortiz et al., 2003). Interestingly, a preliminary search of public databases using the DYT1 protein sequence with the BLAST program showed that the DYT1 protein has the highest similarity to the Arabidopsis AMS protein (Sorensen et al., 2003).

To gain additional insights into the phylogenetic relationship between DYT1, AMS, and other close homologs, we performed phylogenetic analysis of bHLH genes, including DYT1, AMS and a recently reported rice gene, UNDEVELOPED TAPETUM (OsUDT1), which is required for normal tapetum development (Jung et al., 2005). The phylogenetic analyses with the bHLH domain region alone, with both bHLH domain and conserved regions (Figure 4.4C), or with the full-length sequences, yielded trees with very similar topologies (others not shown). Our result indicates that DYT1, AMS, OsUDT1, Os02g02820 and PtDYT1-Like (Populus trichocarpa) form a separate clade within a group of related members of the bHLH gene family. Among them, AMS and Os02g02820 are supported as an orthologous pair, as are DYT1 and PtDYT1-like. In Arabidopsis, DYT1 and AMS are most closely related to each other, in agreement with the preliminary BLAST results. The rice OsUDT1 gene could be the ortholog of the DYT1 and PtDYT1-like genes.
Figure 4.4. The DYT1 gene structure and annotated conserved domain.

(A) The genomic region of the DYT1 gene. The dyt1 insertion is flanked by a direct repeat of 6-bp ACTTCT, which correspond to nucleotides 109-104 upstream of the annotated ATG (No.1 to 3 nucleotides) codon. The DYT1 gene has three exons represented as open boxes. (B) The annotated amino acid sequence of the DYT1 protein, with 207 amino acid residues. From Phe29 to the Gln78 is the conserved bHLH domain, which corresponds to the dark region in the schematic box image and is underlined in the amino acid sequence. (C) An unrooted Neighbor Joining tree of Arabidopsis, rice, and Poplar bHLH genes in the same subfamily as DYT1. Gene ID numbers starting with “At” indicate genes from Arabidopsis thaliana; names of genes with functional information are given after the gene ID numbers. Gene ID numbers starting with “Os” indicates genes from rice (Oryza sativa); UDT1 is shown as OsUDT1. “Pt” indicated genes from Poplar (Populus trichocarpa), with temporary names given according to gene ID from the Floral Genome Project. Bootstrap values are show near the relevant nodes.
**DYT1 is preferentially expressed in tapetal cells**

To determine the *DYT1* expression pattern, we performed a real-time PCR experiment (Figure 4.5A). *DYT1* expression was detected at a low level in young inflorescences with meiotic cells and in siliques, and at a high level in wild type anthers from stage 4 to 7, but not in vegetative tissues, or in the post-meiotic stage-12 flower. In addition, RNA in situ hybridization experiments showed that *DYT1* expression was detectable in the floral meristem and early anther primordia (Figure 4.5C), and in archesporial cells at stage 2 (not shown). From stage 4 to early stage 5, a weak signal was detected in precursors of the middle layer, tapetum and meiocytes (Figure 4.5D). From late stage 5 to early stage 6 (Figure 4.5E and H), the *DYT1* expression reached its highest level in the tapetum and is at a low level in meiocytes. However, at late stage 6, the *DYT1* expression signal was drastically reduced to background levels (Figure 4.5F). The *DYT1* expression pattern is consistent to the observed tapetal defects in the *dyt1* mutant anther. In the gynoeceum, a weak signal was detected (data not shown). In the *dyt1* anther, there was a low level expression, which was not specific to the tapetum (Figure 4.5I). Therefore, the insertion upstream of the *DYT1* ATG start codon caused a great reduction of the level of *DYT1* expression in the tapetum. The weak expression suggests that the *dyt1* allele may have some residual function, which might account for the occasional seeds that were observed.
Figure 4.5. The DYT1 expression pattern. (A) Detection of DYT1 expression using real-time PCR in Ler background. DYT1 expression was not detected in any vegetative tissues or stage-12 flower, detected at low levels in the young inflorescence and siliques and at the highest level in the anther. (B) Detection of DYT1 expression using real-time PCR in Ler, ems1 and spl inflorescences. The DYT1 expression was not detected in spl, and detected in ems1 at about 17% of the normal level. (C) to (K) RNA in situ hybridization with a DYT1 probe. (C) to (F) and (H) DYT1 expression in the Ler background. (C) The DYT1 signal was detected in the floral meristem. (D) An anther at stage 4 to early stage 5. The DYT1 signal can be detected mainly within the newly formed tapetum and meiocytes. (E) and (H) At late stage 5, a strong signal is detected in the tapetal cell layer, whereas the signal in the meiocytes is much weaker. (F) At late stage 6, the DYT1 signal is greatly reduced, with residual expression in some meiocytes. (I) A dyt1 mutant anther at late stage 5; The DYT1 signal is low and non-specific in the entire anther. (J) A spl mutant anther at late stage 5. The DYT1 signal is at the background level. (K) An ems1 mutant anther locule at late stage 5. Uniformly weak DYT1 signal can be detected in meiocytes and little signal in cells surrounding the meiocytes. (G) The sense control with a Ler late stage 5 anther. Only background is seen. Rt, root; Sm, stem; Lf, leaf; Se, siliqua; S12, stage 12 flower; Inf, inflorescence; Ar, anther; WT-Inf, wild type inflorescence; ems1-Inf, ems1 inflorescence; spl-Inf, spl inflorescence. T, tapetum; Ms, meiocytes; I, indeterminate cells; E-Ms, excess meiocytes. Scale bars: 20 µm in C-K.
DYT1 expression is positively regulated by SPL/NZZ and EMS1/EXS

To test whether the DYT1 expression was affected by any known early anther development genes, we performed real-time PCR experiments with RNA from spl and ems1 mutant floral buds. Our results showed that the DYT1 expression was barely detectable in the spl inflorescences and that the expression level in the ems1 inflorescences was only ~17% of the wild type level (Figure 4.5B). These results suggest that DYT1 might be downstream of SPL/NZZ and EMS1/EXS. We performed RNA in situ experiments to verify this possibility. In the spl anther, little DYT1 expression could be detected (Figure 4.5J). A weak signal could be detected in the meiocytes in the ems1 anther at late stage 5 (Figure 4.5K) and early stage 6, but, unlike the wild type anther, the strong tapetal signal was not found in the ems1 anther (compare Figure 4.5H with 5K). Therefore, the strong DYT1 expression in the tapetum requires EMS1/EXS. Both the real-time PCR and the RNA in situ hybridization results indicate that SPL/NZZ is essential for DYT1 expression and that EMS1/EXS promotes the high-level DYT1 expression specific to the tapetum.

The expression of anther genes in the dyt1 mutant is altered

The finding that DYT1 encodes a bHLH-type putative transcription factor suggests that DYT1 controls gene expression required for normal anther development. To test this hypothesis, we performed RT-PCR using primers for anther genes. We obtained results for a total of 32 genes (Figure 4.6); among these, SPL/NZZ, EMS1/EXS and TPD1 are known early anther development genes (Schiefthaler et al., 1999; Yang et al., 1999; Canales et al., 2002; Zhao et al., 2002; Yang et al., 2003; Yang et al., 2005). The other genes were chosen for their tapetum-preferential expression according to either previous reports or unpublished gene expression data.
obtained in our lab (Table 4.2). We found that for 21 genes out of 32, the expression was significantly reduced in the \textit{dyt1} mutant compared to the wild type (Figure 4.6), indicating that indeed the normal expression of a large number of genes depends on the \textit{DYTI} gene function. In particular, two regulatory genes, \textit{MS1} and \textit{AMS}, which are important for tapetum development, exhibited greatly reduced levels of expression, suggesting that \textit{MS1} and \textit{AMS} act downstream of \textit{DYTI} (Wilson et al., 2001; Ito and Shinozaki, 2002; Sorensen et al., 2003). In contrast, some tapetum preferential genes, such as A6 and A9, were still expressed in the \textit{dyt1} mutant at slightly reduced levels. In addition, the expression of \textit{SPL}, \textit{TPD1}, \textit{EMS1/EXS}, \textit{AtMYB33} and \textit{AtMYB65} was not dramatically different in the \textit{dyt1} mutant, indicating that their expression does not require \textit{DYTI}. To verify the RT-PCR results, selected genes were further analyzed using real-time PCR and the results (Figure 4.6D) support the conclusion that \textit{AMS} and \textit{MS1} expression requires \textit{DYTI} function.

In addition to the tapetum-preferential genes, we also tested the expression of two meiosis-specific genes: \textit{SDS} and \textit{RCK/AtMER3} (Azumi et al., 2002; Chen et al., 2005; Mercier et al., 2005). Although the expression level of \textit{SDS} did not change, the expression of \textit{RCK/AtMER3} was significantly reduced. \textit{SDS} is known to act earlier than \textit{RCK/AtMER3} in prophase I, suggesting that the \textit{dyt1} mutation might affect the expression of late prophase I genes more than early prophase I genes.
### Table 4.2. Microarray data of the genes selected in RT-PCR.

<table>
<thead>
<tr>
<th>gene name</th>
<th>locus</th>
<th>oMC</th>
<th>Av-WT</th>
<th>Av-ems1</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPL</td>
<td>At4g27330</td>
<td>2044/2045</td>
<td>402.50</td>
<td>283.52</td>
<td>SPOROCYTELESS/NOZZLE</td>
</tr>
<tr>
<td>EMS1</td>
<td>At5g07280</td>
<td>509/511</td>
<td>1062.95</td>
<td>146.89</td>
<td>EXCESS MICROSPOROCYTES1/EXTRA SPOROGENOUS CELLS</td>
</tr>
<tr>
<td>TPD1</td>
<td>At4g24972</td>
<td>2046/2047</td>
<td>348.88</td>
<td>9.39</td>
<td>ABORTED MICROSPORES</td>
</tr>
<tr>
<td>MS5</td>
<td>At4g20900</td>
<td>2050/2051</td>
<td>75.46</td>
<td>23.20</td>
<td>male sterility MS5; pollenless3</td>
</tr>
<tr>
<td>MS5 like</td>
<td>At5g44330</td>
<td>2052/2053</td>
<td>90.26</td>
<td>1.51</td>
<td>MS5 family protein</td>
</tr>
<tr>
<td>MS1</td>
<td>At5g22260</td>
<td>2054/2055</td>
<td>300.93</td>
<td>63.67</td>
<td>male sterile protein 1</td>
</tr>
<tr>
<td>AtMYB32</td>
<td>At4g34990</td>
<td>2056/2057</td>
<td>161.19</td>
<td>179.94</td>
<td>putative transcription factor (MYB4)</td>
</tr>
<tr>
<td>AtMYB4</td>
<td>At4g38620</td>
<td>2058/2059</td>
<td>90.26</td>
<td>1.51</td>
<td>MYB-like protein myb-related protein Y49</td>
</tr>
<tr>
<td>MS2</td>
<td>At3g11980</td>
<td>2060/2061</td>
<td>690.78</td>
<td>12.25</td>
<td>male sterility protein 2 (MS2)</td>
</tr>
<tr>
<td>LPT12</td>
<td>At3g51590</td>
<td>2062/2063</td>
<td>446.85</td>
<td>40.33</td>
<td>lipid transfer protein</td>
</tr>
<tr>
<td>ATSTP2</td>
<td>At1g07340</td>
<td>2064/2065</td>
<td>174.90</td>
<td>60.65</td>
<td>hexose transporter</td>
</tr>
<tr>
<td>ATA1</td>
<td>At3g42960</td>
<td>2066/2067</td>
<td>2090.38</td>
<td>29.80</td>
<td>alcohol dehydrogenase (ATA1)</td>
</tr>
<tr>
<td>PAB5</td>
<td>At1g71770</td>
<td>2068/2069</td>
<td>823.25</td>
<td>106.96</td>
<td>polyadenylate-binding protein 5</td>
</tr>
<tr>
<td>At3g28470</td>
<td>2070/2071</td>
<td>758.75</td>
<td>22.98</td>
<td>myb transcription factor</td>
<td></td>
</tr>
<tr>
<td>At3g13220</td>
<td>2072/2073</td>
<td>794.97</td>
<td>14.42</td>
<td>ABC transporter</td>
<td></td>
</tr>
<tr>
<td>At1g69500</td>
<td>2074/2075</td>
<td>1636.47</td>
<td>16.85</td>
<td>cytochrome P450</td>
<td></td>
</tr>
<tr>
<td>At3g23770</td>
<td>2076/2077</td>
<td>497.53</td>
<td>17.17</td>
<td>beta-1,3-glucanase</td>
<td></td>
</tr>
<tr>
<td>At1g01280</td>
<td>2078/2079</td>
<td>430.88</td>
<td>1.96</td>
<td>cytochrome P450</td>
<td></td>
</tr>
<tr>
<td>FLF</td>
<td>At5g10140</td>
<td>2080/2081</td>
<td>251.54</td>
<td>20.77</td>
<td>MADS box protein FLOWERING LOCUS F (FLF)</td>
</tr>
<tr>
<td>ATMYB103</td>
<td>At5g56110</td>
<td>2084/2085</td>
<td>93.20</td>
<td>36.35</td>
<td>Regulate the tapetum and trichome development, anther specific gene.</td>
</tr>
<tr>
<td>A6</td>
<td>At4g14080</td>
<td>2098/2099</td>
<td>3692.11</td>
<td>41.2</td>
<td>Glycosyl hydrolase family 17 protein / anther-specific protein (A6)</td>
</tr>
<tr>
<td>A9</td>
<td>At5g07230</td>
<td>2100/2101</td>
<td>5099.79</td>
<td>81.5</td>
<td>lipid transfer protein, tapetum-specific protein A9</td>
</tr>
<tr>
<td>MYB33</td>
<td>At5g06100</td>
<td>2102/2103</td>
<td>238.76</td>
<td>176.9</td>
<td>myb family transcription factor (MYB33)</td>
</tr>
<tr>
<td>MYB65</td>
<td>At3g11440</td>
<td>2104/2105</td>
<td>335.48</td>
<td>203.4</td>
<td>myb family transcription factor (MYB65)</td>
</tr>
<tr>
<td>SDS</td>
<td>At1g14750</td>
<td>271/272</td>
<td>80.48</td>
<td>27.18</td>
<td>SOLO DANCER, a putative cyclin</td>
</tr>
<tr>
<td>RCK/AtMER3</td>
<td>At3g27730</td>
<td>1963/1964</td>
<td>81.68</td>
<td>37.16</td>
<td>ROCK-N-ROLLERS/AtMER3, a ATP-dependent DNA helicase</td>
</tr>
<tr>
<td>UBQ1</td>
<td>At3g52590</td>
<td>U1/U2</td>
<td>4509.45</td>
<td>4007.97</td>
<td>UBIQUITIN EXTENSION PROTEIN 1 (UBQ1)</td>
</tr>
</tbody>
</table>

The genes were selected as described. The microarray intensity value of the gene for the

*ems1/exs* anther is at most half of that from the wild type anther. Genes known to be important

for early anther development are regulatory genes *SPL/NZZ*, *TPD1*, *AtMYB33* and *AtMYB65*. 

Constitutive controls are *AtMYB4* and the *UBQ1*. 

---

158
Figure 4.6. Expression of anther and tapetum genes in the wild type and dyt1 mutant. (A) Each of the 11 genes, including the controls UBQ1 and MYB4, show either normal or increased expression in the dyt1 mutant background compared to the wild type. (B) and (C) Twenty one genes show significantly decreased expression in the dyt1 mutant background. (D) Real-time PCR of selected anther genes in both wild type and dyt1 backgrounds. (E) Effects of overexpression of DYT1 on selected anther genes. No significant effects were observed by DYT1 overexpression even though the DYT1 expression levels were elevated tens of folds. (F) Expression pattern of selected putative regulatory genes in spl, ems1 and dyt1 background.
**DYT1 is not sufficient for tapetum development**

Both *ems1/exs* and *dyt1* mutations affect tapetum development. Previous reports and our results suggest that *DYT1* acts downstream of *EMS1/EXS*. To test this further, we generated the *ems1 dyt1* double mutant and examined its early anther development. We found that the double mutant resembles the *ems1* mutant in that the double mutant anther also completely lacks the tapetum (Figure 4.7), suggesting that indeed *EMS1/EXS* is upstream of *DYT1* in the same pathway. In addition, to test whether *DYT1* is sufficient to alter anther development, we generated transgenic plants carrying a 35S-*DYT1* fusion in wild type, *spl/nzz* and *ems1/exs* backgrounds. Transgenic lines with *DYT1* over-expression were identified by RT-PCR (Figure 4.6E and data not shown) and analyzed for their anther morphology. In all cases, the 35S-*DYT1* transgenic anthers had morphologies resembling those of the corresponding genotypes without the transgene (Figure 4.8). Therefore, the over-expression of *DYT1* was not able to suppress the *spl/nzz* and *ems1/exs* mutant phenotypes, indicating that other genes acting downstream of *SPL/NZZ* and *EMS1/EXS* are likely required for normal tapetum development. Because *DYT1* was found to be required for normal expression of a number of tapetum genes, we tested selected genes by real-time PCR to determine whether the 35S-*DYT1* transgene was able to stimulate the expression of these genes. Our results indicate that the 35S-*DYT1* transgene did not alter the expression of these genes substantially (Figure 4.6E), consistent with the morphological results.
Figure 4.7. The *ems1 dyt1* double mutant anther morphology at late stage 5. (A) Wild type. (B) *dyt1*. (C) *ems1*. (D) *ems1 dyt1* double mutant. The double mutant lacks the tapetum. T, tapetum; ML, middle layer; Ms, meiocytes; E-Ms, excess meiocytes; I, indeterminate cells. Scare bar: 10 µm.
Figure 4.8. Phenotypes of wild type and mutant stage 5 anthers. (A) Ler at early stage 5. (B), (D), (F) and (H) are anther sections from 35S-DYT1 transgenic plants in wild type (B, early stage 5), wild type (D, late stage 5), ems1 (F) and spl (H) backgrounds. (C) Ler at late stage 5. (E) ems1 at late stage 5. (G) spl at stage 5. The 35S-DYT1 transgene did not seem to alter anther morphology. T, tapetum; Ms, meiocytes. Scale bar: 10 µm.
Our results and previous studies suggest that additional genes other than \textit{DYT1} probably function downstream of \textit{SPL/NZZ} and \textit{EMS1/EXS} to promote tapetum development. It is possible that some of these genes might encode regulatory proteins. For example, the \textit{AtMYB33} and \textit{AtMYB65} genes are known to be important for tapetum development, and might also function downstream of \textit{SPL/NZZ} and \textit{EMS1/EXS}. To test these ideas, we examined the expression in wild type and mutants of these genes and other genes, which were identified to be tapetum-preferential from our microarray data (W.Z., Y.S., H.M., unpublished data) and encode bHLH or WD-40 proteins. Real-time PCR results (Figure 4.6F) indicate that the expression of two bHLH genes (At2g31210 and At1g06170) was greatly reduced in the \textit{spl} mutant, and one of them (At2g31210) was expressed at a lower level in the \textit{ems1} mutant. In contrast, the expression levels of \textit{AtMYB33}, \textit{AtMYB65} and two other genes were either nearly normal or increased.

4.5 Discussion

\textit{DYT1} is important for tapetum development and function

Our analysis of the \textit{dyt1} mutant phenotype indicates that \textit{DYT1} is required for normal tapetum development following its formation. Furthermore, \textit{DYT1} is expressed preferentially in the tapetum at the later stage 5, consistent with its role in this layer. The \textit{EMS1/EXS}, \textit{SERK1/SERK2}, and \textit{TPD1} genes (Canales et al., 2002; Zhao et al., 2002; Yang et al., 2003; Albrecht et al., 2005; Colcombet et al., 2005; Yang et al., 2005) are important for the formation of the tapetal layer; therefore, these genes likely act at earlier stages than \textit{DYT1}, as supported by the observed strong tapetal expression of \textit{EMS1/EXS} and \textit{TPD1} (Canales et al., 2002; Zhao et al., 2002; Yang et al., 2003) prior to the strong \textit{DYT1} expression in the tapetum. Our expression and
double mutant analyses support the hypothesis that DYT1 acts downstream of SPL/NZZ and EMS1/EXS.

AMS and MS1 are also important for tapetum function, but they are required at post-meiotic steps (Wilson et al., 2001; Ito and Shinozaki, 2002; Sorensen et al., 2003). In ams anthers, meiosis is normal and microspores are formed; however, the newly formed microspores soon degenerate. In ms1 anthers, pollen development is abnormal and no normal mature pollen is produced. Compared with AMS and MS1, DYT1 acts at an earlier stage, before the completion of meiosis. Therefore, DYT1 is required for a key step in tapetum development. In other words, tapetum development requires the combined activities of the EMS1/EXS, SERK1/SERK2, TPD1, DYT1, AMS and MS1 genes: first EMS1/EXS, SERK1/SERK2 and TPD1 specify the tapetal cells as distinct from meiocytes at the time of the cell division that form the tapetal cells, then DYT1 is required to promote correct tapetal cell fate for proper function, and finally AMS further regulates the tapetal cell function supporting normal microspore development.

**DYT1 is required for normal levels of the expression of tapetum genes**

As DYT1 encodes a bHLH putative transcription factor, it is likely that it regulates the expression of tapetal genes. We found that the expression of a majority of tapetum-preferential genes tested depends on DYT1. The greatly reduced expression in the dyt1 mutant of many tapetum-preferential genes, particularly those encoding transcription factors, supports the idea that DYT1 is a key component of a regulatory step in normal tapetum development. The Arabidopsis bHLH gene family has over 140 members, making this the second largest gene family of transcription factors (Toledo-Ortiz et al., 2003). Although DYT1 and AMS clearly have non-redundant and distinct functions, they are members of the same subfamily. Phylogenetic
analysis performed here including the closest rice homologs of DYT1 indicates that the rice gene, OsUDT1, is also a member of this subfamily and a putative DYT1 ortholog. A mutation in the OsUDT1 gene results in a defective tapetum, similar to the dyt1 tapetum. In addition, the expression pattern of the OsUDT1 gene is slightly different from that of DYT1 (Jung et al., 2005). Therefore, this bHLH subfamily contains phylogenetically and functionally distinct members.

Some tapetum marker genes, such as A6 and A9, were expressed in the dyt1 mutant at slightly reduced levels. It is possible that other genes are also important for the activation of some tapetum genes. In addition to DYT1, previous reports described mutants with similar phenotypes, such as fat tapetum, gne1 and gne4 (Sanders et al., 1999; Sorensen et al., 2002). Although the molecular nature of the FAT TAPETUM, GNE1 and GNE4 genes are unknown at this time, it is likely that additional loci are involved in defining the tapetal cell fate. In other words, DYT1 is essential, but not sufficient, for the specification of the tapetum cell fate, as supported by our observation that over-expression of DYT1 did not alter anther phenotypes in wild type or mutant backgrounds. Recently, it was reported that the AtMYB33 and AtMYB65 genes redundantly facilitate tapetum development, with the double mutant having tapetum defects before the completion of meiosis (Millar and Gubler, 2005). It is known that bHLH transcription factors can form homodimers or heterodimers with other bHLH proteins. In some cases, it has been shown that bHLH proteins can form complexes with MYB proteins and WD-40 proteins (Ramsay and Glover, 2005). It is possible that the AtMYB33 and AtMYB65 proteins may form heterodimers with DYT1 to regulate tapetum-preferential gene expression. This idea is consistent with our result that the expression of these two MYB genes is not altered in the dyt1 mutant.
Our analysis of gene expression in the spl mutant suggest that two other genes (At2g31210 and At1g06170) (Figure 4.6F) encoding bHLH proteins might also act downstream of SPL/NZZ and that these bHLH genes might in turn regulate other genes. Indeed, an examination of upstream sequences (500 bp from the ATG start codon) of 163 putative tapetum-preferential genes (identified from microarray data, W. Z., A. Wijeratne, Y. S., and H. M., unpublished results) revealed that 143 of them have potential binding sites for bHLH proteins, and 69 genes have putative MYB-binding sites, with 59 genes have both types of elements. These observations support the hypothesis that bHLH and MYB proteins regulate over-lapping but non-identical sets of tapetum-preferential genes. In particular, 18 of 21 genes that exhibit reduced expression in the dyt1 mutant have putative bHLH-binding sites; in addition, between 500-1000 bp upstream of the AMS ATG start codon, there is a bHLH-binding consensus site. Further experiments are needed to test whether these genes are direct targets of DYT1 and/or other bHLH proteins.

**DYT1 supports completion of meiosis**

It is known that a functional tapetum is required for normal pollen development following meiosis, as shown by molecular ablation studies and the characterization of mutants such as ams (Aarts et al., 1997; Wilson et al., 2001; Ito and Shinozaki, 2002; Sorensen et al., 2003). In the ems1/exs, serk1 serk2, and tpd1 mutants, the tapetum is missing and excess meiocytes occupy the position of the tapetum (Zhao et al., 2002; Albrecht et al., 2005; Colcombet et al., 2005; Yang et al., 2005). Nevertheless, meiotic nuclear divisions still occur in the ems1 mutant, indicating that the tapetum is not required for meiotic nuclear events. On the other hand, the ems1 meiocytes do not undergo cytokinesis, suggesting that tapetum might be needed for the completion of meiosis. The dyt1 mutant phenotypes provide further support for
this idea. In addition to a morphologically abnormal tapetum, the dytl mutant meiocytes were found to have thinner callose cell walls than normal, suggesting that normal tapetum function is needed for the formation of the callose wall. Nevertheless, DYT1 expression was detected at a low level in the meiocytes and expression of some meiotic genes was reduced; therefore, it is possible that DYT1 might also function in meiocytes.

A model for DYT1 function and the control of tapetum identity

This and previous studies support a model for the genetic control of tapetum development and function (Figure 4.9), although evidence for biochemical interactions is not yet available. SPL/NZZ is required for the formation of sporogenous cells and surrounding somatic cell layers, including the tapetum (Yang et al., 1999). Recently, Ito et al. showed that AG is a direct activator of SPL/NZZ expression (Ito et al., 2004). EMS1/EXS, TPD1 and SERK1/SERK2 are required for the formation of tapetum (Zhao et al., 2002; Yang et al., 2003; Albrecht et al., 2005; Colcombet et al., 2005). In addition, phenotypic changes caused by TPD1 overexpression are dependent on EMS1/EXS (Yang et al., 2005). We show here that SPL/NZZ and EMS1/EXS positively regulate expression of DYT1. The AtMYB33 and AtMYB65 genes are expressed in the tapetum and their expression does not require the SPL/NZZ, EMS1/EXS, or DYT1 gene.

In addition, it is likely that the genes depending on DYT1 for normal expression support the tapetum function that produces the enzymatic activities and materials needed for pollen development (Dickinson and Bell, 1976; Liu and Dickinson, 1989; Hernould et al., 1998; Rubinelli et al., 1998; Taylor et al., 1998; Zheng et al., 2003; Scott et al., 2004). Among the genes regulated by DYT1 are AMS and MSI, which also encode transcription factors that likely regulate late tapetum genes. Analysis of DYT1 overexpression transgenic plants and expression
studies in various mutants suggest strongly that other genes are needed for normal tapetum
development. Phenotypic similarities suggest that $AtMYB33$ and $AtMYB65$, as well as potentially
others (such as $GNE1$) (Sorensen et al., 2002), may act at approximately the same step as $DYT1$.
Therefore, $DYT1$ is a critical component at a key step in the regulatory network responsible for
tapetum development and function (Figure 4.9).
Figure 4.9. A model for *DYT1* function and tapetum specification. The thin arrows indicate a positive genetic regulation. The arrowheads represent gene functions controlling a developmental stage. The open arrows indicate development from one stage to the next.
Acknowledgments

We thank J.-P. Vielle Calzada for identifying the Ds line carrying the dyt1 mutation, W. Yang, D. Zhao, and D. Ye for spl, ems1 and tpd1 mutants, A. Omeis and J. Wang for plant care, D. Grove and A. Price for DNA sequencing and real-time PCR, and G. Ning, R. Haldeman, and M. Hazen for assistance in preparing anther sections and for use of the Olympus BX-51 microscope and camera. We thank A. Wijeratne for help in real-time PCR and B. Feng, C. H. Hord, W. Hu, A. Surcel, G. Wang, A. Wijeratne, and L. Zahn for helpful comments on the manuscript. This work was supported by a grant from the US Department of Energy (DE-FG02-02ER15332) to H. Ma and used plant materials generated with support from a National Science Foundation grant (0215923). H. Ma gratefully acknowledges the support of the John Simon Guggenheim Memorial Foundation.
References


crossover classes cohabit in *Arabidopsis*: one is dependent on *MER3*, whereas the other one is not. Curr. Biol. 15, 692-701.


Chapter 5

Conclusions and Perspective
In this thesis, I have performed the microarray experiments, the molecular evolutonal study of the KFB family, and the functional study of DYT1 to understand anther development/meiosis in Arabidopsis.

By performing microarray experiments with mRNA from wild type anthers at stages 4 to 6 and wild type young inflorescences (floral stages 1 to 9), 707 genes were identified to be expressed at least 2 fold higher in anthers than in young inflorescences and were defined as A/I differential genes. The study of gene expression patterns also found that 250 of 707 genes are anther preferential genes, which are expressed in anthers at least 2 fold higher than in any other tissues. Among 250 genes, 106 genes are anther specific genes, which are only expressed in anthers. The cluster analysis of expression patterns of these 707 genes uncovered that 359 genes formed a co-expressed cluster named as the C4 cluster. Most of the C4 cluster genes are anther preferential genes and are expressed higher in anther than in any other tissues tested. Almost all the known genes that were reported to be essential for early anther development and/or male meiosis are in the C4 cluster. The analysis of anther microarray data of wild type, spl and ems1 further revealed that among 359 genes in the C4 cluster, 330 (85%) genes were down regulated in spl, 288 (74%) genes were down regulated in ems1, and 272 (70%) were down regulated in both spl and ems1 (Yang et al., 1999; Zhao et al., 2002). A gene was defined as the down regulated gene in a mutant when its expression level in the mutant decreased more than 2 fold than that in the wild type. The spl anther lacks both tapetal cells and meiotic cells (Yang et al., 1999), whereas the ems1 anther lacks only tapetal cells (Zhao et al., 2002). These results suggest that genes in the C4 cluster might specify the features of early anther development. It is possible that many of these genes might be involved in the formation and development of tapetal cells and meiotic cells. The analysis of gene functional categories and known functional domains further
uncovered that a number of genes in the C4 cluster encode putative transcription factors, protein kinases, chromosome remodeling related proteins, putative meiosis proteins and SCF components. Future studies of their biological functions through reverse genetic approaches could provide more information to understand early anther development and male meiosis in Arabidopsis. The available T-DNA or Ds-transposon insertional lines of these genes can be obtained to investigate their functions. Double mutants or triple mutants can also be generated for functional analyses of closely related homologs or for analyzing the relationship of genes in the putative genetic networks of anther development or male meiosis. For some candidate genes without any mutant lines, the RNA-interference (RNAi) method can be used to determine their possible functions.

The comparison of the anther microarray data (anther stages 4 to 6) of wild type with the meiotic mutants sds, mmd1 enabled to identify 240 genes regulated by SDS and 39 genes regulated by MMD1 (Azumi et al., 2002; Yang et al., 2003). Among 240 genes regulated by SDS, 231 genes were down regulated in the sds anther and were defined as SDS-dependent genes. This result indicates that SDS might be a positive regulator of gene expression during meiosis. Among the SDS-dependent genes (231 genes), a total of 99 genes were identified as co-expressed genes within a cluster called as S6 by the gene cluster analysis. All the known function genes regulated by SDS were in the S6 cluster, suggesting that genes in the S6 cluster might specify the features of male meiotic cells in Arabidopsis. Interestingly, most of MMD1-depedent genes (33 of 39 genes) are also SDS-dependent genes, suggesting that MMD1 might play a role downstream of SDS functional pathway during male meiosis. The analysis of mutant phenotypes also supports this idea. In sds, meiotic cells are defective in homologs paring, synapsis and chromosome recombination, and mutant homologs form univalent, instead of bivalents (Azumi et al., 2002).
The \textit{mmd1} meiotic chromosome is normal during the meiosis prophase I. Soon after diakinesis, \textit{mmd1} meiotic cells die with chromosome fragmentation and cytoplasmic shrinkage (Yang et al., 2003). To further investigate the relationship of \textit{SDS} and \textit{MMD1}, the \textit{sds mmd1} double mutant needs to be generated and analyzed phenotypically.

Previous studies have shown that cyclins are key regulators of cyclin-dependent kinases (CDKs), which control cell cycle progression by phosphorylating their downstream targets, including some transcription factors (Morgan, 1997). As the only known meiotic specific cyclin, \textit{SDS} might regulate the transcription of these \textit{SDS}-dependent genes through a cyclin-CDK pathway (Azumi et al., 2002; Wang et al., 2004). The \textit{cis}-element analysis of putative promoter sequences of the co-expressed genes in the S6 culster needs to be performed in the near future to identify the enriched binding motifs of some transcription factors, which could provide clues to the \textit{SDS}-CDK downstream targets. Further functional studies of candidate genes involved in the \textit{SDS} pathway can provide better understanding of male meiosis in Arabidopsis.

The functional study of \textit{ASK1} supported the idea that the ubiquitin dependent protein degradation pathway might play essential roles during male meiosis in Arabidopsis (Zhao et al., 1999; Yang et al., 2006; Zhao et al., 2006). Since \textit{ASK1} is expressed ubiquitously, its function in male meiosis might be acquired through its meiotic specific partner(s), the F-box protein(s), which specifies the meiotic important protein(s) for degradation (Zheng et al., 2002; Zhao et al., 2003). Some meiotic specific genes encoding F-box proteins were identified in the C4 cluster discussed above. Among them, several genes encode Kelch-repeat containing F-box proteins (KFBs). Previous studies have reported three KFBs (members of the ZTL family) that play critical roles in controlling flowering time and circadian oscillator (Nelson et al., 2000; Somers et al., 2004; Yasuhara et al., 2004). Another F-box protein, \textit{UNUSUAL FLORAL ORGANS} (UFO),
which contains a degenerated Kelch motif, plays essential roles in both floral meristem and floral organ development (Ingram et al., 1995; Samach et al., 1999; Zhao et al., 2001). UFO can activate \textit{APETALA3} and \textit{PISTILLATA} genes required for petal and stamen identities (Ingram et al., 1995; Wilkinson and Haughn, 1995; Samach et al., 1999). All these results suggest that some of the anther preferential KFBs might play some roles related with early anther development and/or male meiosis.

To gain further insights into KFBs, I performed genome-wide analysis of the KFB family to investigate its evolutionary history. My studies found that the origin of KFBs might occur before the divergence of pants and animals. Unlike animal genomes encoding a single copy of KFB, plant genomes encode a large number of KFBs with various numbers of Kelch motifs. Among the 18 plant KFB subfamilies, most of them are well conserved in land plants. The G5 subfamily underwent rapid gene duplication in Arabidopsis and \textit{Brassica}. Most of the G5 KFBs are present as tandem repeats and are expressed below reliable levels in the selected tissues, while other groups are ubiquitously expressed at much higher levels in Arabidopsis. The results suggest that the mechanism of the flower timing control and circadian oscillator may be conserved in flowering plants. The presence of six anther-specific \textit{AtKFBs} in G5 subfamily suggests that they might gain some new functions in early anther development and/or male meiosis. Three of them \textit{AtKFB29}, \textit{AtKFB32}, \textit{AtKFB75} were in the C4 cluster of A/I differential genes discussed above.

Besides tandem duplications, two other gene duplication types, the segmental duplication and retrotranspositional gene duplication, have been reported to be important for the evolution of gene families. Further analysis needs to be done to investigate these types of gene duplications as possible mechanisms for the birth of KFBs (Vision et al., 2000; Baumbusch et al., 2001; Cannon
et al., 2004). In addition, since the G5 subfamily has evolved rapidly, it is necessary to investigate if they evolved through positive selection by performing the dN/dS analysis (Nei and Gojobori, 1986). Also it would be informative to investigate the functions of the recently evolved anther specific KFBs through reverse genetic approaches discussed above.

Since early anther development and/or male meiosis mutations often result sterile plants or plants with reduced fertility, forward genetic approaches have been effectively used to identify genes that are important for early anther development and/or male meiosis (Ma, 2005). In collaboration with Wei Zhang in our lab, I successfully identified the DYT1 gene, which controls normal tapetum development and function during male meiosis. The functional study of DYT1 further supports the previous idea that tapetum is not essential for early male meiosis, but is required for male meiotic cytokinesis. The observation that the phenotype of the dyt1 ems1 double mutant is same as the ems1 single mutant suggests that DYT1 is a downstream component of the EMS1 pathway (Yang et al., 1999; Zhao et al., 2002). Both gene expression and phenotypic studies indicate that DYT1 acts downstream of both SPL and EMS1.

DYT1 encodes a bHLH transcription factor. Our expression studies indicate that DYT1 is required for the normal expression of some tapetum preferential genes, including AMS and MS1 that are required for the normal tapetum function after meiosis (Wilson et al., 2001; Ito and Shinozaki, 2002; Sorensen et al., 2003). Both gene expression and functional studies support the hypothesis that DYT1 acts up stream of AMS and MS1. The bHLH domain sequence analysis indicates that DYT1 is likely an E-box binding protein. The finding that the putative promoter sequences of AMS and MS1 contain several E-box motifs suggests that both AMS and MS1 might be the direct targets of DYT1 (Wilson et al., 2001; Ito and Shinozaki, 2002; Heim et al., 2003; Sorensen et al., 2003; Toledo-Ortiz et al., 2003). The fact that AMS and MS1 are expressed
strongly when DYT1 expression is rapidly reduced is also consistent with this idea (Ito, 2002 #50; Sorensen, 2003 #76; Wilson, 2001 #80). To investigate the downstream relationships of DYT1 with AMS and MSI, the dyt1 ams and dyt1 ms1 double mutants need to be generated and analyzed phenotypically. To further test this idea, a yeast one-hybrid experiment can be used to test the interactions of the DYT1 protein and the putative promoter sequences of AMS and MSI. This hypothesis can be further tested by the chromatin immunoprecipitation (ChIP) experiment to test protein-DNA interaction in vivo (Odom et al., 2004). In addition, the dexamethasone (DEX) inducible experiment can also be used to identify the direct targets of DYT1 (Duanmu et al., 2002). To identify additional DYT1 downstream targets, microarray experiments need to be performed by using the mRNA isolated from anthers (stages 4 to 7) of wild type and dyt1. The microarray data and cis-element analysis would identify some candidate genes, which could also be tested by both ChIP and DEX-inducible experiments.

Previous studies have reported three Arabidopsis mutants, fat tapetum, gne1 and gne4, whose phenotypes are similar to that of dyt1. Although these genes are not cloned yet, GNE1 and GNE4 have been reported to be located on different chromosomes and GNE4 is on chromosome 4, same as DYT1 (Chaudhury et al., 1994; Sanders et al., 1999; Sorensen et al., 2002). These results indicate that at least one additional gene that has a function similar to that of DYT1 might be present in Arabidopsis. Many bHLH proteins are known to form heterodimers with bHLH proteins (Heim et al., 2003; Toledo-Ortiz et al., 2003). Therefore, it is possible that another DYT1-like protein might be the partner of DYT1. Some bHLH proteins can interact with MYB proteins to form transcriptional complexes to regulate gene expression (Goff et al., 1992; Heim et al., 2003; Toledo-Ortiz et al., 2003). The anther highly expressed MYB proteins might be the good candidates for a DYT1 partner. For example MYB33 and MYB65, which have redundant
function in regulating the tapetum function during meiosis, and the double mutant anther shares some similar phenotypes to the dyt1 anther (Millar and Gubler, 2005). Yeast two-hybrid experiments can be performed to screen and identify putative DYT1 partners in future.

Overall, I have used three different approaches to understand anther development and meiosis in different aspects. Some future experiments mentioned in this chapter will hopefully shed more light on anther development and meiosis. Especially, functional studies of some novel genes assumedly involved in anther development/meiosis and putative DYT1 partners and target genes will attribute essentially to the genetic network of anther development/meiosis in Arabidopsis.
References:


EDUCATION

2002 – Present       Ph.D. Candidate Department of Biology, the Pennsylvania State University
                     University Park, PA 16802
1993 – 1997          B.S. Department of Biochemistry and Biophysics, College of Life
                     Sciences, Wuhan University, P. R. China

PUBLICATIONS


Wang G, Sun Y, Ma H. Characterization of SOLO DANCERS (SDS) in yeast and in *Arabidopsis*: evidence for SDS as a cyclin gene and the requirement of its N-terminal region for SDS normal function. (Manuscript in preparation)