UNIVERSITY OF NOTTINGHAM
SCHOOL OF BIOSCIENCES
DIVISION OF PLANT AND CROP SCIENCES

ARABIDOPSIS MS1 FUNCTIONS AS A HUB IN THE TRANSCRIPTIONAL REGULATORY NETWORK OF LATE TAPETUM DEVELOPMENT

By Wenzhe Yin, BSc

Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy, Dec 2016

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Professor Zoe A. Wilson, for giving me the opportunity taking this PhD project. Without her support and mentorship, it would not have been possible for me to finish my postgraduate study. I am also greatly indebted to Professor Dabing Zhang and Professor Wanqi Liang in Shanghai Jiao Tong University for giving me continuous support and advice on my study.

My sincere thanks also go to my assessor Professor Malcolm Bennett for taking the time to provide constructive critiques of my work.

I would also like to thank all the other members of Wilson Lab and Zhang lab for creating an inspiring and lovely working space.

To my family and friends, your contributions are hugely appreciated, not only during my PhD studies but at every stage of my life. Great thanks to Dr Alice Burgess and Dr Alison Ferguson for doing the proofreading of my thesis and giving helpful advice on my academic writing. The most special thanks go to my girlfriend Yangfan Gao, for always being encouraging and supportive.

Last but not the least, Great respects and thanks to the Internet, the Google search engine and the people who have the great spirit of sharing on the Internet. It would be impossible for me to gain and develop in-depth knowledge and expertise in sciences without the help of the all-powerful Internet!

ABSTRACT

The development of the pollen grains within the anther locule relies upon the nourishing and secretory properties of a tissue layer termed tapetum. The transition of the post-meiotic phase of tapetum development depends on the *MALE STERILITY 1*. In the *ms1* mutant tapetum development is arrested post-meiosis and lacks subsequent biological processes, such as biosynthesis and secretion of pollen wall/coat components and the tissue programmed cell death process. *MS1* exhibits a transient expression pattern, which is tightly regulated and critical for tapetum development and viable pollen formation. Therefore, understanding the genetic control of MS1 is key to uncover the regulation of post-meiotic tapetum development. During this project, three regulation levels of MS1 were studied: (i) transcriptional activation, (ii) auto-repression and (iii) post-translational proteolysis. Phylogenetic footprinting analysis and molecular promoter dissection was used to investigate the transcriptional control of expression and a distal upstream sequence (−2900 to −2066 bp) was found to be essential for the activation of MS1. Three evolutionarily conserved non-coding sequences (CNS), enriched with unusually long consensus motifs, and binding site combinations of MS1 upstream transcription factors (TFs) were found within the −2 kb MS1 upstream sequence. These may serve as essential cis-regulatory elements (CREs) for *MS1* expression. ChIP experiments were used to investigate MS1 autorepression; the MS1 protein was shown to bind to the second exon of its genomic locus and to repress its own expression. Post-translational proteolysis was investigated using a triple mutant of the MS1 interacting gene that encodes for an E3 ubiquitin ligase LRB1 and its two paralogs *LRB2* and *LRB3*; which exhibits a novel

tapetum phenotype that may be induced by altered removal of MS1 protein in the *lrb123* tapetum.

The MS1 protein possesses a Plant Homeotic Domain (PHD) finger and belongs to a plant-specific C-terminal PHD contained protein (CPCP) family. Although extensive research has been carried out on the tapetum regulation role of MS1, very little is known about the underlying molecular mechanism. A series *in-silico* comparative analysis of the CPCP sequences in this thesis found that this family originated from green algae. Besides the PHD, two evolutionarily conserved domains, termed MS1/MMD1 Associated Domain 1 (MAD1) and MAD2, were identified in the protein. Molecular modelling of the MS1 PHD domain predicted a histone reader role with high affinity to H3K4me2/3 histone peptides. Super-resolution STED confocal observation showed that subnuclear localisation of the MS1 protein is distinctive with canonical TFs and aggregates at rounded speckles that resemble Polycomb bodies. A meta-data-analysis of MS1 microarrays found that most MS1 immediately responding genes are repressed by MS1, which is on the contrary to the previously proposed activator role of MS1. MS1 may therefore be a unique plant-specific histone reader family protein that participates in gene repression as a co-repressor.

MYB99 has previously been hypothesised to be a direct target of *MS1*, regulating late tapetum development. Comprehensive phenotyping was carried out on two *MYB99* null alleles; however, no defects were identified, probably due to high function redundancy among the MYB family TFs. In addition, no evidence of direct activation by MS1 was observed by yeast one-hybrid and ChIP analysis. Interestingly, a PCD indicator gene was down-regulated in the *myb99* mutant, suggesting a tapetal PCD regulatory role for *MYB99*.

TABLE OF CONTENTS

Acknowledgements i

Abstract iii

Table of Contents vi

List of Figures ix

List of Tables xii

Abbreviations xiii

1.1.1 Developmental stages of Arabidopsis anther 3

1.1.2 Pollen wall formation 9

1.2.1 Developmental phase of Arabidopsis tapetum 12

1.2.2 Genetic regulation of tapetal cells. 15

1.3.1 Potential role of MS1 in anther development 24

1.3.2 Regulation of MS1 by tapetal transcription factors 30

1.3.3 Rice and barley orthologues of MS1 31

2.1.1 Plant general growth condition 34

2.1.2 Agrobacterium-mediated transformation by floral dipping 34

2.1.3 Germination, growth and antibiotic selection 35

2.2.1 Genomic DNA extraction from Arabidopsis 36

2.2.2 Standard PCR 37

2.2.3 Colony PCR 37

2.2.4 PCR Cloning of fragments 37

2.2.5 Agarose gel electrophoresis 38

2.2.6 DNA gel purification 38

2.3.1 Total RNA extraction 39

2.3.2 cDNA synthesis 39

2.3.3 Semi-quantitative PCR 39

2.3.4 Quantitative (real-time) PCR 40

2.4.1 DNA restriction digestion and DNA ligation 41

2.4.2 TOPO Clone and Gateway Recombination Cloning 41
2.4.3 E. coli transformation 42
2.4.4 Agrobacterium transformation 42
2.4.5 Plasmid DNA extraction and sequencing 43
2.5.1 Protein gel electrophoresis 43
2.5.2 Western blot 44
2.6.1 Scanning electron microscope 45
2.6.2 Confocal microscopy 45
2.6.3 Confocal image analysis 45
2.7.1 Homology detection 46
2.7.2 Molecular modelling 47
2.7.3 Microarray data analysis 47
2.7.4 Protein Domain Analysis 48
3.1.1 A contradictory expression pattern of MS1 54
3.1.2 Long non-coding RNAs 64
3.1.3 Summary and aims 65
3.2.1 Phylogenetic footprinting of the upstream sequence of MS1 67
3.2.2 Promoter dissection assay of the MS1 upstream sequence 76
3.2.3 Yeast one-hybrid assay of the binding of tapetum TFs to the MS1 genomic locus 78
3.2.4 ChIP experiment confirmed MS1 is binding to its second exon 80
3.2.5 Expression profiling of lncRNAs transcribed at the MS1 locus 84
3.2.6 A novel tapetum secretory structure was absent in the lrb123 mutant 85
3.3.1 Activation of MS1 expression needs a combination of CREs 88
3.3.2 MS1 auto-repress via binding to the gene of body of itself 92
3.3.3 LRBs might mediated the degradation of MS1 94
3.3.4 Summary and future work 95
4.1.1 PHD domain functions as histone readers 98
4.1.2 The evolutionary origin of the tapetum and MS1 102
4.1.3 Nuclear bodies and transcription in plant nuclei 104
4.1.4 Transcriptomes of the Arabidopsis Anther 110
4.2.1 Molecular phylogenetic study of MS1 protein family 113
4.2.2 Characterisation of The MS1 protein functional domains 116

4.2.3 The MS1 protein has a specified subnuclear localisation 121
4.2.4 Meta-analysis of the anther and MS1 microarrays 125
4.3.1 MS1 belongs to an ancient Viridiplantae-specific CPCP family 137
4.3.2 MS1 possesses two MADs and a possible H3K4me2/3 binding PHD finger 138
4.3.3 MS1 proteins localise at subnuclear discrete speckles resembling PcG bodies 140
4.3.4 A significant portion of MS1 IRGs are up-regulated in the ms1 mutant. 142
4.3.5 Summary and future work 144
5.2.1 T-DNA insertion mutants of MYB99 lack phenotype 149
5.2.2 Precise phenotyping of the myb99 mutant using SEM 150
5.2.3 Transcriptomic profiling of the myb99 anther 152
5.2.4 Yeast one-hybrid assay of the binding of tapetum TFs to the MYB99 locus. 155
5.2.5 ChIP experiment verifying if MS1 directly binds the MYB99 genomic locus 157
5.3.1 MYB99 may initiate the late tapetum PCD progress 159
5.3.2 MYB99 is an indirect downstream target of MS1 160
5.3.3 Summary and future work 162

References 175

LIST OF FIGURES

Figure 1.1 The structure of Arabidopsis stamen. 2

Figure 1.2 Schematic diagram for the model of Arabidopsis microsporangial cell layer development. Taken from Scott et al., (2004). 5

Figure 1.3 Wild-type Arabidopsis thaliana anther development from Stage 1 to 8. 7

Figure 1.4 Wild-type Arabidopsis thaliana anther development from Stage 9 to 14. 8

Figure 1.5 Surface structure and cross-section morphology of the mature Arabidopsis pollen wall. 11

Figure 1.6 A gene regulation network of the Arabidopsis tapetum. Adapted from Gómez et al., (2015) 21

Figure 1.7 Transmission Electron Micrographs of Cross Sections through Wild-Type (A, C and E) and ms1 Mutant Anthers (B, D and F). 26

Figure 1.8 Transmission electron micrographs of cross-sections through microspore 27

Figure 3.1 Expression patterns of MS1 from various sources. 55

Figure 3.2 The expression pattern of MS1/PTC1 alleles and the MS1 gene map (Yang et al., 2007; Li et al., 2011). 63

Figure 3.3 Genomic view of the four MS1 lncRNAs transcribed from the genome locus of MS1. 65

Figure 3.4 Comparative analysis of promoter regions of MS1 orthologs 69

Figure 3.5 The overall schematic view of positions of the MEME analysed motifs and PlantPAN identified TFBSs within the MS1 distal CNS and two core CNSs. 71

Figure 3.6 The core CNS1 of MS1 contains four consensus motifs generated by the online application of MEME 4.11.2. 73

Figure 3.7 The core CNS2 of MS1 contains three consensus motifs generated by the online application of MEME 4.11.2. 73

Figure 3.8 The distal CNS of MS1 contains four consensus motifs generated by the online application of MEME 4.11.2. 74

Figure 3.9 Motif 2 of MS1 core CNS1 contains a consensus MYB binding sequence and a canonical bHLH binding E-Box. 74

Figure 3.10 Motif 2 of MS1 core CNS2 contains a consensus DYT1 binding sequence and a canonical bHLH E-box. 75

Figure 3.11 Motif 3 of MS1 core CNS1 contains a consensus sequence of GAGA-repeats. 75

Figure 3.12 Schematic diagram showing truncated versions of the MS1 promoter in the promoter dissection assay. 77

Figure 3.13 The yeast one hybrid indicated no interaction persists between selected tapetal TFs and the MS1 genomic sequence. 79

Figure 3.14 A western blot analysis of isolated chromatin for MS1-GFP ChIP. 82

Figure 3.15 MS1 ChIP enrichment at the MS1 locus. 83

Figure 3.16 Semi-quantitative RT-PCR of four MS1 lncRNAs in flower buds and leaf tissues. 85

Figure 3.17 SEM images showing the pollen wall and tapetum cells of the lrb123 mutant and the WT. 87

Figure 3.18 Proposed model of the three levels of regulation of MS1. 97

Figure 4.1 Simplified phylogeny of tapetum and spore/pollen wall evolution. 103

Figure 4.2 A landscape of the diversity of interphase mammalian nuclear bodies. 105

Figure 4.3 Phylogenetic tree of the C-terminal PHD proteins of selected representative species of Chlorophyta and Embryophyta. 115

Figure 4.4 Comparative study and molecular modelling of MS1 PHD finger. 119

Figure 4.5 PRALINE alignment of MS1 and the related Arabidopsis proteins and the graphical representation of MS1 protein. 120

Figure 4.6 STED confocal images of the subnuclear distributions of the MS1-GFP and TDR1-mTFP1 fusion reporter. 124

Figure 4.7 The regulatory pathway of selected Arabidopsis tapetum key TFs. 125

Figure 4.8 Venn diagram dissections and biological process GO analysis of the 2-fold DEGs of MS1, MYB80 and TEK. 127

Figure 4.9 Microarray Clustering and heat map analysis of up-regulated IRGs and two proposed paradigmatic expression patterns. 129

Figure 4.10 Microarray Clustering and heat map analysis of the down-regulated Stage8 IRGs and a proposed paradigmatic expression patterns. 131

Figure 4.11 Microarray Clustering and heat map analysis of the down-regulated St.11 IRGs and a proposed paradigmatic expression patterns. 134

Figure 5.1 Time-course induction analysis of the six MS1 downstream targets using DEX-inducible MS1pro:MS1-GR/ms1-1 transgenic plant. 148

Figure 5.2 The SEM observation of the anthers of the WT Arabidopsis and the ms1 and myb99 mutant. 152

Figure 5.3 The quantitative RT-PCR of four MS1 immediately responding genes in the anther of the myb99-1 mutant and the WT. 154

Figure 5.4 The yeast one hybrid indicated no interaction persists between selected tapetal TFs and the MYB99 genomic sequence. 156

Figure 5.5 No MS1 ChIP enrichment was detected in the MYB99 genomic locus. .158

LIST OF TABLES

Table 1.1 Summary of the major events for stages of Arabidopsis thaliana anther development. 6

Table 1.2 Transcription factors and cell receptors involved in Arabidopsis anther development. (Parish and Li, 2010) 22

Table 1.3 The mutant alleles of MALE STERILITY1 in different Arabidopsis ecotype backgrounds. 29

Table 2.1 The ssp primer list for RT-PCR detection of the MS1 lncRNAs. 52

Table 4.1 Selected genes in clustered groups of each of the ms1 IRG sets. 135

Table 5.1 The growth conditions for the detection of myb99 conditional phenotypes. 150

ABBREVIATIONS

AMS	ABORTED MICROSPORES
A	archesporial cell
AG	AGAMOUS
AGP	ARABINOGALACTAN PROTEIN
AL	Alfin1-like
AtMYB33	Arabidopsis thaliana MYB DOMAIN PROTEIN 33
AtMYB65	Arabidopsis thaliana MYB DOMAIN PROTEIN 65
BES1	BRI1-EMS-SUPPRESSOR 1
bHLH010	basic helix-loop helix transcription factor 10
bHLH089	basic helix-loop helix transcription factor 89
bHLH091	basic helix-loop helix transcription factor 91
BPC	BASIC PENTACYSTEINE
BR	brassinosteroid
BRRE	BR response element
BTB	Bric-a-Brack/Tramtrack/ Broad
BZR1	BRASSINAZOLE-RESISTANT 1
CaMV 35S	cauliflower mosaic virus 35S
CAP-D3	Chromosome Associated Protein D3
CB	Cajal body
cDNA	complementary DNA
CDS	coding DNA sequence
ChIP	Chromatin Immunoprecipitation
CNS	conserved non-coding sequence
CPCP	C-terminal PHD contained protein
CRE	cis-regulatory element
CYP703A2	CYTOCHROME P450, FAMILY 703, SUBFAMILY A, POLYPEPTIDE

	2
DCL1	DICER-LIKE1
DEG	differentially expressed gene
DEX	dexamethasone
DNA	deoxyribonucleic acid
DYT1	DYSFUNCTIONAL TAPETUM 1
E	epidermis
EMSA	electrophoretic mobility shift assay
EN	endothecium
ER	endoplasmic reticulum
FP	fluorescent protein
GA	gibberellin
GFP	green fluorescent protein
GPI	glycosylphosphatidylinositol
GUS	β-glucuronidase
HYL1	HYPONASTIC LEAVES1
IGC	interchromatin granule cluster
IRGs	immediately responding genes
LB	Lysogeny broth
LCM	laser capture microdissection
LHP1	LIKE HETEROCHROMATIN PROTEIN 1
lncRNA	long non-coding RNA
LTP	lipid transfer protein
ML	middle layer
MMC	microspore mother cell
MMD1	MALE MEIOCYTE DEATH1
MS188/MYB80	MALE STERILE 188/MYB DOMAIN PROTEIN 80
mTFP1	monomeric teal fluorescent protein 1
NB	nuclear body
NCB1	NO CAJAL BODIES 1
NGS	next generation sequencing

OD	optical density
ORF	open reading frame
OsVIL2	rice VIN3-like2
PAL	PHD-associated AL domain
PCD	programmed cell death
PcG	polycomb group
PCR	polymerase chain reaction
PHD	plant homeodomain finger
PMI	pollen mitosis I
PMII	pollen mitosis II
POI	protein of interest
PPC	primary parietal cell
PPI	protein-to-protein interaction
PRC1	polycomb repressive complexes 1
PRC2	polycomb repressive complexes 2
PSC	primary sporogenous cell
PTM	posttranslational modification
qPCR	Real-time polymerase chain reaction
RNA	ribonucleic acid
RNAPII	RNA polymerases II
RNA-Seq	RNA sequencing
RNP	Ribonucleoprotein
RT-PCR	Reverse transcription polymerase chain reaction
SD	Synthetic Defined
SEM	scanning electron micrographs
SFC	splicing factor compartment
SNP	single nucleotide polymorphism
SPL/NZZ	SPOROCYTELESS/NOZZLE
STED	stimulated emission depletion microscopy
T	tapetum

TDF1/MYB35	DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION 1/ MYB DOMAIN PROTEIN 35
TDM1	THREE DIVISIONMUTANT 1
TEK	TRANSPOSABLE ELEMENT SILENCING VIA AT-HOOK
TEM	transmission electron micrographs
TF	transcription factor
TFBS	transcription factor binding site
TSS	transcription start site
WRKY43	WRKY DNA-BINDING PROTEIN 43
WT	wild type

CHAPTER 1 INTRODUCTION

Sexual Reproduction in plants was first detected in 1694, by the German botanist Camerarius Rudolph Jacob (Willemse, 2008). Sixty-nine years later, German botanist Joseph Gottlieb Kölreuter, made the first observation of male sterility (MS) in plants (Mayr, 1986). Together, these observations have revolutionised the study of plant reproduction and MS. Since then, MS has been reported in more than 610 plant species. Both manually induced and naturally occurring male sterile plants provide agriculturists with powerful breeding tools to harness heterosis (hybrid vigour). For example, the use of heterosis for rice and wheat crops can offer 20% to over 50% yield increases (Tester and Langridge, 2010). The use of MS plants increases the efficiency and reduces the need for manual labour to create hybrid crop plants with heterosis. Therefore, understanding the genetic and molecular mechanisms of MS has long been the interest of botanists (Raghavan, 1988; Bedinger, 1992; Ma, 2005; Zhang and Wilson, 2009; Ariizumi and Toriyama, 2011; Walbot and Egger, 2016; Shi et al., 2015). In the plant, MS refers to the failure to produce male organs (stamens), functional pollen, dehiscent anthers or viable male gametes, and barriers other than incompatibility preventing pollen from reaching the female ovule (Chen and Liu, 2014). Over the last few decades, a large number of genes causing sporogenous MS were identified through male sterile mutant screens in the model organisms, including *Arabidopsis thaliana*. A remarkable proportion of these genes regulate the development of the inner most somatic layer of the anther wall, called the tapetum, which is essential for viable pollen formation (Wilson and Zhang, 2009). This chapter will provide an overview of the developmental process and the mechanisms of genetic

regulation of the tapetum, and raise the scientific questions, which this study intends to answer.

1.1 ANTHER DEVELOPMENT IN ARABIDOPSIS

An *Arabidopsis* flower has four long and two short stamens, each of which consists of two integral parts: the anther and the filament (**Figure 1.1 A**). The structure of filaments is simple; they are single stalk-like tubes of vascular tissues, which anchor anthers to the base of the flower and transmit water and nutrients to it. Whereas the anther structure is more complex. The anther has four lobes, which are bilaterally symmetrical. In each lobe, a microsporangium (pollen sac) develops at a given stage, offering a chamber (locule) where pollen grains develop. Each of the four locules is surrounded by four distinct, concentric layers of cells, from inner to outer; the tapetum, the middle layer, the endothecium and the epidermis (**Figure 1.1 B**).

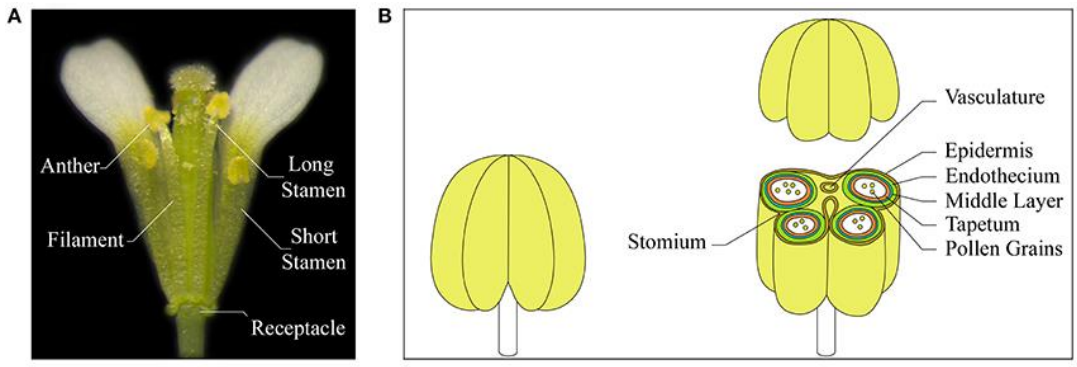


Figure 1.1 The structure of Arabidopsis stamen.

(A) An anthesis flower showing short and long stamens, with all sepals and two petals removed. (B) Cartoon of a post-meiotic stamen and the anther transverse section, with differentiated tissues, indicated in different colours. Adapted from (Cardarelli and Cecchetti, 2014)

1.1.1 Developmental stages of *Arabidopsis* anther

Due to the importance of anther development in MS, the development of *Arabidopsis* anthers has been thoroughly described and divided into 15 developmental stages. These developmental stages are distinguished by morphological and cellular differences during anther development (Scott et al., 2004; Sanders et al., 1999). The major events (**Table 1.1**) and transverse semi-thin section images for each anther developmental stage (**Figure 1.3** and **Figure 1.4**) are well summarised and illustrated by Sanders et al., (1999). At stage 1 (Stage 1), the floral meristem is composed of three histogenic cell layers: L1 (epidermis), L2 (subepidermis) and L3 (core). The stamen (anther) primordium initiates within the L2 (**Figure 1.3 Stage 1**). From Stage 2 to 4, the anticlinal cell division of L1 results in expanded epidermis (E) of the anther primordium, while the cell division and differentiation within L3 layer form the connective and vascular tissues. The L2 cells cluster within the four anther primordia corners to give rise to the primary parietal cell (PPC) layer and the primary sporogenous cells (PSC) via a series of periclinal and anticlinal divisions. Cells of the PPC layer then differentiate into the endothecium (En), the middle layer (ML) and the tapetum (T) by the end of Stage 4, while the PSC differentiate into microspore mother cells (MMCs) (**Figure 1.2 and Figure 1.3 Stage 2-4**). (Canales et al., 2002; Sanders et al., 1999). At Stage 5, the anther development completes its morphogenesis, resulting in a bilaterally symmetrical four-locule morphology. In each locule, MMCs are surrounded by four non-reproductive layers (E, En, ML, T), which are the somatic cell layers (**Figure 1.3 Stage5).** MMCs undergo meiosis between Stage 6 and 7 and this generates tetrads of haploid microspores, which are covered by a layer of callose wall. In the tapetum, these two stages become vacuolated while the middle layer

begins to degenerate (**Figure 1.3 Stage 6 and 7**). At Stage 8, meiosis is completed, and the microspores are released from the tetrads along with the degradation of the callose wall (**Figure 1.3 Stage 8**). From Stage 9 to 10, the microspores go through a general increase in size, a dramatic pollen wall expansion, nucleus polarisation and cytoplasmic vacuolation. The tapetum at this stage degrades and becomes less vacuolated and more densely stained (**Figure 1.4 Stage 9 and 10**). Tricellular pollen grains form after the first asymmetric mitosis and the second symmetric mitosis of microspore at Stage 11 and 12 respectively. After the first mitosis, a uninucleate microspore converts into a bicellular pollen with a larger vegetative cell and a smaller generative cell. Then the generative cell undergoes the second mitosis to form two sperm cells, which are for the plant-specific double fertilisation. At Stage 12, the tapetum cells completely degenerate into cellular debris, which is released into the space of the locules (**Figure 1.4 Stage 11 and 12**). At Stage 13, anther dehiscence occurs, followed by the dispersal of mature pollen grains from the locule. The senescence of stamen and shrinkage of anther cells occurs at Stage 14. At the final Stage 15, the stamen falls off senescing flower (**Figure 1.4 St 13-15**) (Ma, 2005; Sanders et al., 1999).

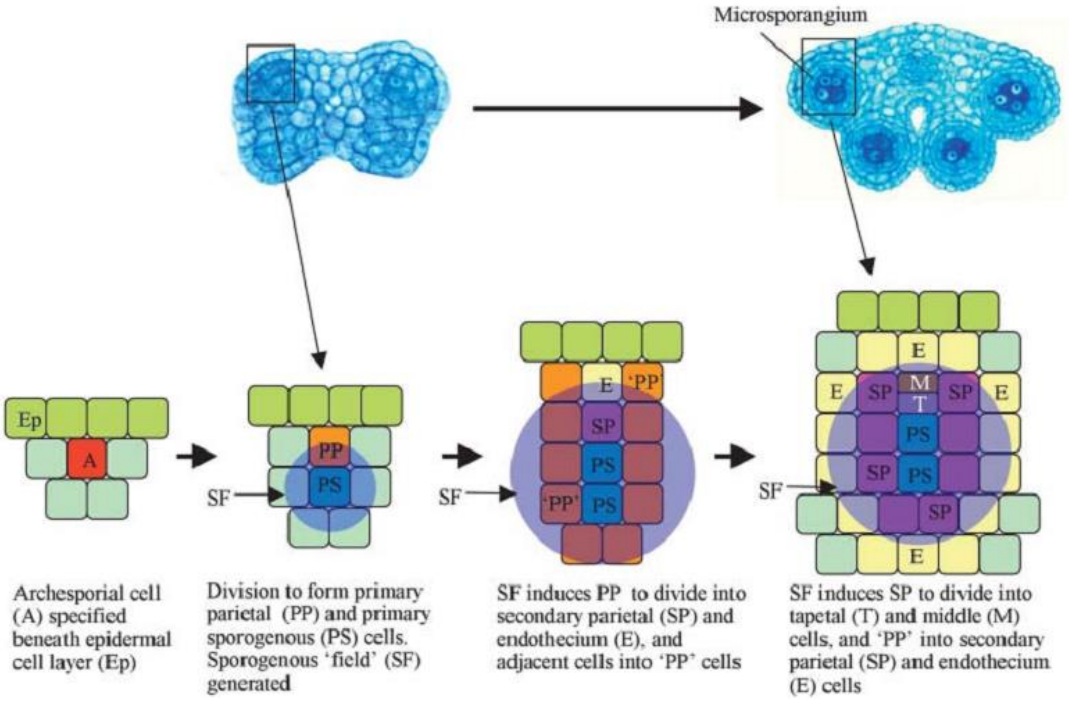


Figure 1.2 Schematic diagram for the model of Arabidopsis microsporangial cell layer development. Taken from Scott et al., (2004).

Table 1.1 Summary of the major events for stages of Arabidopsis thaliana anther development.

Anther stage	Major events and morphological markers[b]	Tissues present[c]	Flower stage[d]	Pollen stage[e]
1	Rounded stamen primordia emerge	L1, L2, L3	5	
2	Archesporial cells arise in four "corners" of L2 layer. Change in shape of primordia to more oval.	E, Ar		
3	Four regions of mitotic activity. 1° parietal and 1° sporogenous layers derived from archesporial cells. Further divisions of each layer generate the 2° parietal layers and sporogenous cells, respectively.	E, 2°P, Sp	7	
4	Four-lobed anther pattern with two developing stomium regions ("notch") generated. Vascular region initiated.	E, En, ML, T, Sp, C, V	8	
5	Four clearly defined locules established. All anther cell types present and pattern of anther defined. Microspore mother cells appear.	E, En, ML, T, MMC, C, V	9	3
6	Microspore mother cells enter meiosis. Middle layer is crushed and degenerates. Tapetum becomes vacuolated and the anther undergoes a general increase in size.	E, En, ML, T, MC, C, V		
7	Meiosis completed. Tetrads of micropsores free within each locule. Remnants of middle layer present.	E, En, ML, T, Tds, C, V		4
8	Callose wall surrounding tetrads degenerates and individual microspores released.	E, En, T, MSp, C, V	10	5
9	Growth and expansion of anther continue. Microspores generate an exine wall and become vacuolated. Septum cells can be distinguished at the level of the TEM. [f]	E, En, T, MSp, C, V, Sm		6–7
10	Tapetum degeneration initiated.	E, En, T, MSp, C, V, Sm	11–12	
11	Pollen mitotic divisions occur. Tapetum degenerates. Expansion of endothecial layer. Secondary thickenings or "fibrous bands" appear in endothecium and connective cells. Septum cell degeneration intiated. Stomium differentiation begins.	E, En, T, PG, C, V, Sm, St		8–9
12	Anther contains tricellular pollen grains. Anther becomes bilocular after degeneration and breakage of septum below stomium. Differentiated stomium seen in TEM. [f]	E, En, PG, C, V, St		10
13	Dehiscence. Breakage along stomium and pollen release.	E, En, PG, C, V	13–14	
14	Senescence of stamen. Shrinkage of cells and anther structure.	E, En, C, V	15–16	
15	Stamen falls off senescing flower.		17	

Adapted from Sanders et al, (1999) (a). The flower developmental staging was based on the work of Smyth et al. (1990) and Bowman et al. (1991b). Pollen developmental stages were based on the work of Regan and Moffatt (1990). (b). The differentiation of the cell-types within each locule of the anther was not synchronized during Stage 1 to 4 (**Figure 1.3**). From Stage 5 onwards development of locule cell types was consistent within an anther (**Figure 1.3** and **Figure 1.4**). (c). Ar, archesporial; C, connective; E, epidermis; En, endothecium; L1, L2, and L3, the three cell layers of the stamen primordia; MC, meiocyte; ML, middle layer; MMC, microspore mother cell; MSp, microspore; 2°P, secondary parietal layer; PG, pollen grains; Sm, septum; Sp, sporogenous cells; St, stomium; T, tapetum; Tds, tetrads; V, vascular.

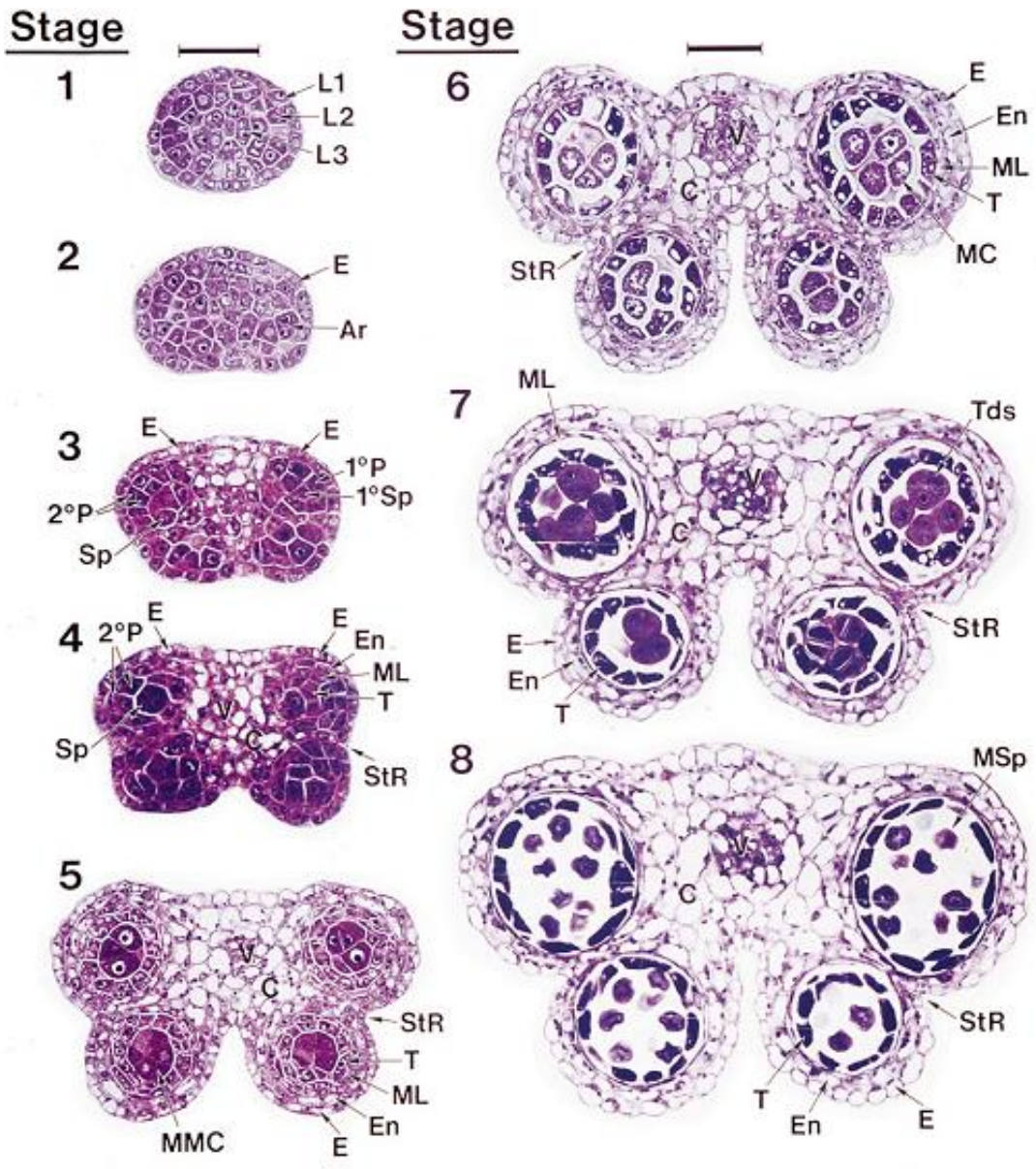


Figure 1.3 Wild-type *Arabidopsis thaliana* anther development from Stage 1 to 8.

Developmental stage based semithin transverse sections of resin embedded *Arabidopsis* anthers. Stain: toluidine blue. Ar, archesporial cell; C, connective; E, epidermis; En, endothecium; L1, L2, and L3, the three cell-layers in stamen primordia; MC, meiotic cell; ML, middle layer; MMC, microspore mother cells; MSp, microspores; 1°P, primary parietal layer; 2°P, secondary parietal cell layers; 1°Sp, primary sporogenous layer; Sp, sporogenous cells; StR, stomium region; T, tapetum; Tds, tetrads; V, vascular region. Scale bar is 25μm. Figure is taken from Sanders et al., (1999)

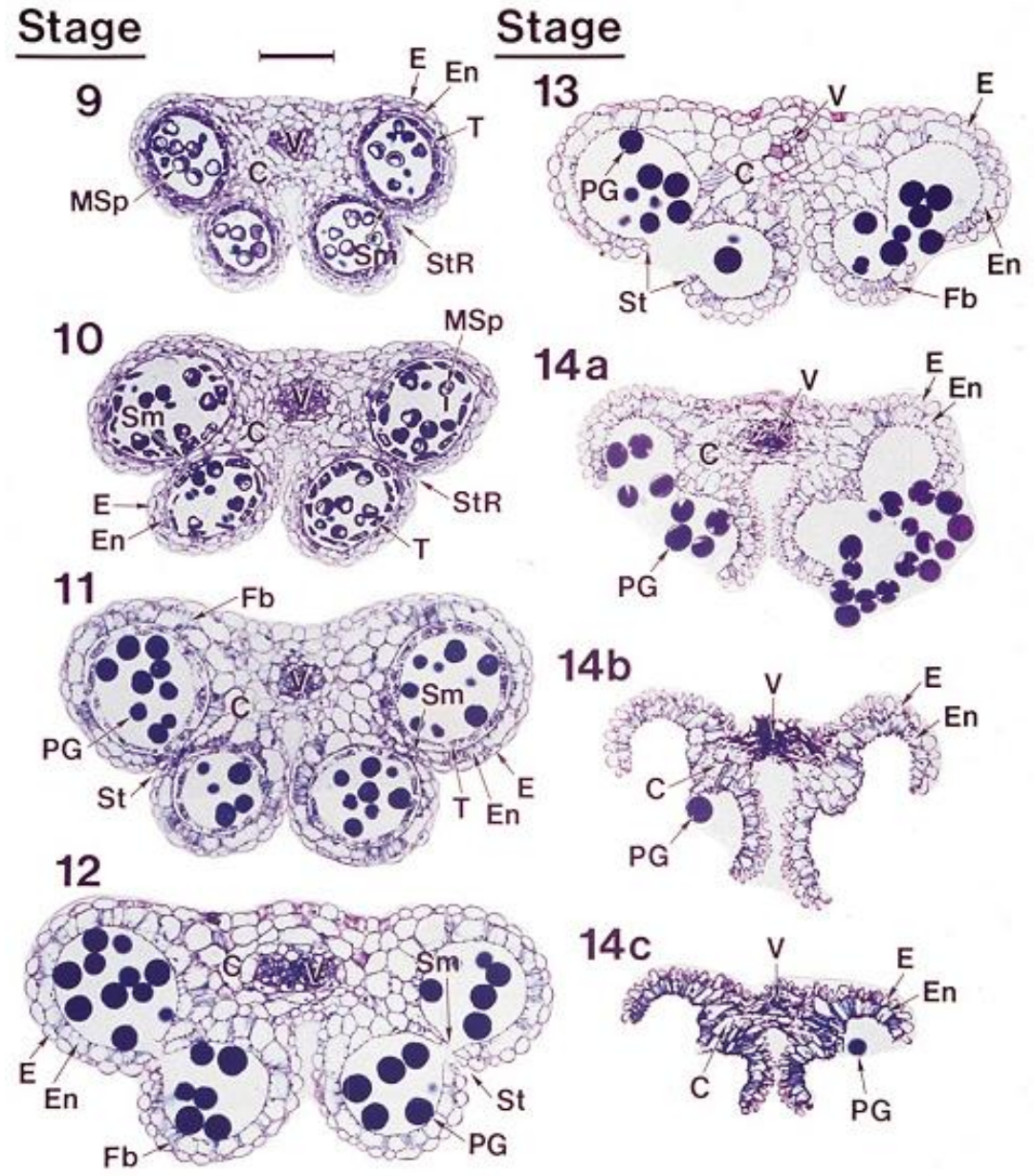


Figure 1.4 Wild-type *Arabidopsis thaliana* anther development from Stage 9 to 14.

Developmental stage based semithin transverse sections of resin embedded Arabidopsis anthers. Stain: toluidine blue. C, connective; E, epidermis; En, endothecium; Fb, fibrous bands; MSp, microspores; PG, pollen grains; Sm, septum; St, stomium; StR, stomium region; T, tapetum; V, vascular region. Scale bar is 50µm for Stage 9 to 14. Figure is taken from Sanders et al., (1999).

1.1.2 Pollen wall formation

Pollen grains play a vital role in plant reproduction: a pollen grain is a vector delivering two sperm cells to the female reproductive organ for double fertilisation. To protect pollen from harsh conditions, flowering plants evolved a highly-specialised pollen encasement, consisting of an inner wall (intine), and an outer wall (exine) which contains a chemically non-reactive and highly physically stable biopolymer called sporopollenin (**Figure 1.5 A and B**) (Ariizumi and Toriyama, 2011). In many species, the exine is intricately decorated, showing a species-specific sculpturing (shape and structure). In *Arabidopsis thaliana*, the exine shows a regular hexagonal pattern with a radially directed rod structure, termed bacula, and an external tangentially orientated edge structure, termed tectum (**Figure 1.5 C and D**). Beyond this outer wall, additional layers of a hydrophobic mix of compounds, termed pollen coat or tryphine, cover both the surfaces and within the chambers of the patterned sporopollenin framework (**Figure 1.5 D**) (Quilichini et al., 2014b).

Pollen wall formation relies heavily on the somatic tissues of the anther, especially the tapetum (Gómez et al., 2015). Early in microspore development, callose deposits between the primary cell wall and the plasma membrane of the MMCs, forming a transient callose wall, which subsequently becomes very thick and fully covers the meiocytes or microspores in the tetrads. The callose is a permeable layer allowing sporopollenin precursors secreted by the tapetum to deposit on the membrane surface of microspores to form primexine. The growth of this primary primexine pollen wall layer is relatively slow (Renzaglia et al., 2014). The primexine structure is regularly established, consisting of undulated microspore plasma membrane and the primexine matrix, which contains probaculae and protecta situated at regular intervals

(Nishikawa et al., 2005). The primexine is essential for the deposition and sculptured patterning of sporopollenin over the following course of pollen development. After meiosis, the callose wall is dissolved by β-1,3 glucanase secreted by tapetum, resulting in the release of microspores from tetrads (Lu et al., 2014). In a subsequent uninucleate microspore stage (Stage 8-10), exine rapidly establishes. The sporopollenin precursors secreted by the tapetum percolate the fluid of the locules, and polymerise on the pollen surface, guided by the sites of probaculae and protecta. When the exine formation approaches the end, the primexine recedes and disappears. Together, the formation and dissolution of callose wall, primexine establishment and microspore plasma membrane undulation are all essential for the sporopollenin anchoring and exine patterning of pollen wall (Ariizumi and Toriyama, 2011).

In the tricellular pollen stage (Stage 11 and 12), an inner intine pollen wall, composed of pectin, cellulose, and hemicellulose, appears to be visible between the exine and pollen plasma membrane (**Figure 1.5 C and D**) (Blackmore et al., 2007). Following the tapetal cell breakdown (Stage 12), a mixture of waxes, lipids, sterols, sugars, and proteins derived from the tapetum, deposit on the outermost surface of the pollen grain, forming the pollen coat (Fellenberg and Vogt, 2015).

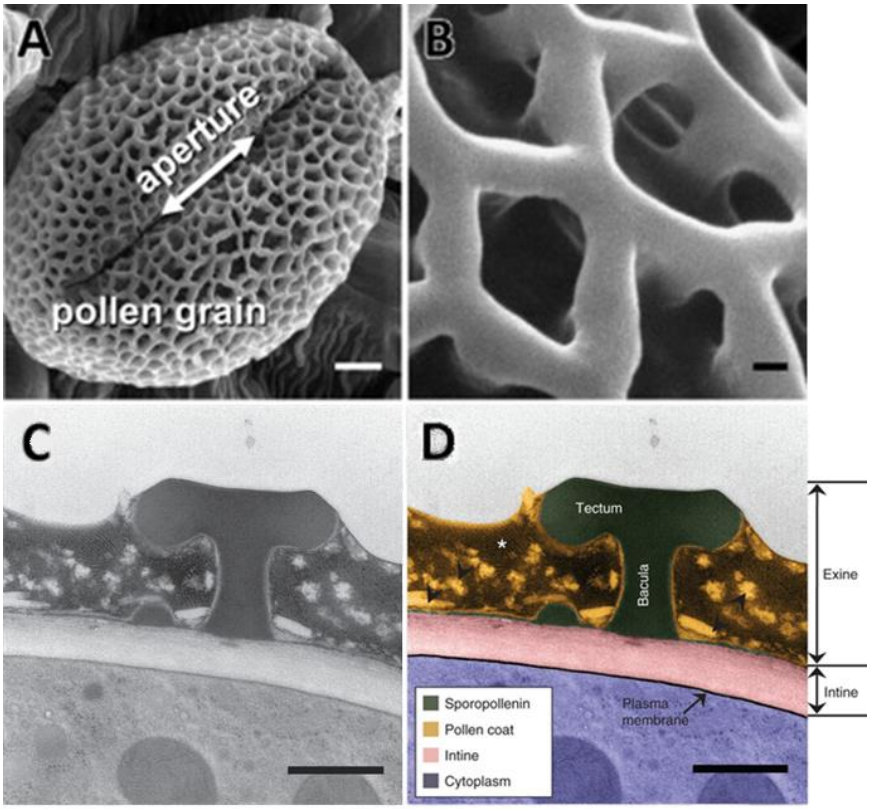


Figure 1.5 Surface structure and cross-section morphology of the mature *Arabidopsis* pollen wall.

Scanning electron micrographs (A, B) and transmission electron micrographs (C, D) of a wild-type pollen grain in cross-section. Bars = 2 μm (A, B) and 500 nm (C, D). (D) False coloured micrograph. Adapted from (Quilichini et al., 2014a, 2014b).

1.2 ARABIDOPSIS TAPETAL CELL

The tapetum is the innermost somatic layer of the anther wall, directly surrounding the developing reproductive cells, and plays a vital role in the development of pollen grains (Parish and Li, 2010). Although tapetal cells derive from the same parental cells as microspores in the developing anther, they undergo a strikingly different differentiation program and show a distinct cellular morphology. The tapetal cell is binucleate, strongly polarised and only partially enclosed with the cell wall.

Furthermore, the late tapetum also goes through Programmed Cell Death (PCD) and eventually degrades (Wallace et al., 2011).

Although tapetum is an ephemeral tissue, its function is crucial for land plant reproduction. To avoid the use of water as the medium through which the male gametes are conveyed, land plants need to nourish and strengthen spores or pollen grains in the liquid-filled cavity within microsporangia or anthers before they are released into the environment. To protect the male gametes, land plants have developed the protective structure of the spore/pollen wall. The function of the tapetum is closely linked with spore/pollen wall formation, which indicates that the tapetum is a very ancient nurse tissue, whose evolution can date back to the emergence of land plants. The simplest type of tapetum recognised, with its cell wall maintained, is found in true mosses (Bryopsida) (Blackmore and Knox, 2016). While two principal types of tapetum are found in seed plants (Spermatophyta), namely secretory tapetum and amoeboid tapetum. *Arabidopsis* has a secretory type of tapetum, which has a short lifespan but a conserved three-phase development (Liu and Fan, 2013).

1.2.1 Developmental phase of Arabidopsis tapetum

Tapetum development can be divided into three phases based on the cellular and morphological differences. As the three phase transitions respectively coincide with the onset and accomplishment of the microspore meiosis, they are termed as the early (pre-meiotic), middle (meiotic) and late (post-meiotic) phases (Owen and Makaroff, 1995).

During the early stages of archesporial cell differentiation, tapetal cell fate is determined by signalling between reproductive and non-reproductive cells (Sun et al., 2007; Yang et al., 2003). After the cell fate determination, newly specified tapetum cells enter the early phase. The tapetum in this phase has a thin cellulosic cell wall, which is often permeated by plasmodesmata, connecting the tapetal cell with surrounding tapetal and sporogenous cells (Quilichini et al., 2014a). The tapetal cell organelles, such as mitochondria, Golgi apparatus, vesicles, plastids and granular endoplasmic reticulum with ribosomes, could be clearly identified (Echlin and Godwin, 1969). When the locule of the anther appears within the four corners of the anther, the tapetum cells produce locular fluid and nourish the MMCs.

As soon as microspore meiosis begins, the tapetum transits into next phase, the meiotic phase. The tapetal cells first go through mitosis, but without cytokinesis, resulting in the characteristic binucleate cells. At the same time, the plasmodesmatal links between the tapetum and the microspores are severed by callose walls which enclose microspores (Ünal et al., 2013; Renzaglia et al., 2014). In tapetal cells, a significant vacuolation occurs while the organelles undergo a remarkable enhancement both in population and size, migrating towards the inner tangential plasma membrane, which makes the tapetal cells polarised during this phase (Smyth, 2001). Subsequently, the tapetal inner tangential walls and radial walls gradually break down, and an aggregation of Golgi vesicles and ER cisternae shows at the locule side of the cell, which indicates a high metabolic and secretory activity of the sporopollenin precursors (Quilichini et al., 2014b). Consistent with the cellular morphology, the genes involved in the biosynthesis and transportation of sporopollenin are actively expressed in this phase of tapetum development (Li et al., 2015).

After the accomplishment of microspore meiosis, the release of microspores from tetrads is noticeably synchronised with the tapetal middle-to-late phase transition. After the degradation of callose wall by the enzyme mixture secreted by the tapetum, sibling microspores are released into anther locules (Lu et al., 2014). As a consequence, microspores can get direct contact with the sporopollenin precursors, resulting in a dramatic expansion of pollen exine wall (Hu et al., 2014). The tapetum in the late phase is still metabolically active, secreting substances into the locule by diffusion or exocytosis. However, the cellular morphology of the tapetum begins to change. Cells become less vacuolated compared to theprevious phase, with more dense organelles accumulating (Quilichini et al., 2014a). At the more advanced post-meiotic phase, the tapetum initiates programmed cell death (Parish and Li, 2010), and the tapetal sporopollenin biosynthetic activity gradually quenches, instead, the biosynthetic activity of pollen coat components dominates (Ischebeck, 2016). As the pollen approaches maturity, the shape of tapetal cells is no longer maintained and the cytoplasm is disorganised, with an increased vacuolation and progressive disappearance of mitochondria, plastids, and nuclear envelopes (Shukla et al., 1998). On the contrary, two types of highly specialised tapetum organelles, tapetosomes and elaioplasts, remain numerous and intact in the senescing tapetal cells (Rejón et al., 2016). Eventually, the tapetum is completely degenerated, with all of the tapetal cell remnants released into the locules, and integrated into the pollen wall as pollen coat material (Papini et al., 1999).

In brief, the tapetum goes through three phases of development, which is consistent with its roles in different pollen developmental stages.

1.2.2 Genetic regulation of tapetal cells.

The entire course of tapetal development is completed in a relatively short time period. However, the regulation of the phase transition timing is strictly coordinated and well-synchronized with the microspores. This synchrony is crucial for the formation of viable mature pollen grains (Gómez et al., 2015; Shi et al., 2015; Cardarelli and Cecchetti, 2014; Zhang and Yang, 2014). For example, when β-1,3 glucanase secretion is deregulated, anthers fail to produce viable pollen grains (Hird et al., 1993), and a delay of tapetal PCD initiation always leads to male sterility (Feng and Dickinson, 2010).

In organisms, transcriptional factors play a main role in exerting rapid and precise spatial-temporal regulation of biological processes (Pires and Dolan, 2010). Since the tapetal developmental program is tightly regulated, transcription factors undoubtedly play an important role. Several transcription factors involved in *Arabidopsis* tapetal development have been uncovered (Zhang and Yang, 2014; Liu and Fan, 2013; Gómez et al., 2015).

In early tapetum development (early phase), *AGAMOUS (AG)* and *SPOROCYTELESS/NOZZLE (SPL/NZZ)* are the only reported transcription factors that regulate the development of tapetum. *AGAMOUS (AG)* encodes a MADS-box family necessary for the initiation of cell division and differentiation of stamen and carpel. *AG* expression initiates at the central region of the floral meristem, and is later restricted to developing stamens and carpels (Bowman et al., 1991a). In anthers, *AG* expression only occurs in the tapetum and connective tissues. In developing pollen grains, no *AG* mRNA is detected (Sieburth, 1997). *AG* directly activates the

expression of *SPOROCYTELESS/NOZZLE (SPL/NZZ)*, which can maintain a continuous expression without *AG* (Ito et al., 2004), indicating that *SPL/NZZ* and *AG* may form a regulatory feedback loop (Wijeratne et al., 2007). The *SPL/NZZ* gene encodes a putative novel transcription factor. The *SPL* first expresses in the anther primordium, with prolonged expression in all newly differentiated anther tissues, suggesting that it acts in the somatic cell fate specification (Chang et al., 2011).

During the middle (meiotic) phase tapetal development, various groups of transcription factors responsible for tapetum development were extensively studied, in particular the MYB and bHLH transcription factor families, revealing their critical roles in regulating various aspects of tapetum development.

The *AtMYB33* and *AtMYB65* encode two redundant R2R3 MYB domain transcription factors, which have similar expression patterns and functions in anthers (Millar and Gubler, 2005; Allen et al., 2007). This redundant gene pair is induced by the phytohormone Gibberellin acid (GA), and negatively regulated by two microRNAs (miRNAs) from the family of miR159 (Tsuji et al., 2006). *AtMYB33* and *AtMYB65* are together responsible for a vast array of GA-induced gene expression in anthers (Tsuji et al., 2006; Aya et al., 2009). The double *myb33/myb65* mutant has a hypertrophied tapetum, which expands to engulf the entire locule, resulting in microspore collapse (Kaneko et al., 2004; Millar and Gubler, 2005). Moreover, callose degradation is defective in the mutant, and therefore the tapetum fails to secret β-1,3 glucanase. *AtMYB33* and *AtMYB65* are expressed normally in the *spl/nzz* mutant, suggesting they are under a different regulation pathway (Gocal et al., 2001; Millar and Gubler, 2005).

The *DYSFUNCTIONAL TAPETUM 1 (DYT1)* encodes a putative basic helix-loop-helix (bHLH) transcription factor, which is essential for normal

development of the tapetum (Zhang et al., 2006a). In the *dyt1* mutant, tapetal cells go through a severe premature vacuolation, resulting in an arrest in the meiotic nuclear division, which eventually leads to the absence of pollen grains (Zhang et al., 2006a). *DYT1* acts genetically downstream of both *AG* and *SPL/NZZ* (Schiefthaler et al., 1999; Yang et al., 1999; Du et al., 2009), but upstream of more than 20 genes encoding transcription factors. These transcription factors include most reported key tapetal transcription factors, *i.e. AMS*, *MS1*, *TDF1/MYB35*, *TEK* and *MYB80*. As *DYT1* expresses in the tapetum at the onset of meiotic phase and dominates the tapetal cell differentiation, absence of DYT1 causes a failure for the tapetum to differentiate into a functional meiotic phase or a subsequent post-meiotic phase tapetum. Hence, more than 1000 genes are down-regulated in *dyt1* mutant anthers compared with that of the wild-type (Feng et al., 2012; Zhang et al., 2006a).

Three redundant bHLH transcription factors, *bHLH010*, *bHLH089*, and *bHLH091*, act downstream of *DYT1* and form a heteromeric complex with DYT1 to regulate tapetum development (Zhu et al., 2015a). The expression of these three genes is relatively high in the tapetum, suggesting their important roles in tapetum development. However, single mutants of each of the bHLH010, bHLH089 and bHLH091 are developmentally normal due to their functional redundancy. The various double and triple mutants exhibit defects in tapetum morphology, callose degeneration, and pollen meiosis. Besides the conserved bHLH domain of bHLH010/bHLH089/bHLH091, a C-terminal conserved region named bHLH Protein Interaction and Function (BIF) domain facilitate DYT1 function via enhancing its nucleus localisation. The *bHLH010/089/091* forms a positive feedback regulatory motif with *DYT1* in the tapetum regulation network (Cui et al., 2016).

The *DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION 1* (*TDF1*)/*MYB DOMAIN PROTEIN 35* gene encodes a putative R2R3 MYB transcription factor, which is directly activated by *DYT1* (Gu et al., 2014). *TDF1/MYB35* is predominantly expressed in the tapetum, and low levels of expression were also observed in meiocytes and middle layer cells (Zhu et al., 2008). Like the *dyt1* mutant, the absence of *TDF1/MYB35* causes hypertrophied tapetum and middle layers, collapsed microspores, and eventually a sporophytic male sterility. Strong evidence shows that *TDF1/MYB35* acts upstream of *AMS* and *MYB80* (Chang et al., 2011).

The *ABORTED MICROSPORES* (*AMS*) gene encodes a tapetum-specific bHLH transcription factor, which directly regulates genes associated microspore compartmentalization, callose dissociation, sporopollenin precursor synthesis and transport, as well as pollen coat formation (Sorensen et al., 2003; Xu et al., 2014a). *AMS* acts downstream of both *DYT1* and *TDF1/MYB35*, with a prolonged expression which lasts from microspore meiosis to the second pollen mitosis (Xu et al., 2010; Wang et al., 2010). The *ams* mutant displays prematurely separated and abnormally enlarged tapetal cells, resulting in a complete degeneration of microspores at late anther developmental stages (Ma et al., 2012). AMS was believed to act as a principal master regulator of pollen wall formation based on its characteristic expression and downstream target genes (Xu et al., 2014a; Xiong et al., 2016).

The *MALE STERILE 188/MYB DOMAIN PROTEIN 80* (*MS188/MYB80*) gene encodes an R2R3 MYB transcription factor which is strongly expressed in the tapetum from microspore meiosis to the first pollen mitosis (Li et al., 1999; Higginson et al., 2003). Knockout of *MYB80* leads to a down-regulation of several key sporopollenin synthesis genes, including *CYP703A2*, a fatty acid hydroxylation

monooxygenase which catalyses in-chain hydroxylation of medium-chain fatty acid (C10–C14) (Morant et al., 2007; Dobritsa et al., 2009b). This down-regulation is consistent with the complete absence of an exine layer in *myb80* mutants (Zhang et al., 2007). Tapetal cells in the *myb80* mutant display more pronounced vacuolation and less secretory organelles compared with that of wild-type plants (Zhang et al., 2007b). Callose dissolution and tapetal cell wall degradation are also arrested (Li et al., 2007). It was reported that MYB80 is a direct regulatory target of *AMS*, and forms a complex with AMS to activate the expression of CYP703A2 (Phan et al., 2011; Xiong et al., 2016; Lou et al., 2014).

The *TRANSPOSABLE ELEMENT SILENCING VIA AT-HOOK* (*TEK*) encodes an AT-hook nuclear-localized protein, which is highly expressed in the tapetum during the tetrad stage and directly regulated by *AMS* (Lou et al., 2014). In plants, AHL proteins present in the nuclear matrix bind to the matrix attachment regions of chromosomal DNA, thereby regulating gene expression (Xu et al., 2013a, 2013b). In the *tek* mutant, nexine is absent from the pollen wall while the sexine formation remains normal (Lou et al. 2014). The expression of four members of *ARABINOGALACTAN PROTEIN (AGP)* family, namely, *AGP6*, *AGP11*, *AGP23* and *AGP40*, are induced by TEK. Jia et al. (2015) demonstrated that AGPs form the backbone of the nexine layer as an expression of AGP6 in the *tek* mutant can partially restore nexine formation.

Although lots of research has been carried out on early and middle tapetum development, the regulation of the late (post-meiotic) tapetal development has barely been reported. The *MALE STERILITY1 (MS1)* gene is transiently expressed at the time point when the tapetum goes through the second phase transition, thus playing an

essential role in late tapetum development. The details of *MS1* gene will be reviewed in the next section. *MYB99* acts downstream of *MS1*, and has been long postulated as a direct target of MS1, however, due to a lack of a clear phenotype of the mutant, its role in tapetum development remains mysterious (Dubos et al., 2014; Alves-Ferreira et al., 2007).

The specification and differentiation of tapetal cells is tightly regulated by the fine-tuned expression of transcription factors and small RNAs, phytohormone receptors, as well as protein ligand-receptor signalling (Sun et al., 2007). In the **Table 1.2**, information on these regulators was summarised, including aspects such as mutant phenotypes, expression patterns and putative functions in tapetal cell development (Parish and Li, 2010). A gene regulation network of tapetum development was also proposed by (**Figure 1.6**) (Gómez et al., 2015).

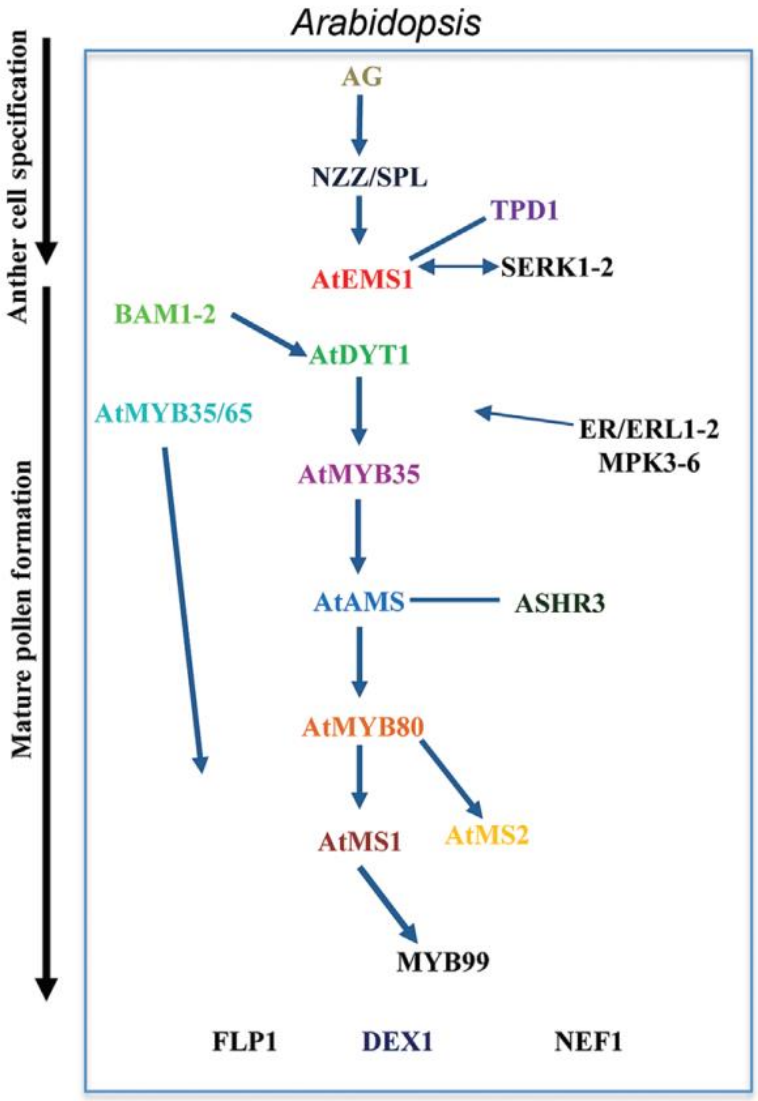


Figure 1.6 A gene regulation network of the *Arabidopsis* tapetum. Adapted from Gómez et al., (2015).

Table 1.2 Transcription factors and cell receptors involved in *Arabidopsis* anther development. (Parish and Li, 2010)

Proteins	Expression pattern	Mutant phenotype	Putative function in anther development
Transcription factors			
AGAMOUS (AG) MADS box	Primordia; tapetum and filament.	Primary sporogenous cells fail to form tapetum and endothelium layer.	Regulates SPL/NZZ
SPOROCYTELESS/NOZZLE (SPL/NZZ)	Initially in anther primordium; in both parietal and sporogenous cells; becomes restricted to tapetum and microsporocytes/meiocytes at stage 5.	Primary sporogenous cells do not form microsporocytes and anther cell wall layers; primary parietal cells do not form tapetum.	May act in L2 layers driving archesporial division. Formation of sporogenous cells and surrounding somatic cell layers; may control genes required for microsporocyte and tapetal differentiation.
AtMYB33 AtMYB65	Initially in all anther cell layers; following meiosis, largely confined to tapetum.	Redundant genes; abnormal tapetum at meiosis, undergoes hypertrophy at pollen mother cell stage.	Tapetal differentiation.
DISFUNCTIONAL TAPETUM1 (DYT1) (bHLH)	Strong in late stage 5 to early stage 6 tapetum, low in meiocytes.	Male sterile. Tapetal cells with enlarged or excess vacuoles; meiocytes lack thick callose wall and do not complete meiotic cytokinesis and collapse.	Regulates expression of tapetum genes.
TAPETAL DEVELOPMENT AND FUNCTION 1) (TDF1) (MYB	Tapetum, meiocytes and microspores.	Stage 6: middle layer vacuolated, tapetal cells abnormally vacuolated and enlarged; excess tapetal division, microspores collapse, no pollen formed. Stage 8: tetrads enveloped by vacuoled tapetum and the middle layer.	Tapetal late stage development;
ABORTED MICROSPORES (AMS) (MYC class of bHLH)	Tapetum, postmeiosis. (Possibly also in microspores).	Male sterile. Tapetal cells become large and vacuolated. Premature tapetum and microspore degeneration.	Tapetal and microspore development.
AtMYB103/80	Tapetum, postmeiosis (stage 7) until tapetum degeneration initiated (late stage 9).	Defects in tapetal cell wall; impaired section; premature tapetal breakdown; male sterile, microspore development defective.	Transition of tapetum to secretory type.
MALE STERILITY 1 (MS1) (Leu-Zipper, PHD-finger)	Tetrad stage in tapetum (stage 7), postmeiosis, commences as callose begins to degrade.	Male sterile, exine not formed; abnormal vacuolation of tapetum and microspores; no tapetal PCD at multinucleate and bicellular stages. Tapetal degradation occurs by necrotic breakdown.	Regulates exine formation and pollen coat development; tapetal development.

Continued table of **Table 1.2**

Proteins	Expression pattern	Mutant phenotype	Putative function in anther development
AtMYB99	Tapetal cells; commences at tetrad stage (stage 7) and strong in stages 8 and 9. Direct target of MS1.	Thin tapetal cells; reduced pollen grain viability.	Unknown.
AtMYB32	All major plant organs, strong expression in tapetum.	Collapsed and unviable pollen.	Regulates genes in PAL pathway.
AtMYB26	Low level throughout anther, highest level at pollen mitosis 1 and bicellular stages.	Male sterile due to failure of anther dehiscence.	Determines endothelial cell development and acts upstream of lignin biosynthesis pathway.
Leucine-rich repeat receptor-like kinases (LRR-RLKs)			
EXCESS MICROSPOROCYTES1/ EXTRA SPOROGENOUS CELLS (EMS1/EXS)	Predominantly in tapetum.	Additional L2 cells undergo archesporial differentiation. Stage 5: completely lacks tapetum and middle layer, produces excess microspores. Stage 6: microspores enlarged, still attached.	Tapetum formation.
BARELY ANY MERISTEM 1& 2 (BAM1/2)	Archesporial cells from stage 2; preferentially in sporogenous cells and pollen mother cells at later stages.	Redundant genes; double mutant lacks endothelium, middle layer and tapetum; extra microspores.	Signals differentiation of archesporial cells; development and function of pollen mother cells.
SOMATIC EMBRYOGENIC RECEPTOR KINASE 1 and 2 (SERK 1/2)	All aerial organs; broad expression in anthers at stage 6; later restricted to tapetum (stage 9).	Redundant genes; double mutant same as *ems/exs.*	Tapetum development and microspore maturation.
RECEPTOR-LIKE PROTEIN KINASE 2. (RPK2)	Tapetum stages 8 and 9.	Middle layer not formed; tapetum abnormally enlarged, degradation restricted.	Controls middle layer determination and tapetal cell fate during microspore maturation; may facilitate subsequent tapetal degradation.
Ligand			
TAPETUM DETERMINATION1 (TPD1) (176 amino acid protein, probably secreted)	Microspores.	Same as *ems1.*	Secreted by microspores (or their precursors), binds and activates EMS receptor kinase on tapetum; tapetal formation induced or may repress genes required for microsporocyte differentiation in tapetum precursors.

1.3 ARABIDOPSIS MALE STERILITY1 (MS1) GENE

Among the *Arabidopsis thaliana* tapetum regulation network revealed up to now, *MALE STERILITY1* (*MS1*) gene is the most downstream transcription factor, which encodes a plant homeodomain finger (PHD) type transcription factor. The MS1 protein belongs to a small plant-specific protein family, whose members have the PHD domain located at their most C-terminal end and the other regions of the proteins lack any known protein domain, which have been identified. Previous studies uncovered its crucial role in late tapetum development and function and viable pollen formation (Wilson et al., 2001; Ito and Shinozaki, 2002; Yang et al., 2007; Ito et al., 2007). A series of the mutant alleles of *MS1* in different *Arabidopsis* ecotypes were identified and studied, showing extremely similar phenotypes, suggesting plant background does not have a significant impact on MS1 function. Detailed information on MS1 alleles is summarised in (**Table 1.3**).

1.3.1 Potential role of *MS1* in anther development

The mutant of *ms1* displays complete male sterility and lack of mature pollen grains under a range of environmental conditions. The other aspects of plant development are not affected, which is consistent with the tapetum-specific expression pattern of *MS1* (Ito et al., 2007; Yang et al., 2007). For early pollen development, the key biological process appears normal, such as callose production, pollen meiosis, tetrad formation, callose dissolution and microspore release. The mutant defects appear immediately after the tetrads are released in three dimensions: tapetum, exine wall

and integrity of pollen grain, which highlights the role of MS1 in these aspects of anther development.

The cytoplasm of tapetal cells firstly becomes granular and abnormally vacuolated at the tetrad stage. The proplastids within the tapetal cells, which would later develop into elaioplasts or tapetosome in wild type (WT), were completely arrested in the mutant. Later, the *ms1* tapetal cells lose secretory properties and polarity, with no vesicles fusing to the radial and inner tangential membranes. In the late phase of tapetal cell development in the mutant, vesicles progressively fuse to form an extensive network of large vacuoles, which effectively engulf the entire cell. Tapetal cells of the mutant then go through the autophagic breakdown of the vacuoles and necrotic-based cell death, as opposed to the tapetal PCD process in wild-type (**Figure 1.7**) (Vizcay-Barrena and Wilson, 2006; Yang et al., 2007).

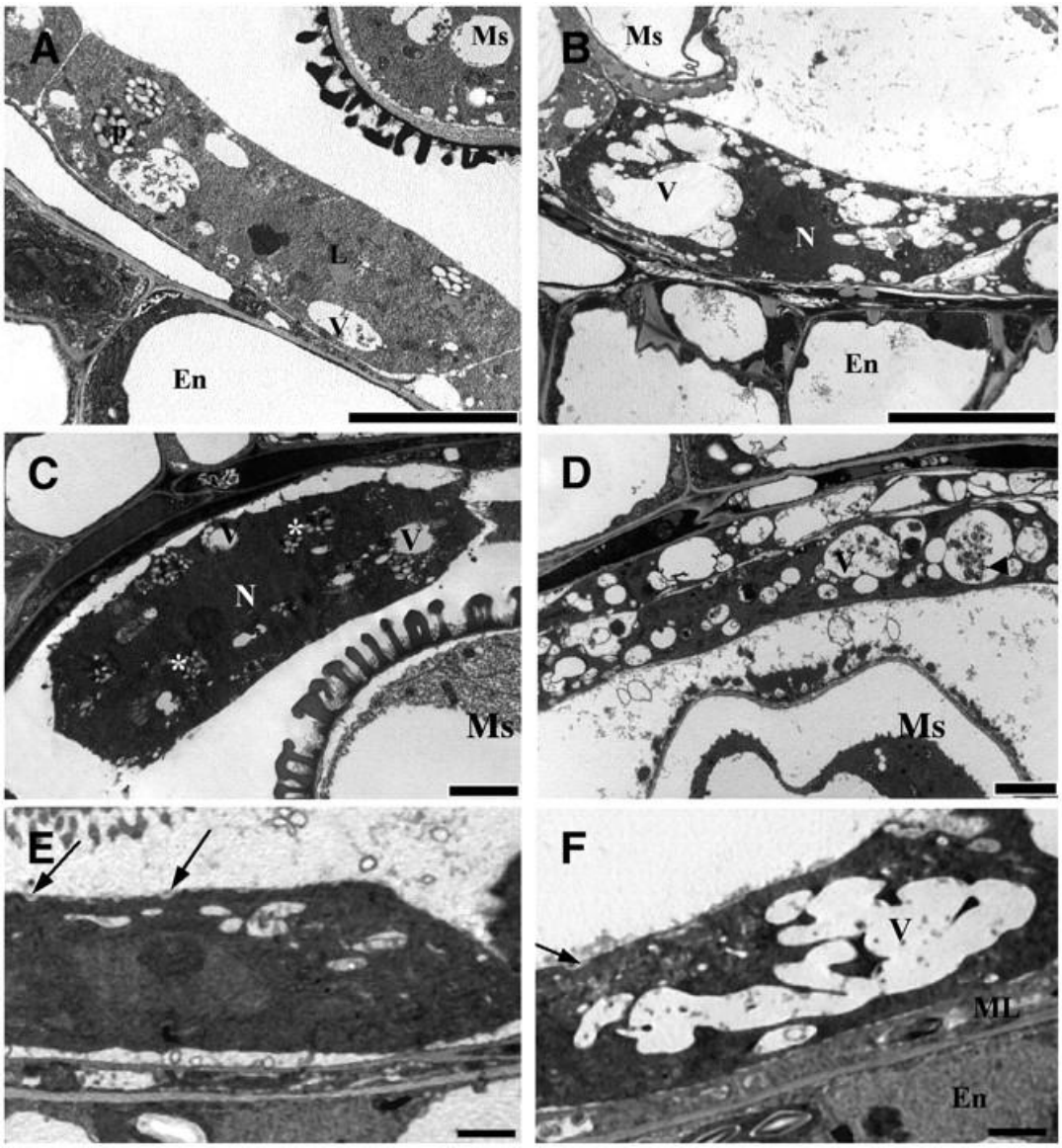


Figure 1.7 Transmission Electron Micrographs of Cross Sections through Wild-Type (A, C and E) and *ms1* Mutant Anthers (B, D and F).

Anther material at Ring-Vacuolated Stage/Microspore Mitosis. Bar=0.5 μm. Ms, microspore; c, callose; m, mitochondria; p, plastid; ER, endoplasmic reticulum.
Figure is taken from Yang et al., (2007).

The exine deposition in the *ms1* mutant is reduced, and the exine structure is abnormally patterned, however, the primexine matrix and undulated microspore membrane surface are normal (**Figure 1.8 C and D**). The deposited sporopollenin forms irregular-shaped spike-like structures that do not resemble the regularly patterned, rod-like shape of the wild-type bacula, and *ms1* mutant exine lacks tecta

(**Figure 1.8 A and B**). Aberrations in exine structure in *ms1* mutant suggest an unusual chemical composition of the sporopollenin, which is linked to the defects in the tapetal biosynthesis or secretory pathways of the sporopollenin. In this regard, the defect in exine is a secondary phenotype caused by the aborted tapetal development in *ms1* mutants (Yang et al., 2007).

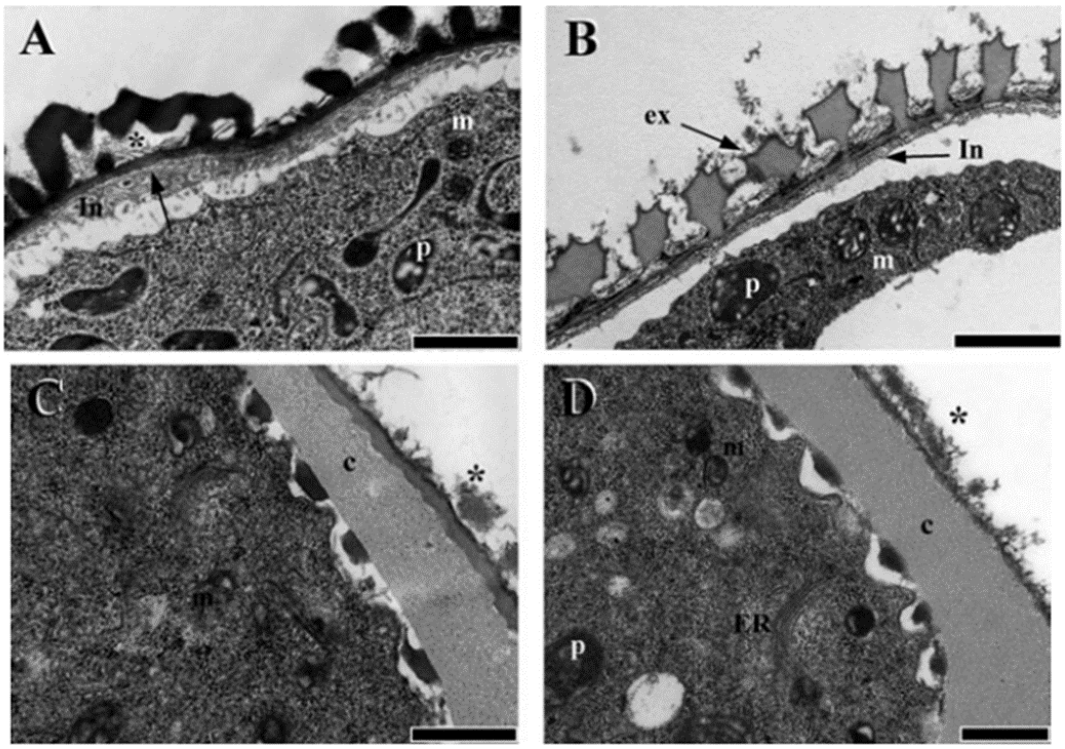


Figure 1.8 Transmission electron micrographs of cross-sections through microspore

Adapted from Vizcay-Barrena and Wilson, (2006). Wild-type (A, C) and the *ms1* mutant (B, D); at the ring-vacuolated pollen stage (A, B, Bar = 1μm) and at the tetrad stage (C, D, Bar = 5μm). (A, B). In, intine; m, mitochondria; p, plastid; ex, exine; V, vacuole; N, nucleus; Nu, nucleolus; Ms, microspore; c, callose; m, mitochondria; p, plastid; ER, endoplasmic reticulum.

As soon as pollen grains approach their maturity, the intine wall and pollen coat are formed. As the biosynthesis of pollen coat material occurs in the tapetum (Mayfield, 2001), it is not surprising that in the *ms1* mutant, the abnormal tapetum cannot normally coat the pollen grains. Strikingly, the intine wall, which is believed to be

synthesised by microspores, fails to form or has limited development in the *ms1* mutant, suggesting a reduction of the pollen viability. Late in anther development, the microspores in the *ms1* mutant are extremely prone to lysis and finally, collapse before the dehiscence of the anther. This indicates that the integrity of the pollen wall and pollen coat is essential for the maintainance of pollen grains in the mature, dehydrated anther locule (Yang et al., 2007; Ito et al., 2007).

Microarray expression profiling confirmed that a large group of tapetum-specific genes were down-regulated in the *ms1* mutant, most of which are involved in cell PCD, wall biosynthesis, lipid metabolism and secondary metabolism. In summary, MS1 has a critical role in directing the middle-late tapetum phase transition, thereby initiating tapetal PCD, and accomplishing the pollen wall formation and pollen coating. (Alves-Ferreira, Wellmer et al., 2007; Ito, Nagata et al., 2007; Yang, Vizcay-Barrena et al., 2007).

Table 1.3 The mutant alleles of *MALE STERILITY1* in different *Arabidopsis* ecotype backgrounds.

MS1 allele (ecotype)	**Mutagen**	**Mutation site [1]**	**Effects on MS1**	**Ref.**
ms1-1 (Ler)	EMS	G882 to A	dysfunction of the splice donor site of the second intron	(Van der Veen and Wirtz,1968)
ms1-2 (Ler)	EMS	C1500 to T	caused premature termination of the polypeptide in the third exon	(Dawson et al., 1993)
ms1-3 (Ler)	EMS	C1380 to T	caused premature termination of the polypeptide in the third exon	(Dawson et al., 1993)
ms1-4 (Ler)	EMS	G1036 to A	dysfunction of the splice acceptor site of the second intron	(Dawson et al., 1993)
ms1-5 (Ler)	X-rays	T215 deleted	caused premature termination of the MS1 polypeptide	(Dawson et al., 1993)
ms1-6 (Ler)	Gamma-rays	G882 to A	dysfunction of the splice donor site of the second intron	(Kalantidis et al., 1994)
ms1-8 (Nossen)	Ds-insertion	G1856 Ds	fully knocked out	(Ito and Shinozaki, 2002)
ms1-9/hkm(Col)	T-DNA	C1258 to CC	fully knocked out	(Ariizumi et al., 2005)

[1] The mutation positions are indicated in relation to the first base of ATG start codon, which is set to 1.

1.3.2 Regulation of MS1 by tapetal transcription factors

A collection of transcription factors was reported to activate the expression of *MS1*. For example Ye et al., (2010) demonstrated that *BR-INSENSITIVE-EMS-SUPPRESSOR1* (*BES1*), an important BR transcription factor effector, directly regulates *MS1* by binding to the BRRE (BR response element)-box of the MS1 promoter region. The phytohormone BR has a significant effect on plant fertility, with the BR biosynthetic and signalling mutants leading to abnormal tapetal development (Zhu et al., 2015b).

Additionally, MS1 acts downstream of a series tapetum-specific transcription factors, namely *DYT1*, *AMS*, *bHLH010*, *bHLH089*, *bHLH091*, *TDF1*/*MYB35* and *MYB80*. These transcription factors are key transcription factors that are essential for tapetum development and function (Zhu et al., 2011, 2015a; Lou et al., 2014). In the single mutants of *DYT1*, *AMS*, *TDF1*/*MYB35, MYB80* and a triple mutant of *bHLH010/089/091, MS1* has either no or very low levels of expression (Zhu et al., 2015a; Xiong et al., 2016; Cui et al., 2016; Gu et al., 2014). *DYT1* is transiently expressed at the tapetum specification stage and is important for tapetal cell fate determination. *bHLH010*, *bHLH089*, *bHLH091* acts downstream of *DYT1* and each of them can form a heteromeric complex with DYT1 and significantly activate the expression of MS1 *in vitro*. *AMS* has a prolonged expression pattern in the tapetum, from the microspore meiosis to the second pollen mitosis stage, which fully covers the expression pattern of the MS1 (Xu et al., 2010). A considerable proportion of the *AMS* regulatory targets are remarkably down-regulated in the *ms1* mutant, indicating AMS achieves its regulation function on downstream targets in a progressive manner, probably by protein-to-protein interaction (PPI) with different transcription factors.

More recently, AMS was reported to interact with MYB80 and binds to the promoter region of CYP703A2, thereby activating its expression (Xiong et al., 2016). CYP703A2 and MYB80 are highly co-expressed proteins with MS1 (Ehlting et al., 2008). Considering that the expression pattern of MYB80 temporally and spatially approximates to that of MS1, it is possible that MS1 is simultaneously or successively regulated by several of its upstream transcription factors, while the MYB80 acts like the final gatekeeper, which ultimately controls the activation of MS1 expression.

1.3.3 Rice and barley orthologues of MS1

In monocot crops, the role of MS1 in regulation of late tapetum development appears to be conserved. The rice (*Oryza sativa*) MS1 is called PERSISTANT TAPETAL CELL1 (PTC1), which is specifically expressed both in tapetal cells and microspores during rice anther stage 8 and 9, regulating programmed male reproductive development and functional pollen formation. Similar to *ms1* mutants, *ptc1* mutants showed delayed tapetal degeneration, a lack of tapetal DNA fragmentation, abnormal pollen wall formation and aborted microspore development. Strikingly, *ptc1* mutants display uncontrolled tapetal cell proliferation and a subsequent necrosis-like cell death. This novel phenotype may represent additional regulatory roles of PTC1 in rice tapetum development compared with MS1, which might explain the fact that PTC1 can only partially rescue the homologous *ms1* mutation in the complementation assay (Li et al., 2011). The barley (*Hordeum vulgare*) MS1, termed as HvMS1, has an anther-specific tapetal expression between the stage of late tetrad and early microspore release, playing a critical role during barley pollen development. Expression of HvMS1 in the homozygous *ms1 Arabidopsis* mutant driven by the native *Arabidopsis* MS1 promoter, can fully complement the male sterility phenotype.

In barley, both silencing and overexpression of HvMS1 leads to male sterility (Fernández Gómez and Wilson, 2014).

1.4 SUMMARY AND AIMS OF THIS PROJECT

Spore and pollen grain development is the result of interactions between two tissues, which acquire their identity after archesporial differentiation, namely microspore mother cells and tapetal cells. The tapetum is a conserved tissue in most land plants, which plays nourishing and secretory roles in spore or pollen development. The specification and development of the tapetum in *Arabidopsis* has been extensively studied. Previous research identified *MS1* as an essential transcription factor directing late tapetum development. Moreover, several early-tapetum-specific transcription factors were reported to regulate the expression of *MS1* by binding to its promoter region (reviewed in **Section 3.1.1.1**). However, *MS1* is transiently expressed only at a relatively late stage rather than activated within early tapetum and constitutively expressed throughout the tapetal development. This contradiction in the regulation of MS1 expression has not been investigated. Additionally, *MS1* protein lacks an annotated DNA binding domain, which means that the mechanism by which *MS1* exerts the regulation on downstream targets remains unclear. This study was aimed at addressing the genetic regulation and the molecular mechanisms of *Arabidopsis MALE STERILITY1*, this comprised of the following specific objectives:

- Identifying the functional regions of the upstream sequence of MS1 that play regulatory roles in the expression of *MS1* using phylogenetic footprinting and promoter dissection analysis, and verifying the interaction between these functional regions and the early tapetum TFs in a yeast-one-hybrid assay.

- Verifying if the MS1 is directly auto-regulated.

- Characterizing a complete loss-of-function line of E3 ubiquitin ligases *LRBs*, one of which encodes an MS1 protein interacting protein, to uncover if LRBs mediated the proteolysis of the MS1 protein.

- Analysis of the phylogeny, protein sequence, protein subnuclear localization of MS1 and its homologs to predict the working molecular mechanism underlying its regulation role.

- Re-analysis of *MS1* and other tapetal TF microarray data combined with the cell type-specific expression profiling data of wild type anthers, to isolate the possible direct regulatory targets of the MS1.

- Detailed phenotyping of two mutant alleles of the *MYB99* gene, a possible direct *MS1* regulatory target, to further resolve the function of *MS1* in tapetum development.

- Verifying the direct binding of MS1 to the genomic locus of *MYB99* via yeast one-hybrid and ChIP-PCR, to confirm the direction regulation of MYB99 by MS1.

CHAPTER 2 GENERAL MATERIALS AND METHODS

2.1 PLANT MATERIALS AND GROWTH CONDITIONS

2.1.1 Plant general growth condition

The *Arabidopsis thaliana* and *Nicotiana benthamiana* seeds of wild-type and mutant were obtained from the Wilson lab, University of Nottingham, or Zhang lab, University of Shanghai Jiao Tong University. All T-DNA insertion mutants were obtained from the Nottingham Arabidopsis Stock Centre (NASC). Seeds were sown onto the Levington M3 compost (The Scott's Company Ltd, UK).

Plants were grown in plastic pots or trays at a long day condition: 23/18°C (day/night), 16-hour photoperiod and 180±20 μmol/s/m^2 light intensity. 8-week old plants were no longer watered. The seed were collected when the siliques had dried out completely. The collected seeds were stored in a dry atmosphere at room temperature.

2.1.2 Agrobacterium-mediated transformation by floral dipping

The floral dipping transformation of Arabidopsis was done based on the procedure of (Zhang et al., 2006b). Specifically:

A 5mL of overnight Agrobacterium starter culture in LB medium containing required antibiotics was inoculated into 100ml LB medium and grown at 28°C with shaking until the OD_{600} had reached 1.0-2.0. The cells were pelleted by centrifugation at

4,000g for 10 minutes and afterwards re-suspended in 50ml of freshly made solution with 5% (w/v) of sucrose and 0.05% (v/v) surfactant Silwet L-77.

Each inflorescence of newly flowering Arabidopsis with any existing siliques removed was dipped into the *Agrobacterium* suspension for about 90 seconds. To ensure the humidity level maintained, the dipped plants were covered in the plastic sleeves overnight. The plants were uncovered the following day and left to grow in the normal condition until the seeds were suitable for harvest.

2.1.3 Germination, growth and antibiotic selection

Seeds were surfaced sterilised in 1.5 ml Eppendorf microcentrifuge tubes of 70% (v/v) ethanol for 10 minutes with gently agitating. The seeds were then collected by centrifuging the tube and discarding of the supernatant ethanol. Such procedure was repeated twice and the seeds then were washed using water. Excessive ethanol and water were absorbed by sterile filter paper in a laminar flow cabinet and allowed to dry for 2 hours.

For germination and antibiotic selection, the sterilised seeds were sown on the surface of ½ MS media (Murashige and Skoog, 1962) solidified with 1.5% (w/v) agar media in 12 x 12 square petri dishes. If required, filter sterilised antibiotics were added at the appropriate concentration in the MS media after the autoclaved media cooled to 50 °C. After 10 days of the growth on MS plate in growth room, the young seedlings of the plant were transferred into plastic pots filled with M3 compost and grown in glasshouse.

2.2 ANALYSIS AND MANIPULATION OF DNA

2.2.1 Genomic DNA extraction from *Arabidopsis*

Genomic DNA extraction from Arabidopsis was using DNeasy Plant Mini Kit (Qiagen) in accordance with the manufacturer's instructions. Approximately 100 mg fresh weight of plant materials for genomic DNA extractions were collected in 1.5 ml Eppendorf microcentrifuge tubes, frozen in liquid nitrogen immediately, and then ground with a micropestle into powder until homogenized texture achieved. The tubes were kept in liquid nitrogen to avoid thawing during grinding process. Samples were then incubated at 65°C for 10 minutes after addition of Buffer AP1 and 4µl 100mg/ml RNase A solution with briefly vortexing mix. Cell lysates added with 130 µl Buffer AP2 were incubated on ice for 5 minutes and then passed through the QIAshredder Mini spin column before being centrifuged at 20,000 g for 2 minutes. The supernatant was transferred into a new microcentrifuge tube with 675µl Buffer AP3/E following with a mix. The sample was then loaded onto a DNeasy Mini spin column in a 2 ml microcentrifuge tube following with centrifugation at 6,000 g for 1 minute. The column loading with 500µl Buffer AW was centrifuged for 1 minute at 6,000 g with the flow-through discarded and the above process was repeated once. Then the column was centrifuged at 20,000 g for 2 minutes. Column loaded with 50µl molecular grade water (Sigma) was incubated at ambient temperature for 5 minutes and then centrifuged at 6,000 g for 1 minute to elute the DNA. NanoDrop Spectrophotometer (NanoDrop Technologies, USA) was used to determine the quality and quantity of genomic DNA. DNA samples were used immediately or stored at −20°C.

2.2.2 Standard PCR

PCR amplification was implemented using RedTaq ReadyMix PCR Reaction Mix (Sigma-Aldrich) following manufactures' introductions in a 10µl reaction volume. Standard PCR reaction composition including 5 µl RedTaq ReadyMix, 0.25 µl Forward primer (10 pmol/µl), 0.25 µl Reverse primer (10 pmol/µl), 0.5µl Template DNA (50 ng) and up to 10 µl water. Standard PCR thermocycle programme was conducted under typical conditions: 94°C for 3 minutes (Initial denaturing), 40 cycles of 94°C for 30 seconds (Denaturing), 55-60°C for 30seconds (Annealing), 72°C for 1minute per kb (Extension) and then 72°C for 4 minutes (Final extension).

2.2.3 Colony PCR

Colony PCR was employed to rapidly detect positive clones of constructed plasmids in bacterial colonies. Overnight incubated the newly transformed bacteria in 37°C and picked the fresh colonies from the media surface using 2µl pipette tip. 5-10 colonies were picked in this way, then each tip was immersed into the PCR reaction tubes. After 10 minutes, tips were removed from the PCR tubes and standard PCR was conducted.

2.2.4 PCR Cloning of fragments

Q5 (New England Biolabs) or KOD-Plus-Neo (TOYOBO) high-fidelity RNA polymerase was used to prepare DNA fragments for cloning. The PCR reaction was

set up in accordance with the manufacturer's instructions. If required, 0.5 µl RedTaq DNA polymerase was added to the 50 µl finished high-fidelity PCR reaction and ran a single extension cycle for 10 minute, for the 3'-end A tail addition. The PCR product was then for corresponding TA cloning procedures.

2.2.5 Agarose gel electrophoresis

Agarose gel electrophoresis was adopted to separate DNA by size for visualization and purification of fragments from PCR or cloning reactions. 0.8-3.5% (w/v) molecular biology grade agarose was dissolved in 0.5x TBE buffer (45 mM Tris-borate, 1mM EDTA) with 0.15µg/ml ethidium bromide. The size of DNA fragment was determined using appropriate DNA marker (HyperLadder I, II or V, Bioline) by mixing the DNA sample with 1/5 volume of 6x NEB Purple Loading Buffer. DNA samples were electrophoresed at an electric field strength of 10 V/cm until DNA fragments were separated sufficiently. DNA was visualized with long wave UV light, and gel images were captured and recorded with SnapGene Software (GSL Biotech LLC).

2.2.6 DNA gel purification

DNA gel purification was done in accordance with the manufacturer's instructions using the E.Z.N.A.® Gel Extraction Kit. DNA bands ranging from 70 bp to 5 kb in length were cut from the agarose gel and dissolved into the binding buffer under a 65 °C water bath. Then, the buffer was applied onto a HiBind® DNA spin column. Following two wash steps using the binding buffer, DNA was eluted with deionized water.

2.3 EXPRESSION AND DNA ANALYSIS

2.3.1 Total RNA extraction

Total RNA was extracted from fresh leaf or inflorescences tissues (approximately 100mg) using TRIzol® Reagent (Life technology). The extraction procedures were done in accordance with the manufacturer's instructions. RNA samples were measured using NanoDrop Spectrophotometer (NanoDrop Technologies, USA) and stored at −80° C.

2.3.2 cDNA synthesis

2μg of total RNA was mixed with 50 μM oligonucleotide $(dT)_{12-18}$, 1μL 10mM dNTP Mix and nuclease free water to a total volume of 13μL. The mixture was then incubated at 65°C for 5 minutes before cooling on ice. 4 μl of 5x First Strand Buffer, 1 μl DTT (0.1M), 0.5 μl Superscript III reverse transcriptase (200 U/μl, Invitrogen, USA) and 1 μl RNaseOUT Recombinant RNase Inhibitor (40 U/μl, Promega, USA) were then added to each sample and incubated at 50°C for 60 minutes. The reaction was inactivated by heating to 75°C for 10 minutes and stored at −20°C. In the later practice, 1 μl of prepared cDNA was used as template in RT-PCR analysis.

2.3.3 Semi-quantitative PCR

Fresh tissues were collected for RNA extraction; each tissue sample was weighed and made sure equal material was used for the RNA extraction. After the RNA extraction, the concentration of each RNA sample was measured, RNA input for each cDNA

synthesis reaction was made equal. RT-PCR reactions were set up as standard PCR reaction except the input cDNA were normalised based on the concentration of the RNA that had been used for cDNA synthesis. PCR products were electrophoresed on 3.5 % agarose gel with GelGreen Nucleic Acid Gel Stain (Biotium company) added in the gel. The gel images were taken using the BioRad Gel Documentation Systems. The relative quantity of the PCR product was analysed using the software of Image Lab Lite 5.2.1 (BioRad company) based on the fluorescence signal intensity of DNA dye. YLS8, Rhfp, GAPC2 and RHIP1 were used as reference gene based on the housekeeping reference gene scoring (Hong et al., 2010).

2.3.4 Quantitative (real-time) PCR

Quantitative real-time polymerase chain reaction (qRT-PCR) was performed for accurate quantitative gene expression analysis. It's a reproducible, and less labour-intensive PCR method. SYBR Green QPCR 2x Master Mix (Thermo Fisher Scientific, UK) was used following manufacturer's instructions and operated on the Light Cycler 480 Real-Time PCR system (Roche Applied Science) controlled by the LightCycler 480 software. The process of amplification took place under the condition of 95 °C for 3 minutes, 40 cycles of 95 °C for 30 seconds, 55 °C for 1 minute, and 72 °C for 1 minute followed by the dissociation process under the circumstance of 1°C per cycle from 55°C to 95°C. All samples were run in duplicate and the retrotransposon gene *TA3* (AT1G37110) was used as a reference gene to normalize amplification for gene expression analysis and. The $2^{-\Delta\Delta CT}$ analysis method is adopted to analyse the relative changes in gene expression (Livak and Schmittgen, 2001).

2.4 MOLECULAR CLONING METHODS

2.4.1 DNA restriction digestion and DNA ligation

Plasmid DNA requiring digestion using the desired restriction endonucleases was incubated in reaction volumes of 50µl at 37°C overnight and then inactivated at 85°C for 20 minutes.

DNA ligation was performed using T4 DNA ligase (Invitrogen) in accordance to the manufacturer's instructions. The reaction underwent overnight incubation at 16°C. Approximately 25-50 ng of plasmid DNA was used with an insert: plasmid molar ratio of 3:1to ensure the ligation efficiency. 5µl of the ligation reaction was transformed into 50µl One Shot Mach1TM competent cells (Invitrogen).

2.4.2 TOPO Clone and Gateway Recombination Cloning

TOPO clone of the PCR product was done in accordance to the manufacturer's instructions using pCR™8/GW/TOPO® TA Cloning Kit. After the TA TOPO cloning, the DNA fragment of interest was recombined onto the pCR8 Gateway entry clone. Then the further cloning procedures were compatible to the Gateway BP/LR cloning.

Gateway cloning was implemented using Gateway™ BP/LR Clonase™ II Enzyme Mix (Invitrogen) in accordance to the manufacturer's instructions with half-quantities of the instructed reaction system as an exception. All BP and LR reactions were incubated at 25°C overnight or 5 hours and the reactions terminated by adding 1µl proteinase K (Invitrogen) and incubating at 37°C for 10 minutes. Re-combined

Gateway reactions were transformed into DH5α competent cells while all empty Gateway vectors were maintained in *ccdB*-resistant DB3.1 competent cells.

2.4.3 *E. coli* transformation

Competent E. coli cell in 50 µl aliquots were thawed on ice for 5 minutes and 50 ng of supercoiled plasmid DNA was added to each aliquot in the meanwhile. The cells were then incubated on ice for 30 minutes before being pre-heated at 42°C for 1 minute. The cells were incubated on ice for another 20 minutes for recovery and 200 µl SOC medium (2% (w/v) Tryptone, 0.5% (w/v) yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl2, 20mM glucose) was added to each aliquot at ambient temperature. The cells were then incubated with shaking for 1 hour at 37℃. A proper amount (50-100 µl) of cell culture was spread onto semi-solid LB agar plate containing appropriate antibiotic and incubated overnight at 37℃.

2.4.4 Agrobacterium transformation

50 µl aliquots of competent *Agrobacterium tumefaciens* cells (strain GV3101) frozen at −80 °C were thawed on ice for 5 minutes. 200ng of supercoiled plasmid DNA was added to each aliquot with gentle mix on the ice. The cells were transferred to a dry electroporation cuvette and pulsed in the electroporation apparatus with 1800V. A 1.25 kV/cm electric field was applied and the cells were transferred into 1ml LB media immediately. The cells were incubated at 28℃ for 2 hours before plating onto LB agarose plate containing appropriate antibiotics. The plates were then incubated at 28 °C for 48 to 72 hours.

2.4.5 Plasmid DNA extraction and sequencing

Plant genomic DNA was extracted using DNeasy Plant Mini Kit (Qiagen, UK) following the manufacturer's instructions. Plasmid DNA was extracted from 5-10 ml *E. coli* cultured overnight in LB media plated containing appropriate antibiotics using QIAprep Miniprep Kit (Qiagen, UK) following the instructions supplied by manufacturer. DNA samples were precipitated on ice for 20 minutes before pelleting by centrifugation at 4°C, 13 000 rpm for 20 minutes. The pellet was then washed twice with 70% (v/v) ethanol and suspended in water. DNA samples were used immediately or stored at −20 °C. NanoDrop Spectrophotometer (NanoDrop Technologies, USA) was used for DNA quality and quantity determination.

Sequencing was performed using the Sanger sequencing services provided by Source Bioscience (Nottingham). Date analysis was performed with SnapGene Sequence Editing Software (GSL Biotech LLC, US).

2.5 PROTEIN METHODS

2.5.1 Protein gel electrophoresis

For each sample, 15 µl solution of protein was mixed with 5 µl 4X SDS-PAGE loading buffer and incubated in 95 °C for 10 minutes before loading on the 12 well NuPAGE® Tris-Acetate gels (Thermo Fisher Scientific Company). The Electrophoresis was finished in the XCell SureLock™ Mini-Cell Electrophoresis Tank in accordance of the manufactures' instructions except 400 ml MES running buffer with antioxidant added and Amersham ECL Full-Range Rainbow Molecular

Weight Markers was used. The electrophoresis was run under 250V voltage for 1 hour. After the electrophoresis, the SDS-PAGE gel was washed in fresh MES buffer and stained with 20ml SimplyBlue™ SafeStain (Thermo Fisher Scientific Company) solution for 40 minutes at room temperature with agitation. Then the gel was put into deionized water to get slight detained until the protein band appears clear. Gel image was taken and the gel was complete detained using fresh MES buffer for the following western blot assay.

2.5.2 Western blot

The protein transfer was conducted using the Trans-Blot® SD semi-dry transfer cell (Bio-Rad company) in accordance to the manufacturer's instruction. Specifically, NuPAGE® Transfer Buffer (Thermo Fisher Scientific Company) was used to desalt the electrophoresed protein gel and pre-soak the filter paper and the nitrocellulose membrane for an equilibration. An 'transfer sandwich' (filter_paper_x2-gel-membrane- filter_paper_x2) was made and put between the two pad of the transfer cell. Air bubbles and excess buffer was removed from the transfer sandwich and the transferring was run at 12V of voltage for 45 minutes. The transferred nitrocellulose membrane was then blocked in 5% (w/v) BSA solution overnight at 4 °C. Then incubated sequentially with anti-GFP antibody and a second antibody conjugated with HRP luminophore. The membrane was washed five times before the chemiluminescent detection using SuperSignal™ West Femto Maximum Sensitivity Substrate Kit (Thermo Fisher Scientific, UK). All the detection steps were carried out in accordance to the manufacturers' instructions in the darkroom.

2.6 MICROSCOPY METHODS

2.6.1 Scanning electron microscope

Fresh inflorescence was taken from 6-week-old plant and immediately fixed into 100% methanol for 1 hour. The dehydration steps were done based on the protocol of (Talbot and White, 2013). Critical point dry and metal coating of the sample was done using the Leica EM CPD300 Critical Point Dryer and Super Cool Sputter Coater respectively in accordance to the manufacturers' instructions. 1.1.1 Scanning electron microscope observation was conducted using Hitachi S3400N VP-SEM under the supervisions of the technicians in SEM division, Instrumental analysis Centre of Shanghai Jiao Tong University.

2.6.2 Confocal microscopy

Freshly prepared tissue sample was put on ice before the microscopy observation using the Leica TCS SP8 STED 3X confocal microscopy. The operation was under the supervision of the technicians in Confocal Microscopy division, Instrumental analysis Centre of Shanghai Jiao Tong University.

2.6.3 Confocal image analysis

Leica Application Suite 3.0.1 was used to process and deconvolute the STED confocal images. Pictures of maximum projections were generated throughout the z-depth confocal stacks obtained within a 7.21 μm optical section.

2.7 BIOINFORMATICS

2.7.1 Homology detection

MS1 orthologs were detected using the protein query TBLASTN search against the annotated and early released genomes in Phytozome using the default settings (Goodstein et al., 2012), except for the MS1 ortholog in *Thellungiella parvula,* which was detected using the ortholog group annotation in PLAZA 3.0 (Proost et al., 2015). A 2.0 kb upstream sequence of the MS1 ATG start site was downloaded from Phytozome or PLAZA 3.0 and submitted to EARS (Picot et al., 2010) for evolutionary conserved non-coding sequence analysis. Then short DNA fragments within each CNS group were analysed through MEME (Bailey et al., 2015) with settings adjusted to search for up to four motifs of 30-150 bp in each CNS region. Sequence logos were downloaded from the html MEME output.

The analysed species in the comparative phylogenetic footprinting included 9 species from the Brassicaceae family (*Arabidopsis lyrata, Arabidopsis thaliana, Arabidopsis halleri, Boechera stricta, Brassica rapa, Capsella grandiflora, Capsella rubella, Eutrema salsugineum, Thellungiella parvula*) and 11 species from the Poaceae family (*Oryza sativa, Oryza brachyantha, Hordeum vulgare, Sorghum bicolor, Setaria italic, Brachypodium distachyon, Aegilops tauschii, Triticum aestivum, Triticum urartu, Panicum virgatum, Zea mays*).

Sequences of MS1 homologous genes were extracted from Phytozome database for omparative green plant genomics (Goodstein et al., 2012). The Phylogenetic tree was adapted from the MS1 orthologue group ORTHO000988 in PLAZA 3.0 (Proost et al.,

2015). Protein sequence alignment was conducted using ClutalOmega and PRALINE online applications (Simossis and Heringa, 2005; Sievers and Higgins, 2014).

2.7.2 Molecular modelling

The 3D molecular prediction of the structure of the MS1 PHD-H3K4me3 complex was built on the automated protein structure homology-modelling SWISS-MODEL server (Biasini et al., 2014 https://swissmodel.expasy.org/). The 612-661 aa protein sequence of MS1 was fed to the server to run blast identifying the most conserved template. The X-ray resolved crystal structure of the human MLL5 PHD finger in complex with H3K4me3 (SMTL ID 4l58.1.A, Ali et al., 2013) showed a Seq Similarity of 0.49 and gave a high Estimation value (0.79) of the Global Model Quality. Therefore, this 4l58.1.A was chosen to be modelling template and the final models of MS1 PHD-H3K4me3 was refined to 1.5 Å resolution. Threaded PDB outputs were visualized in PyMOL v1.7.0.3 (http://www.pymol.org).

2.7.3 Microarray data analysis

Normalised microarray data was retrieved from Arabidopsis Anther/Floral Expression Plots (https://www.cpib.ac.uk/anther/) and clustered using an online microarray data analyser tool CIMminer and default parameters for distance calculation and clustering (https://discover.nci.nih.gov/cimminer). During the clustering and heatmap analysis, the constitutively low-expressed and high expressed genes that clustered at the top and bottom of the heat map respectively, exceeds the microarray dynamic range. Therefore, these genes were ruled out of further analysis. Venn diagram and GO enrichment analysis was conducted using FUNRICH software (Pathan et al., 2015).

For tapetum highly expressed genes, a LMC microarray data (Vizcay Barrena, unpublished) specifically on tapetum cells at tetrad release and PMI stage was re-normalised using a robust microarray averaging procedure (Irizarry et al., 2003). This resulted in a signal saturation at the high-end value of 16 and background value at the low-end of 6. The signal overall average value is 8.7. An empirical value of 12 was used as the cut-off line to filter the tapetum high expressed genes, which generated 1,199 and 1,101 tapetum high expressed genes in tetrad release and PMI stage respectively.

2.7.4 Protein Domain Analysis

Graphical representation of the protein domains and PTMs was drawn using the Illustrator for Biological Sequences 1.0.1 version (Liu et al., 2015b). The *in-silico* ubiquitination sites identification was made via a web-based system (UbiSite http://csb.cse.yzu.edu.tw/UbiSite) (Huang et al., 2016). The nuclear localisation signal peptides were predicted using a combination of online analysis tool, NucPred (Brameier et al., 2007), cNLS Mapper (Kosugi et al., 2009) and NetNES version 1.1 (La Cour et al., 2004).

2.8 PROMOTER DISSECTION

Truncated or full-length versions of the MS1 promoter were PCR amplified from genomic DNA of Arabidopsis L*er* WT. The amplified DNA fragments were first ligated in into pDONR207, generating gateway ENTRY clones. Then the promoters were transferred pBGWFS7 GFP-GUS promoter reporter construct via gateway LR

reaction. Plants were transformed and screened using flower dip method based on procedures described in Zhang et al. (2006b). Floral buds at appropriate stages from positive transgenic plants were dissected to observe the GFP reporter signals under a Leica SP5 TCS confocal microscope.

2.9 YEAST ONE-HYBRID

Full length 2 kb intergenic non-coding DNA between AT5G22260 (MS1) and AT5g22270 (SIED1) excluding the UTR regions was PCR amplified as the template for yeast one-hybrid bait sequence analysis. Six ~430 bp sectors with 60 bp overlap with each other covering this 2-kb region of MS1 promoter were PCR amplified and cloned into the pCR8 gateway ENTRY vector. Similarly, full-length upstream region of MYB99 and AMS were amplified and partitioned into ~400 bp regions and cloned into the pCR8 gateway ENTRY vector. All these yeast one-hybrid baits DNAs were transferred into yeast bait vector pHIS2LEU2GW (University of Warwick) via LR gateway reaction. The CDS region of selected transcription factors were PCR amplified using cDNA extracted from Arabidopsis inflorescence as template and ligated into pCR8 gateway ENTRY clone and LR recombined into pDEST22 (Thermo Fisher Scientific Company) as yeast one-hybrid prey vector.

Yeast strain Y187 (mating type α) and AH109 (mating type a) were cultured and transformed with the constructed bait and prey yeast vectors, respectively, based on the procedures described in Gietz and Schiestl (2007). Yeast transformants were colony-PCR verified and cultured for the hybrid mating. The mating colonies were incubated overnight 30 °C and replicated onto the selection plates using a velvet replication block. The selection plates were arranged as follows:

- SD−2 (−Leu, −Trp), for confirmation of the yeast strain mating
- SD−3 (−Leu, −Trp, −His), for screening the activation of HIS3 reporter
- SD−3 (−Leu, −Trp, −His, with 5, 10, 50 mM 3-AT added), for screening the activation of HIS3 reporter with auto-activation repressed.

The Arabidopsis Circadian Clock transcription factor *CCA1 HIKING EXPEDITION* (*CHE*) and the promoter region of a morning-expressed MYB transcription factor *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) were used as positive controls in the yeast one hybrid as there has been previous reports of strong protein-DNA interaction between them in yeast one-hybrid system (Pruneda-Paz et al., 2009).

2.10 CHROMATIN IMMUNOPRECIPITATION ASSAYS (CHIP)

The ChIP experiments were carried out based on the procedures described in Kaufmann et al. (2010) except the following modifications.

- 4 g initial plant inflorescence tissue from MS1-GFP and L*er* WT lines with opened floral buds removed was used for each precipitation group.
- Three fixate (crosslinking) conditions were used and tested to optimise the ChIP output, *i.e.* fixation in MC buffer for 10 minutes, 15 minutes and fixation in deionised water for 20 minutes.
- Before the immunoprecipitation step, aliquots of the isolated and sheared chromatin were tested in western blot using anti-GFP antibody to verify if tapetal chromatin containing MS1-GFP was not depleted or lost during the early operations.

- An alpaca derived GFP antibody coupled to magnetic microparticles (ChromoTek GFP-Trap®) was used as the antibody to pull down the MS1-GFP bound DNAs.

- qPCR of the eluted DNAs from the ChIP was conducted on Light Cycler 480 Real-Time PCR system (Roche Applied Science) and the Ct values were analysed by a free qPCR statistical analysis software pyQPCR (http://pyqpcr.sourceforge.net/). Percent input method was used to analyse and show the level of ChIP DNA enrichment at a given locus.

- The eukaryotic initiation factor (EIF4), ACTIN2, and retrotransposon gene TA3 were used as ChIP negative control locus (Yamaguchi et al., 2014).

2.11 RETRIEVAL AND EXPRESSION DETECTION OF MS1 LNCRNA

AGI code of MS1, AT5G22260, was used to retrieve the putative MS1 lncRNA sequence from PLncDB (http://chualab.rockefeller.edu/gbrowse2/homepage.html). A window of 3 kb flanking sequence of MS1 on each side was adopted in the genome browser for the putative MS1 lncRNA loci.

A routine strand-specific RT-PCR was carried out to verify the transcription and expression pattern of lncRNAs. PrimeScript RT Reagent Kit was used to conduct the two-step reverse transcription in accordance to the manufacturer's instructions. Strand-specific primers (ssp) were designed and used instead of oligo dT primer to specifically anneal to the 3'-end of each MS1 lncRNAs in the reverse transcription

reaction. RT-PCR primers were designed to cover the 100-150 bp region, which spans an intron-exon junction if possible. The primers used are listed as follow (**Table 2.1**).

Table 2.1 The ssp primer list for RT-PCR detection of the *MS1 lncRNAs*.

Primer Name	Primer Sequence
MS1-lncRNA1-RTF	CAGCCACTAACATGCAGACATCAA
MS1-lncRNA1-RTssp	TCTCATCTTGCGTG
MS1-lncRNA1-RTR	GTTCTCGGGTTCCAAAGTAGCTTTAC
MS1-lncRNA2-RTF	TTTCGCTGACGTAAGACCTATCAAAGAA
MS1-lncRNA2-RTssp	AACAACACGTATCCT
MS1-lncRNA2-RTR	CCAGCGAGGCAGCGATACATTA
MS1-lncRNA3-RTF	ACCCTTAACACACTTCCTCAACACTT
MS1-lncRNA3-RTssp	GTGTGACCATCATCT
MS1-lncRNA3-RTR	TTGGTTTGATCTTGAGAAGAGAGGGAATAG
MS1-lncRNA4-RTF	CGGATTTGGTCACTTGCTCTCTCT
MS1-lncRNA4-RTssp	CTCAAAGGAAACAATG
MS1-lncRNA4-RTR	AGTCTAATTAGTACCTGGCCTTTAAACCG

CHAPTER 3 MULTILEVEL REGULATION ON THE EXPRESSION OF MS1

3.1 INTRODUCTION

Pollen develops in the anther locules, with layers of nourish and secretory somatic cells surrounded. The innermost lay is called the tapetum, whose normal development and function is crucial for the formation of viable pollen grains (Zhang and Yang, 2014). The development of the tapetum is strictly regulated. A vast array of small RNAs, phytohormone receptors, protein ligand-receptors as well as transcription factors are involved in regulating the specification, differentiation and specific functions of the tapetum (Feng et al., 2013; Liu and Fan, 2013). *MALE STERILITY1* (*MS1*), which encodes a putative PHD-type transcription factor, plays a key role in late tapetum regulation through directing the middle-to-late phase transition of the tapetum (Ito et al., 2007; Yang et al., 2007; Vizcay-Barrena and Wilson, 2006; Ariizumi et al., 2005). Absence of MS1 arrests tapetum development during the end of the middle phase, thereby preventing tapetum from initiating late tapetal biological processes, such as programmed cell death, pollen wall precursor synthesis and the differentiation of organanlles essential for pollen coat formation (Vizcay-Barrena and Wilson, 2006). Expression profiling of the *ms1* mutant is consistent with the regulatory role of *MS1*: a large number of genes that are normally expressed within the meiotic tapetum show a significantly prolonged expression while late tapetum (postmeiotic) genes show dramatic down-regulation compared with the same stage WT anther (Yang et al., 2007). Li et al. (2011) speculated that in the rice *ms1/ptc1* mutant, middle-phase tapetum persists longer than that of the WT, and fails to reach

its late development phase and then eventually goes through necrotic breakdown. Hence, MS1 appears to act as a primary switch for the middle-to-late tapetal development programme. Understanding the genetic control of MS1 is key to uncover the regulation network for late tapetum development.

3.1.1 A contradictory expression pattern of MS1

The expression of MS1 is restricted to the tapetum during the tetrad release stage. However, the transcriptional and translational fusions show different expression patterns regarding the duration of expression. For example, a functional translational GFP fusion of MS1 using its full length genomic sequence (from 2.9 kb upstream to 329 bp downstream) only showed GFP signal in the nuclei of the tapetal cells for about 12 hours during tetrad release (**Figure 3.1 E**). Anthers expressing MS1-GFP could only be detected in one bud within an inflorescence as tetrad release is a transient process during anther development (**Figure 3.1 C and E**) (Yang et al., 2007). However, a GUS transcriptional reporter driven by a 3-kb upstream sequence of MS1 displayed a more prolonged expression starting from the tetrad to pollen mitosis stage, showing GUS signals in several buds within one inflorescence (**Figure 3.1 A**) (Ito and Shinozaki, 2002). An *in-situ* hybridization expression analyses of *MS1* showed similar expression pattern to the *MS1-GFP* fusion reporter (**Figure 3.1 B**) (Ito and Shinozaki, 2002), nevertheless, *in-situ* hybridization results from Zhu et al. (2011) were more consistent with the GUS reporter signals, showing the expression of *MS1* begins at tetrads release stage and persists until microgametogenesis (**Figure 3.1 G-J**).

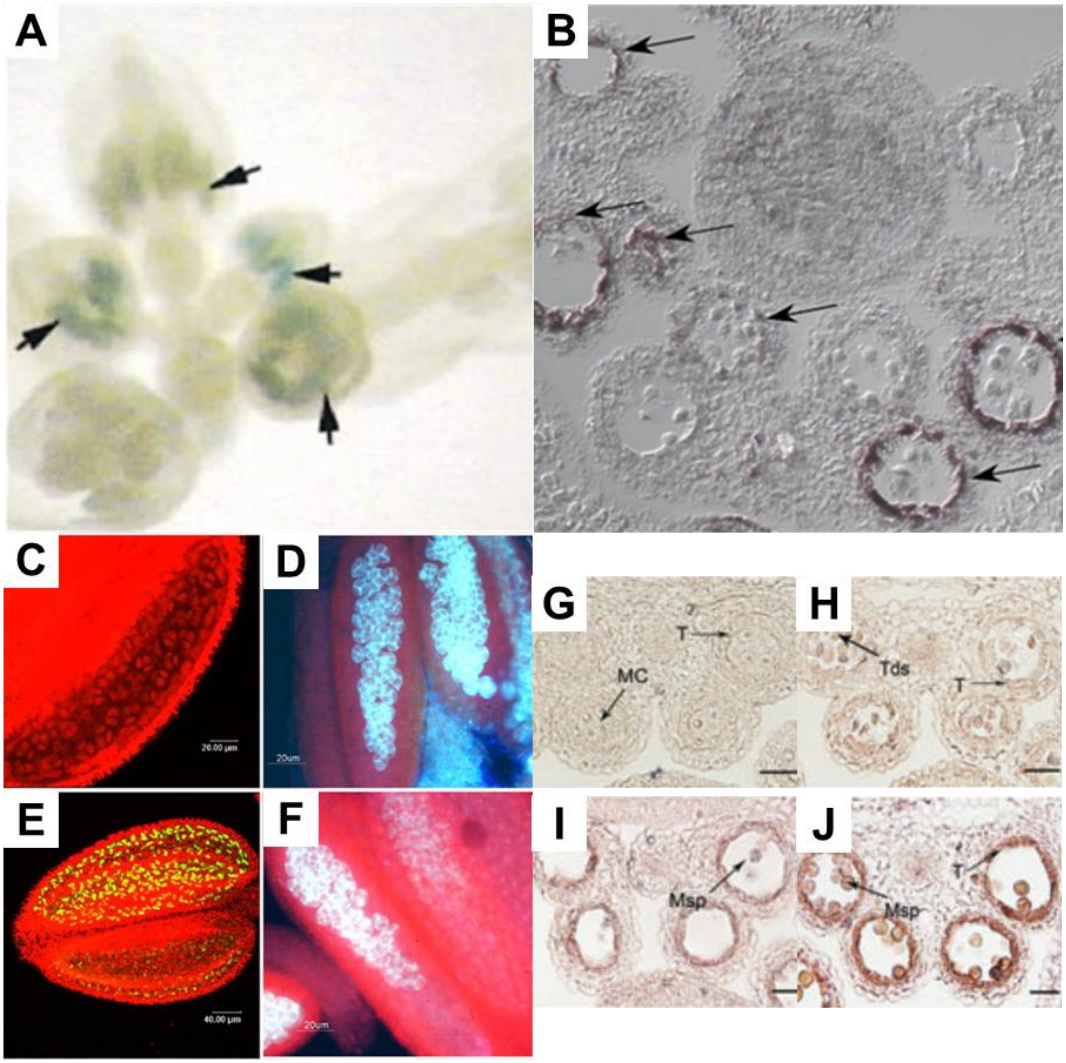


Figure 3.1 Expression patterns of MS1 from various sources.

(A) Whole mounts of entire L*er* inflorescence showing specific GUS staining using a *MS1* promoter fusion (arrows) (Wilson et al., 2001). (B) In situ hybridization analysis of *MS1* adapted from (Ito and Shinozaki, 2002). (C-F) Functional MS1-GFP fusion protein expressing in anther of transgenic lines with aniline blue staining of the callose wall. Figure adapted from (Yang et al., 2007). MS1-GFP is not expressed during early tetrad stage (C and D); at late tetrad stage, MS1-GFP can be detected when the breakdown of callose wall initiates (E and F). (G-J) In situ hybridization analysis of *MS1* showing a prolonged expression from late tetrad stage until pollen mitosis occurs. Figure adapted from (Zhu et al., 2011). Transverse sections of anthers of Stage 6 (G), Stage 7 (H), Stage 8 (I) and Stage 10 (J).

Although the *in-situ* hybridization data are slightly inconsistent, the most informative clue on the regulation of MS1 can be extracted from the MS1-GFP fusion protein as

this can fully rescue the *ms1* mutant phenotype whilst other aspects of the plant development are not affected. Combined with the fact that the MS1 transcriptional reporter shows a different expression pattern from that of the translational reporter, it can be deduced that at least three levels of regulation give the characteristic MS1 expression, namely, an activation of MS1 expression soon after tetrad release, a rapid removal of the protein after a transient expression and transcriptional repression.

3.1.1.1 Activation of MS1

Transcriptional activation in plant is commonly achieved via the recognition and binding of transcription factors to transcription factor binding sites (TFBSs) within gene upstream regions. Several TFBSs has been reported as important for tapetum specific gene expression (Farquharson, 2016; Xiong et al., 2016; Shumin et al., 2015; Liu et al., 2013; Xu et al., 2014a; Ye et al., 2010). As tapetum-specific genes often exert unique transcriptional profiles (Aya et al., 2011b; Pearce et al., 2015; Ehlting et al., 2008), it is most likely that various TFBSs act together to ensure the specific expression in tapetum. For example, a 513 bp upstream sequence of DYT1 was reported as essential for its early tapetum expression. A truncation of the −513 to −506 bp region weakened the expression level but the temporal and spatial pattern of DYT1 expression was maintained and a novel 'CTCC' DNA motif localized at −455 to −452 bp region was showed to be essential for the early tapetum activation of DYT1 (Shumin et al., 2015). A rice tapetum-specific lipid transfer protein (LTP), *OsLTP6*, achieved its characteristic expression pattern through a combination of positive and negative regulatory elements in its upstream sequence. A 26-bp motif upstream of TATA box was key for the transcriptional activation of *OsLTP6* whilst a CAAT box and a GTGA box located between −310 to −350 bp upstream enhanced the

transcription level of OsLTP6. Additionally, two negative regulatory elements exist at two distinct regions, preventing the leaky expression of *OsLTP6* in leaf and stem.

MS1 is under the regulation of a vast array of TFs (reviewed in **Section 1.2.2**), and a series of TFBSs on the upstream sequence of MS1 were also reported as important for the activation of MS1 expression (Farquharson, 2016; Cui et al., 2016; Ye et al., 2010; Xiong et al., 2016; Phan et al., 2012). However, no regulatory model has integrated all these factors and some of the reports of regulation are unclear and quite ambiguous.

In the BR biosynthetic and signalling mutants, the expression of a series of tapetum specific genes are significantly reduced, including *MS1* and *MALE STERILITY2* (*MS2*), which encodes for a tapetum-specific fatty acid reductase and the mutant of which also cause male sterility (Ye et al., 2010). ChIP-PCR experiments showed that BES1 directly bound to the promoter regions of MS1 and MS2, which indicates that the BR induction of the *MS1/MS2* expression is probably due to direct activation by *BES1* (Ye et al., 2010). *BES1*, together with its homologue *BRASSINAZOLE-RESISTANT1* (*BZR1*), are both the important BR signalling effectors that mediate the transcriptional response to BR via their capacity to bind to the E-box (CANNTG) and/or to the BR response element (BRRE, CGTGT/CG) upstream the BR responsive genes (Mar??chal and Brisson, 2010). Ye et al. (2010) claimed that BES1 directly activates the expression of both *MS1* and *MS2* by binding to its upstream E-box based on the ChIP-PCR data. It is worth noting however that, a BRRE and E-box enriched region (−664 to −452) of the *MS1* promoter was not tested in their BES1 ChIP-PCR experiment. On the contrary, a distant upstream region (−1121 to −952 bp) of MS1 which only possesses one E-box on the boundary of the PCR amplicon was tested and a moderate enrichment (two-fold-change to the control)

was obtained. This might explain why ChIP DNA enrichment on MS1 is relatively moderate compared with that of the MS2 (six-fold-change to the control). Moreover, the BES1 protein that binds to the upstream of MS1 is not sufficient to activate MS1 expression, as in the BES1 overexpression line MS1 expression is not significantly changed (Ye et al., 2010), suggesting other factors, possibly BZR1, are necessary for the BR induced MS1 expression.

Cui et al., (2016) and Gu et al., (2014) detected that the −504 to −485 bp upstream region of *MS1,* possessing a *DYT1* specific E-box, was bound by DYT1 *in vitro* in an electrophoretic mobility shift assay (EMSA), in which the addition of the bHLH089 protein was able to enhance protein-DNA interaction. It is believed that the heterodimer complex formed between *DYT1* and any one of *bHLH010/089/091* can enhance the transcription activation activity of DYT1 via promoting its localization to the nucleus (Farquharson, 2016). In an *in vivo* transient dual-luciferase assay, high level of *MS1* expression was activated by the DYT1-bHLH091 complex (Cui et al., 2016). Although these direct regulation data are sound and consistent with the RNA-SEQ or microarray profiling of mutants, the binding of DYT1 and bHLH010/089/091 on MS1 promoter are essential, but not sufficient for activating MS1 normal expression, as in each single mutant of AMS and MYB80, the MS1 expression is lost, while DYT1 and bHLH010/089/091 are normally expressed. Also, DYT1 and bHLH010/089/091 expression occurs at a quite early stage, whether these proteins persist to the tetrad stage tapetum and thus able to activate *MS1* expression is questionable.

AMS was reported as a primary regulator binding to E-Box (CANNTG) motif of promoter sequence of a series of tapetum specific genes and coordinating their

expression. MS1 is not expressed in the *ams* mutant, however, the direct regulation of MS1 by AMS is not described yet. AMS is a homologous gene of DYT1 and as well interacts with bHLH010/089/091 respectively (Xu et al., 2010), hence it is likely that AMS direct regulates *MS1* via forming complexes with bHLH010/089/091. On the other hand, Phan et al. (2011) reported that the MYB80 has a high binding affinity to an MYB core-motif (CCAACCA), which exits at the upstream sequence of MS1 (−662 to −657 bp). Although in their work, the ChIP-PCR of MYB80 protein failed to obtain any binding signals on the promoter sequence (−405 to −37 bp) of MS1, genuine positive binding might have been missed as the ChIP-PCR detected region did not cover the MYB80 binding motif. More recently, *CYP703A2* was reported as a direct target of MYB80 and AMS facilities this regulation by forming a complex with MYB80 and together binding to a region of *CYP703A2* promoter where E-box and degenerated MYB-core motif (AACC) are enriched. Since *MS1* is highly co-expressed with *CYP703A2*, and the E-box and MYB-core Motif are also enriched at the upstream sequence of MS1, it is possible that MYB80 and AMS regulate the expression of MS1 in an analogous way.

Taken together, the abovementioned TFs of early or middle phase tapetum development regulate MS1 expression via context-dependent binding to the TF-binding motifs in the upstream sequence of MS1. Despite being direct regulated by a series of TFs, the basis of the fact that MS1 is not expressed until the tetrad release stage is poorly understood.

3.1.1.2 Removal of The MS1 Protein

Rapid removal of the MS1 protein was observed in a translational MS1-GFP reporter line, and this rapid removal of the protein is possibly required for normal tapetum development progress since the overexpression of MS1 in Arabidopsis or barley both can cause severe fertility reduction, which suggests that excess MS1 protein is not beneficial for normal development of anther cells (Fernández Gómez and Wilson, 2014; Yang et al., 2007). Also, the male sterile phenotype of *ms1* cannot be rescued by overexpression of MS1 (Yang et al., 2007), suggesting that the presence of MS1 protein must be tightly restricted to the specific stage of tapetum to fulfil its regulating role. This contrasts with other tapetal-specific transcription factors, for instance, AMS, which has a long duration of expression throughout tapetal development (Xu et al., 2010; Ferguson et al., 2016).

Ubiquitin-mediated proteolysis is the main non-lysosomal proteolytic pathway for the rapid degradation of intracellular proteins in eukaryotic cells. A previous yeast two-hybrid screen and pull-down assay identified and confirmed a MS1 interacting protein LIGHT-RESPONSE BTB 1 (LRB1), which is a nuclear-localized E3 ubiquitin ligase and was speculated to mediate the MS1 protein degradation (Ni et al., 2014; Yu, 2015). However, knock out analysis of the T-DNA insertion line of *LRB1* showed no observable phenotype in the male fertility (Yu, 2015). More recently, LRB1 and its two conserved homologs, LRB2 and LRB3 were found to dimerize and recruit a nuclear-localized complex of photoreceptor and photo-effector to form an CUL3-LRB complex for rapid polyubiquitination and subsequent degradation via the 26S proteasome pathway (Ni et al., 2014). The finding of the function and redundancy of LRBs sheds light on understanding the mechanism for the rapid removing of MS1

protein in tapetum. Considering the biological significance of rapid degradation of MS1, the lack of phenotype in *lrb1* can be explained by the redundancy among LRBs. Therefore, a phenotypic study on the tapetum and pollen wall in the *lrb1lrb2lrb3* triple-mutant (*lrb123*) would be a most straightforward way to uncover if the interaction between MS1 and LRBs play a part in the regulation of MS1 *in vivo*.

3.1.1.3 Auto-repression of MS1

In the tapetum, in contrast to the well-studied transcriptional activations, transcriptional repression is barely reported. The MS1 induced transcription factors, such as *WRKY DNA-BINDING PROTEIN 43* (*WRKY43*) and *AT3G53310* (*AP2/B3-like transcriptional factor*), showed no observable phenotype in their T-DNA insertion knockout lines (unpublished data in our group). Though it is hard to identify *MS1* transcriptional repressors in *MS1* downstream targets, the auto-repression of *MS1* might offers clues as to how *MS1* is transcriptional repressed.

All nine *ms1* mutant alleles (summarised in **Table 1.3**) showed a prolonged and upregulated expression of the mutant *ms1* transcript compared to the WT *MS1* (**Figure 3.2 A**). Whereas in the heterozygous *MS1/ms1* plants, the mutant transcript expression level was similar to the WT, which indicates that the MS1 protein is necessary for the auto-repression (Yang et al., 2007). On the other hand, the mutant transcript expression could be repressed to the normal level by the complementary MS1-GFP protein driven by 2.9 kb upstream sequence and the gene body (regions spanning gene models) of MS1; however in the cauliflower mosaic virus (CaMV) 35S promoter driven MS1(CDS)-GFP overexpression line, this auto-repression disappeared (Yang et al., 2007). This suggests the full-length genomic sequence of

MS1 and/or its upstream sequence are also essential for MS1 auto-repression. Interestingly, the *ms1* alleles showed varying levels of the *ms1* mutant transcripts. In *ms1-1* and *ms1-6* mutant, *ms1* expression level ranges up to 35-fold to that of the WT where as in *ms1-2* to *ms1-5*, the expression ranges from 5 to 10-fold to the WT (**Figure 3.2 A**) (Yang et al., 2007). Based on the relative positions of SNPs or insertions of the *ms1* mutants, the open reading frame (ORF) of MS1 encountered a series of premature stop codon. The *ms1* gene products were deduced to be a series of truncated MS1 proteins of different length and all without the PHD finger (**Figure 3.2 D**).

The auto-repression regulation of tapetum key TFs, including *PTC1* (**Figure 3.2 B**) (Li et al., 2011), *MYB80* (Xu et al., 2014b) and *DYT1* (Cui et al., 2016; Zhu et al., 2015a), have been widely reported, providing a suggestion that this auto-repression might be of a critical biological significance in tapetum development. Yet the molecular mechanism controlling this remains poorly understood. Especially in the case of MS1, it is not clear whether MS1 is directly auto-repressed and by what underlying mechanism that some alleles have a greater effect than the others on the MS1 auto-repression.

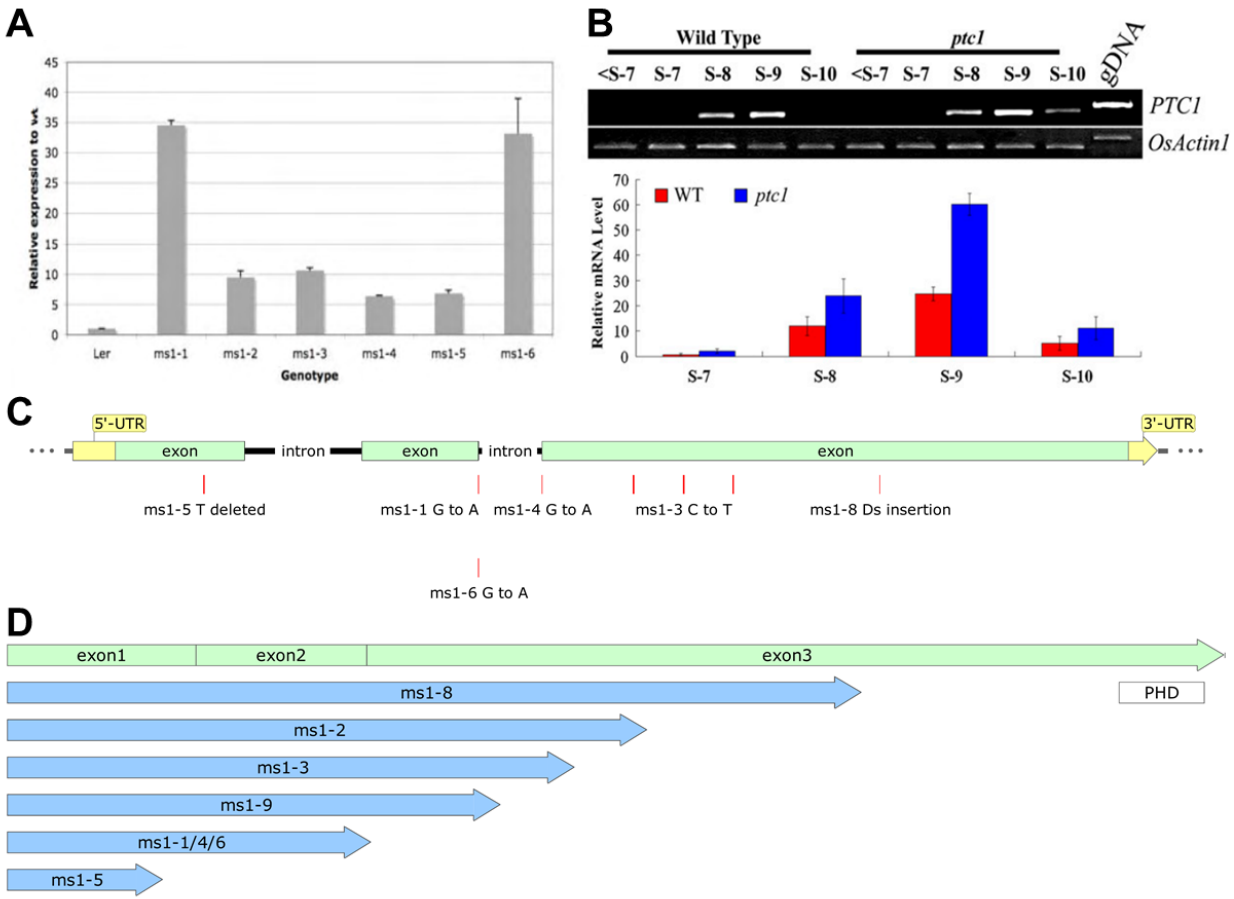


Figure 3.2 The expression pattern of MS1/PTC1 alleles and the MS1 gene map (Yang et al., 2007; Li et al., 2011).

(A) Real time RT-PCR detection of MS1 expression in L*er* and ms1 alleles. Comparative expression is shown relative to wt L*er* expression. Figure taken from Yang et al. (2007). (B) RT-PCR and real time RT-PCR detection of PTC1 expression in rice staged anther of WT and *ptc1* mutant. Comparative expression is shown relative to the expression of *wt* stage 7 anther. Figure taken from (Li et al., 2011). (C) The gene structure of MS1 showing SNPs and insertions of *ms1* mutant alleles. (D) Schematic diagram showing the deduced gene products of *ms1* alleles.

3.1.2 Long non-coding RNAs

Long non-coding RNAs (lncRNAs) are a class of non-coding RNAs (ncRNAs) longer than 200 nt (Liu et al., 2015a). Early transcriptome studies have found the transcription of lncRNAs covering over 90% of the genome across eukaryotes (Böhmdorfer and Wierzbicki, 2015; Kung et al., 2013). In the early studies, there was controversy on whether the pervasively expressed lncRNAs played biological roles or were just genome-wide transcriptionl noise due to their low expression level, however, remarkable progress has been made revealing the critical roles of lncRNAs in diverse biological processes across eukaryotes (Liu et al., 2015a). In plants, many lncRNAs have been reported as key regulators of gene expression, however, the regulation mechanisms are varied. These regulating lncRNAs can function either *in-cis* or *in-trans* by seuqence homology or complementarity to DNA or RNAs, or by their RNA secondary structure, serving as scaffolds for macromolecular assembly (Chekanova, 2015). The lncRNA regulation of gene expression mostly functions at the transcriptional level, however, some exceptional cases have been reported where the lncRNAs regulate gene expression posttranscriptionally (Böhmdorfer and Wierzbicki, 2015).

Wang et al., (2014) identified ~40,000 putative lncRNAs in *Arabidopsis thaliana* by analysing over 200 transcriptomic data sets. The genomic information of identified Arabidopsis lncRNAs were collected in an online Plant long non-coding RNA database (PLncDB), which offers a genome browser for plant lncRNAs search and retrieval (Jin et al., 2013). At the genome locus of *MS1* in the PLncDB genomic browser, four putative lncRNAs were found, namely, *AT5TU027070* and

AT5TU027080, both of which were transcribed from the anti-sense strand of *MS1*, *AT5NC027090*, transcribed from the upstream intergenic region of *MS1* and *AT5TU027100*, transcribed from the antisense strand of the *AT5G22270* gene (**Figure 3.3**). In this thesis, these four lncRNAs will be overall referred to as 'MS1 lncRNAs' and separately as *MS1 lncRNA1, 2, 3, 4* based on their genomic loci in the order of 5' to 3' of MS1 (**Figure 3.3**).

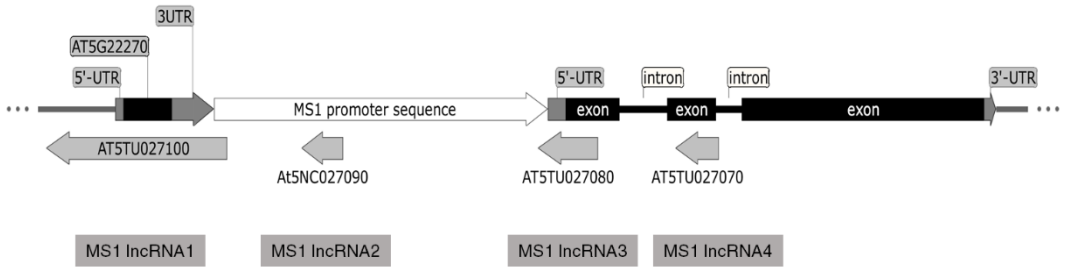


Figure 3.3 Genomic view of the four MS1 lncRNAs transcribed from the genome locus of MS1.

Although it has been a short time since the discovery of lncRNA, the study of lncRNA has significantly expanded our knowledge of gene expression regulation. Considering the tight regulation of *MS1* expression, the lncRNAs transcribed at the MS1 locus might also play a role in ensuring this strict control. In this work, a strand-specific RT-PCR verifying the expression pattern of these predicted lncRNAs was carried out, which is a prerequisite for characterising the lncRNA mediated regulation of *MS1*.

3.1.3 Summary and aims

Although to date there is no general agreement on the regulatory mechanism of *MS1*, at least three level of regulations on MS1 can be deduced from previous translational and transcriptional reporter of MS1, *i.e.* an activation of *MS1* soon after tetrads

release, a rapidly removal of the protein after a short-time of expression and transcriptional repression. Additionally, the recently identified lncRNAs transcribed at the locus of *MS1* might also be involved in this regulation. Understanding the regulation of *MS1* is key to reveal the control mechanism of late tapetum phase transition and late tapetum development.

Here in this chapter, through a phylogenetic footprinting motif search, promoter dissection assay, yeast one-hybrid assay, strand-specific RT-PCR, ChIP-PCR and SEM precise phenotyping, the following fact were uncovered re the multilevel regulation of *MS1*:

a) An upstream conserved non-coding sequence (CNS) is essential for the expression of MS1.

b) The tapetum-specific transcription factors that act upstream of *MS1* (*DYT1*, *TDF1*, *AMS*, *BES1* and *BZR1*) could not bind to any of the CREs of the *MS1* promoter in yeast.

c) MS1 is directly auto-repressed via a binding on the junction of second exon and first intron, where a lncRNA is transcribed.

d) The mutant of LRB123, a MS1 *in vitro* interacting protein which may participate in the MS1 protein degradation, shows a smooth tapetum membrane phenotype.

3.2 RESULTS

3.2.1 Phylogenetic footprinting of the upstream sequence of MS1

TFs regulate gene expression by binding to TFBSs, of which the identification is highly challenging due to the prevalence of great numbers of non-functional TFBSs within the plant genome (Van de Velde et al., 2016). Phylogenetic footprinting, is a powerful comparative genomics approach, which is used for the identification of evolutionarily conserved noncoding sequences (CNSs), which are functionally constrained and can significantly help in eliminating false-positive TFBSs (Van de Velde et al., 2014).

In-vitro evidence strongly supported that *MS1* is directly regulated by an array of TFs, including *DYT1*, *bHLH010*/*089*/*091*, *BES1*, *MYB80* and *AMS*, however, the functional TFBSs that these TFs bind to exerting the regulation of *MS1* are not clear. Therefore, a phylogenetic footprinting analysis was conducted using the promoter regions of MS1 orthologs in a number of genome sequenced Brassicaceae species. Species from the Poaceae family were also included in the analysis initially, as rice and barley MS1 orthologs have been also reported to play a conserved regulation role in tapetum development (Fernández Gómez and Wilson, 2014; Li et al., 2011). The 2.0 kb upstream sequence of ATG start codon from 9 Brassicaceae species including *Arabidopsis thaliana* and 11 species from the Poaceae family including rice (*Oryza Sativa*) and barley (*Hordeum vulgare*) were obtained and submitted online to the EARS server (Picot et al., 2010) for an evolutionary analysis of conserved non-coding sequence (CNS). Although rice *PTC1* and barley *HvMS1* showed a very similar expression pattern with Arabidopsis MS1, no conservation was detected in the

upstream sequence of MS1 or its orthologs between Brassicaceae and Poaceae family (data not shown). However, significantly conserved sequences (p-value smaller than 10^{-24}) were detected both at the adjacent and distal region to the ATG start site within the Brassicaceae and Poaceae family respectively (**Figure 3.4**). In Brassicaceae comparative analysis, three outstanding CNSs were found (**Figure 3.4 A**). Two adjacent CNSs ranged from −50 to −300 bp and −350 to −600 bp were obtained, showing quite sharp p-value mean boundaries and a relatively even distribution within the CNS region, which suggest a long continuous consensus sequence might present adjacent to the ATG start site across Brassicaceae MS1 orthologs. A distal CNS (dCNS) ranged from around −1450 to −1900 bp was identified, which flanked the 3' end of an adjacent gene AT5G22270 (*SIED1*). This CNS showed varying conservation across the region, consisting of four sub-conserved regions (**Figure 3.4 A**). Since CNSs adjacent to the ATG start codon often play a critical role in the regulation of gene expression, the two adjacent CNSs of MS1 will be referred to as core CNS 1 (cCNS1) and core CNS 2 (cCNS1) in the rest of this thesis. The CNS identified within the Poaceae family using the *PTC1* promoter as template is highly distinct from that of the Brassicaceae family (**Figure 3.4 B**). Only one ATG adjacent CNS (core CNS) showed significant conservation while a distal conserved region extended into the coding sequence of an upstream gene, which makes it less likely to have a CRE role. The sharp peak of the cCNS p-value mean indicates that a short and unique consensus sequence might present at the centre of PTC1 core CNS (**Figure 3.4 B**).

Since the upstream CNS distribution of MS1 and PTC1 were highly different, suggesting the regulation mechanism of the MS1 in Arabidopsis and PTC1 in rice might have diverged after the monocot–dicot separation, the promoter sequence of Poaceae family species were no longer used in the later analysis. To predict the

potential CREs over-represented in the MS1 promoter, the promoter sequence of MS1 orthologs, covered by cCNS1, cCNS2 and dCNS were submitted to MEME (Bailey et al., 2015) for DNA motif analysis.

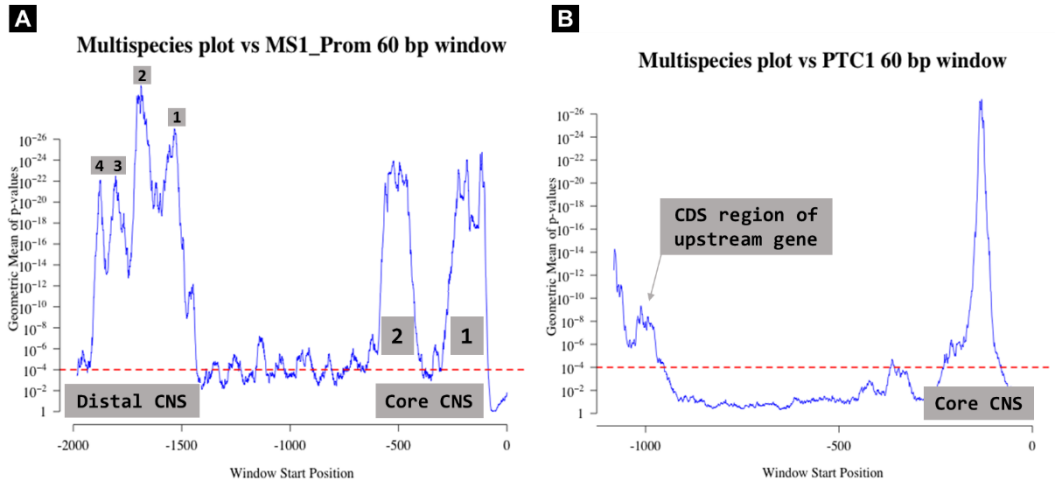


Figure 3.4 Comparative analysis of promoter regions of MS1 orthologs.

Multispecies plots showing comparative analysis of the MS1 ortholog promoter within Brassicaceae family (A) and Poaceae family (B) versus promoters of MS1 and PTC1 respectively, performed with the EARS tool. The red dashed line indicates the default cut-off P-value (0.0001).

Strikingly, for each CNS, unusually long motifs with a highly-conserved sequence were identified ranging from 30 to 70 bp (only motifs with E-value smaller than 10^{-50} were kept (**Figure 3.5** to **Figure 3.8**). Unlike animals, the CRE sequence motif of plants are usually relatively short ranging around 6 to 20 bp (Franco-Zorrilla et al., 2014), long and highly consensus motif as in the MS1 CNSs, might present additional regulatory function besides the TFBSs. To further dissect the function of these motifs, each motif sequence was submitted to a TFBS scanning programme in PlantPAN 2.0 (Chow et al., 2016). A MYB binding motif was found located at cCNS1 motif 2 with a bHLH E-box motifs on its 3'-side (**Figure 3.9**), which might present a binding site for the AMS-MYB80 complex (Xiong et al., 2016). A highly conserved DYT1 motif

and bHLH E-box binding motif were found in cCNS2 motif 2 (**Figure 3.10**), which is consist with previously reported DYT1-bHLH010/089/091 complex binding to the MS1 upstream sequence (Cui et al., 2016). A GAGA-repeats sequence variant was also found in cCNS1 motif 3 (**Figure 3.11**), which is the binding motif of the BPC family protein. Interestingly, there is a 30 bp GA-rich motif that locates to the first intron of MS1, resembling the identified GAGA-repeats. To date, there are no reports on the regulation of MS1 by BPC family members, however, the recurrence of the GAGA-repeats flanking the TSS of MS1 might provide a clue that MS1 is also under regulation of BPCs.

Here lists the DNA sequence of the 30 bp GAGA-repeats MS1 intron:
AGAGAGAGAGAAGGGAAAAGAGAGAGAGGA
and the 36 bp GAGA-repeats identified in MS1 core CNS2 motif 3:
GATGAAAAGGAAAAGTGAGTGAGAGAAAGAGAAAGG.

Although no TFBS was identified for the rest of MS1 CNS motifs, the prominent level of the sequence conservation and the unusual width of the motif imply their evolutionarily constrained function.

The BRRE-box was not identified among the CNS regions. However, in the regions of MS1 upstream sequence beyond the CNSs, three BRRE-box sequences exist. In the upstream sequences of Brassicaceae MS1 orthologs, 3-5 BRRE-Box sequences were detected beyond the CNS regions, while their relative distances to the ATG codon are varied (data not shown).

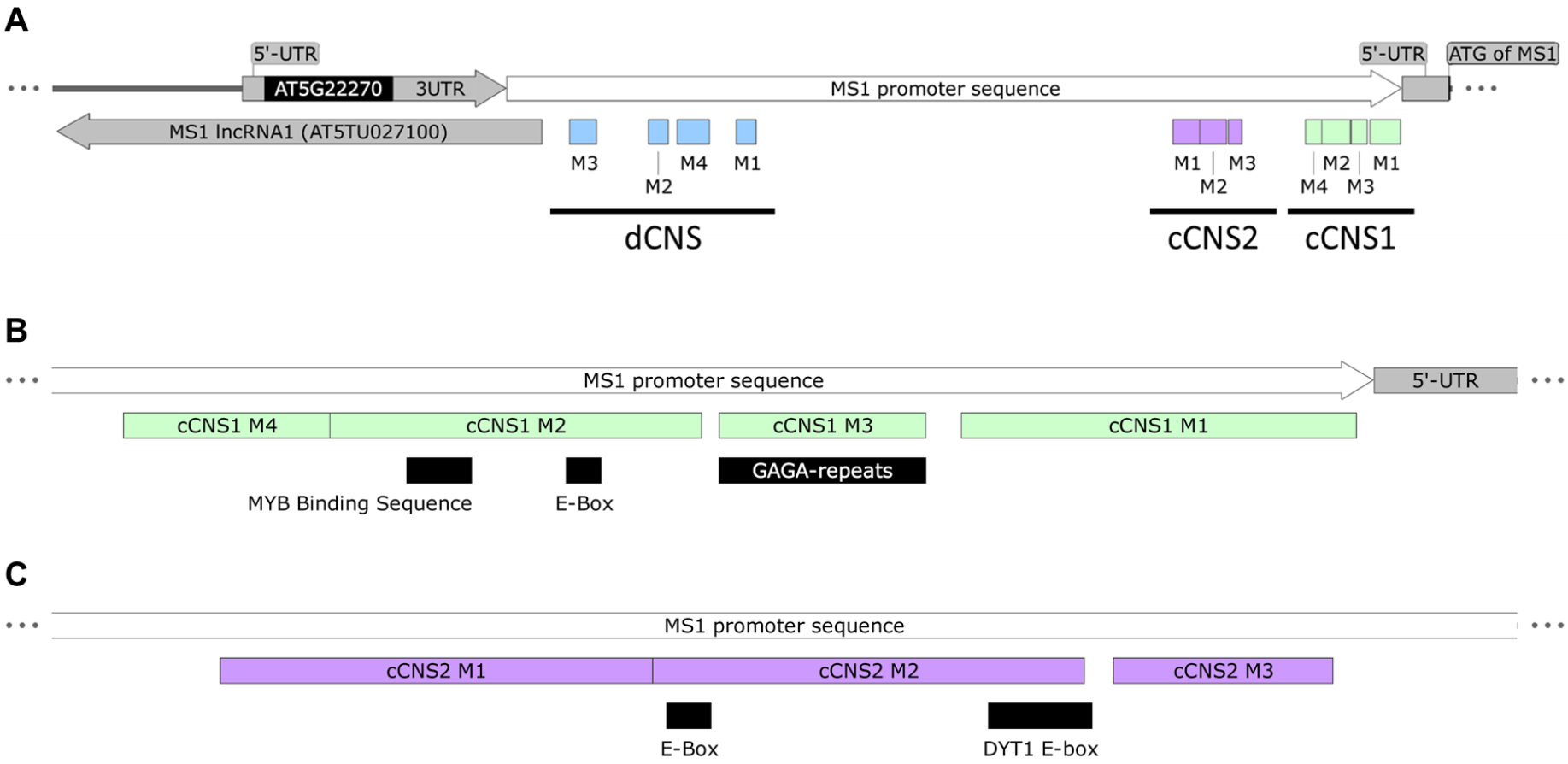


Figure 3.5 The overall schematic view of positions of the MEME analysed motifs and PlantPAN identified TFBSs within the MS1 distal CNS and two core CNSs.

(A) The 3-kb upstream sequence of MS1 showing the core CNS motifs are closely adjacent to each other while the distal CNS motifs are relatively separate; CNS regions are indicated by black lines with labels, motifs of cCNS1, cCNS2 and dCNS are indicated with green, purple and blue boxes respectively with numbered labels. (B) A zoomed view of core CNS1 showing the MYB-binding sequence and an E-Box is located at the core CNS motif 2, and the entire core CNS motif 3 represents GAGA-repeats. (C) A zoomed view of cCNS2 showing the motif 2 possesses a canonical E-Box and a DYT1 E-Box.

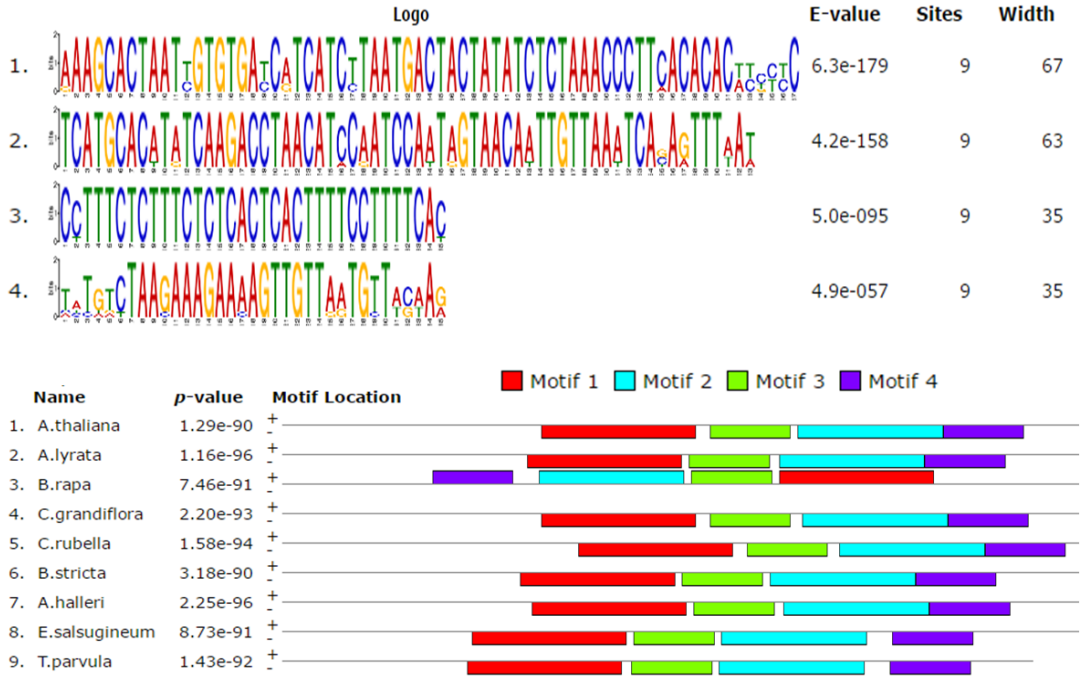


Figure 3.6 The core CNS1 of MS1 contains four consensus motifs generated by the online application of MEME 4.11.2.

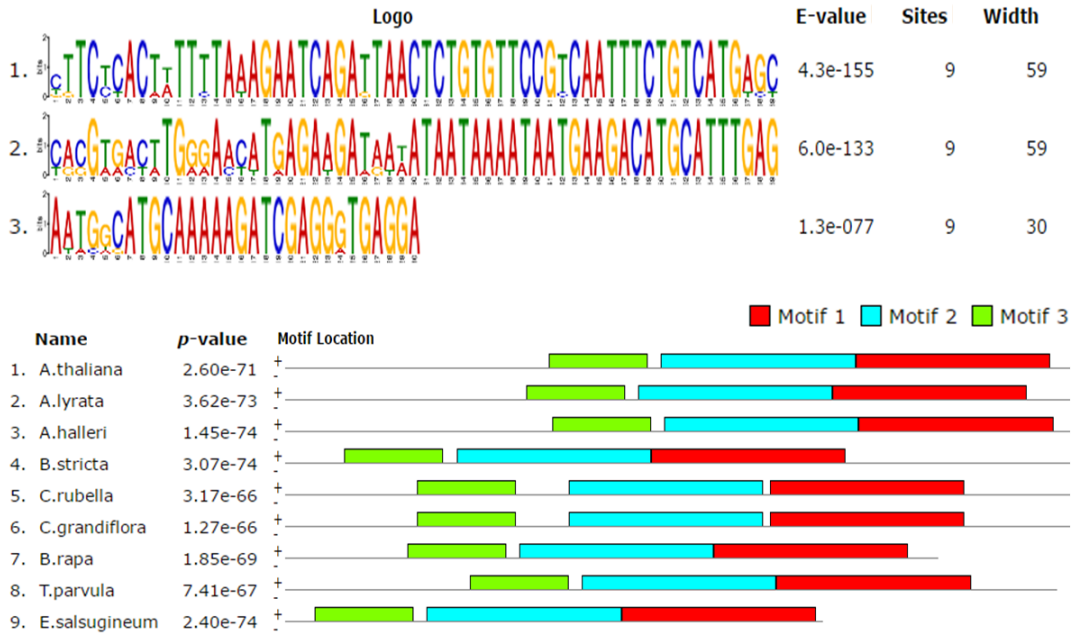


Figure 3.7 The core CNS2 of MS1 contains three consensus motifs generated by the online application of MEME 4.11.2.

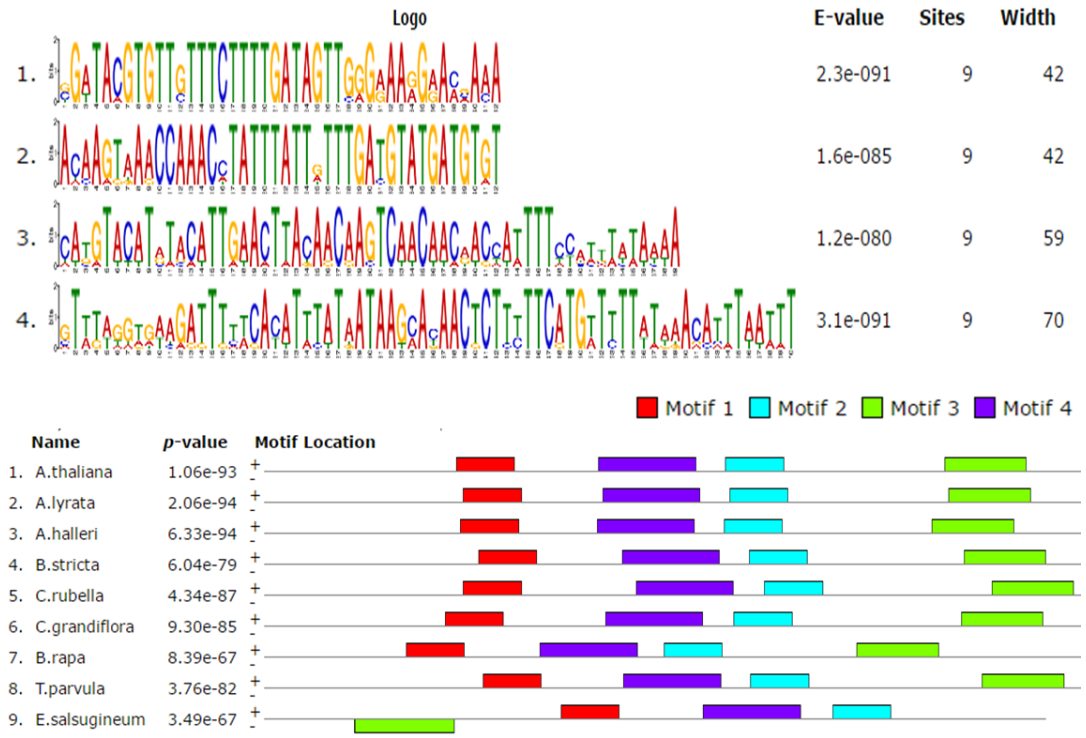


Figure 3.8 The distal CNS of MS1 contains four consensus motifs generated by the online application of MEME 4.11.2.

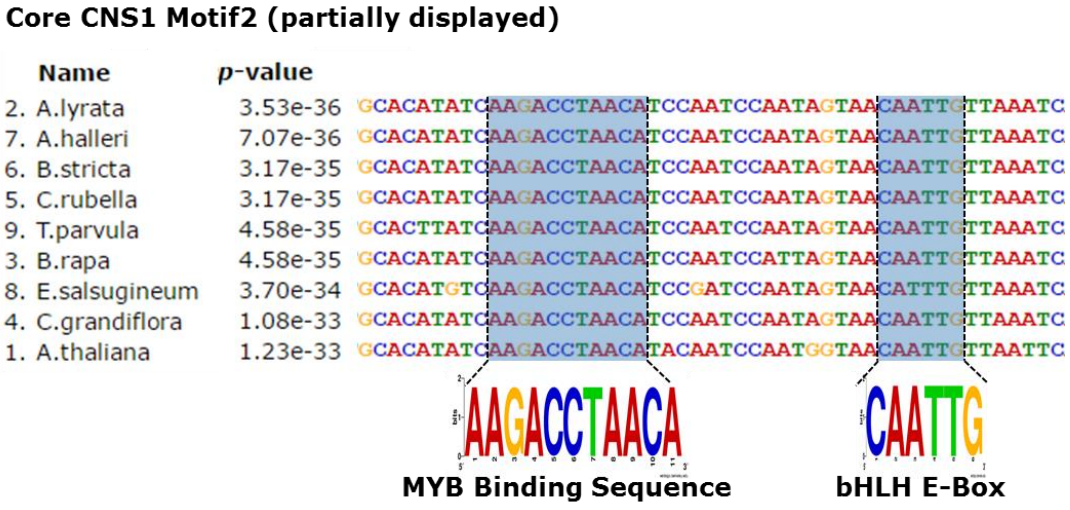


Figure 3.9 Motif 2 of MS1 core CNS1 contains a consensus MYB binding sequence and a canonical bHLH binding E-Box.

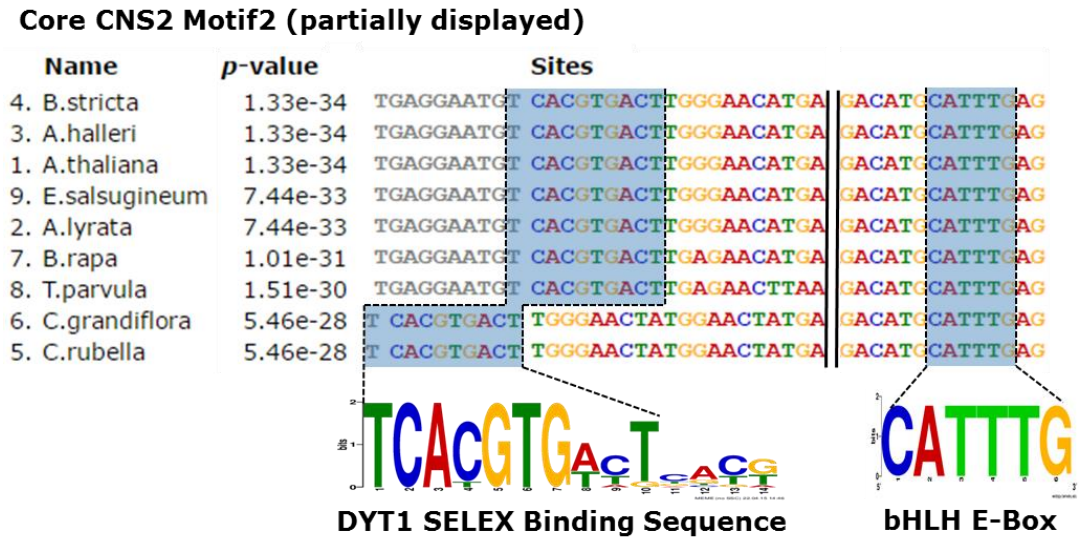


Figure 3.10 Motif 2 of MS1 core CNS2 contains a consensus DYT1 binding sequence and a canonical bHLH E-box.

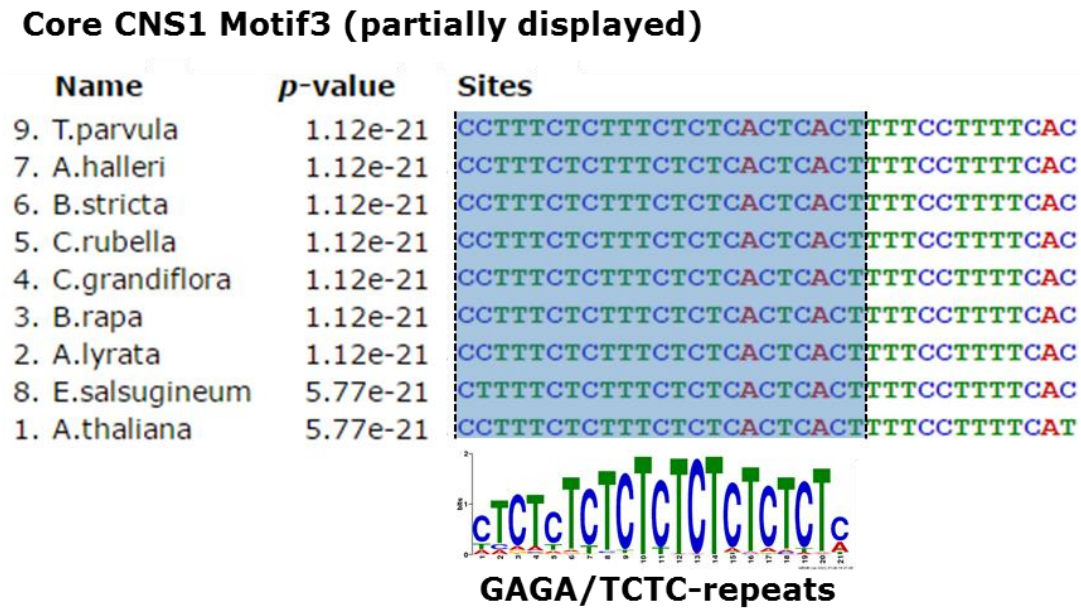


Figure 3.11 Motif 3 of MS1 core CNS1 contains a consensus sequence of GAGA-repeats.

3.2.2 Promoter dissection assay of the MS1 upstream sequence

To further characterise the function of the CNSs identified in the MS1 promoter, a molecular promoter dissection assay was carried out. Previous work has demonstrated 3.0 kb upstream of MS1 can fully rescue the mutant. Therefore, four truncated versions of this 3.0 kb DNA (termed as pro1, pro2, pro3 and pro4) based on the position of CNSs were made using GFP reporter binary vector and transformed into WT Arabidopsis (**Figure 3.12**). After the antibiotic screening of the transgenic seedlings and PCR verification of the T-DNA presence in the plant genome, four independent transformant lines of pro1-GFP and three independent transformant lines of pro2-GFP were obtained. Surprisingly, all of them failed to give any GFP signal in the anther (data not shown). Unfortunately, due to an overheating accident in the growth room, the pro3-GFP and pro4-GFP transformants failed to survive to the anthesis stage. Although the −2900 to −2066 bp region of MS1 upstream sequence mainly possesses gene AT5G22270 and a lncRNA AT5TU027100 (MS1lncRNA), it appears to be essential for the expression of MS1, which indicates additional TFBSs or other type of CREs exist in this region.

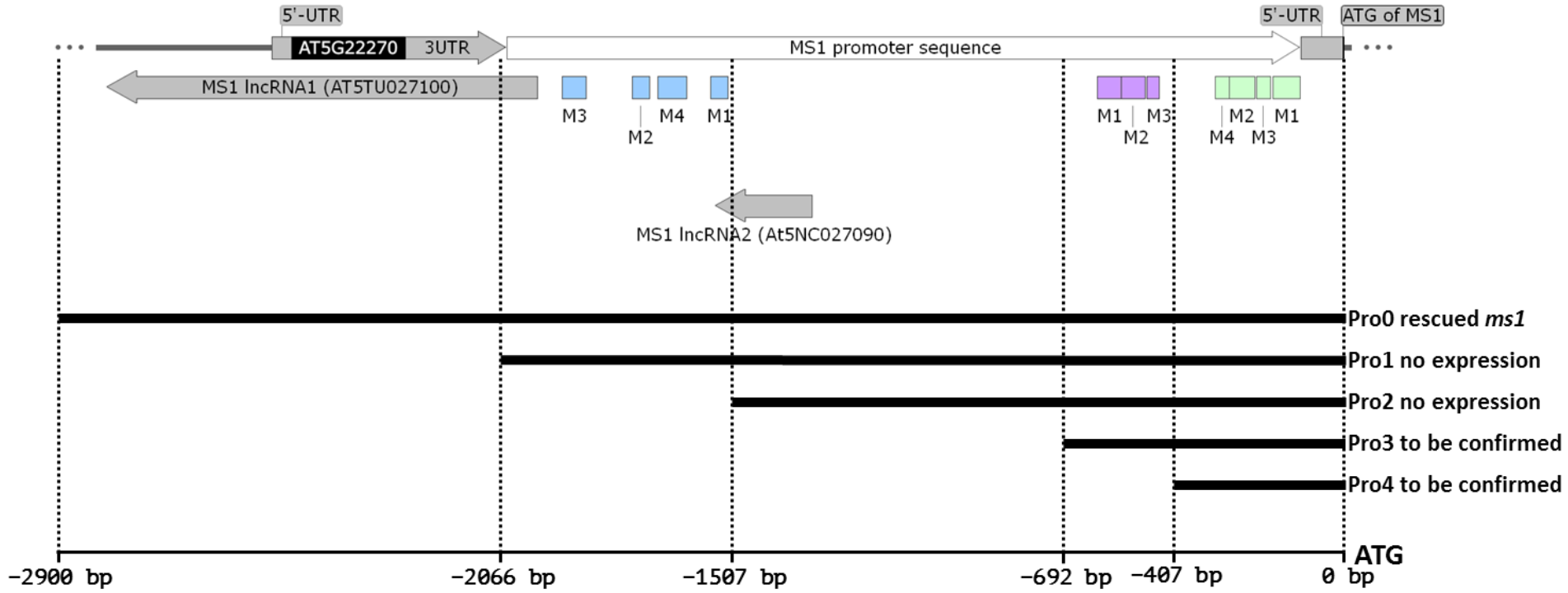


Figure 3.12 Schematic diagram showing truncated versions of the MS1 promoter in the promoter dissection assay.

Motifs of cCNS1, cCNS2 and dCNS are indicated with green, purple and blue boxes respectively with numbered labels. The pro0 (−2900 to −1 bp) presents the upstream sequence that Yang et al. (2007) used in the complementation assay which successfully rescued the *ms1* phenotype. The pro1 (−2066 to −1 bp) has the coding region of the adjacent gene, AT5G22270, truncated, which also abolished the MS1 lncRNA1 transcription. The distal CNS was truncated in pro2 (−1507 to −1 bp), which leads to an abolishment of the MS1 lncRNA2 as well; the intergenic DNA sequence between the distal CNS and the core CNS2 was truncated in pro3 (−692 to −1 bp) and pro4 (−407 to −1 bp) only kept the core CNS1.

3.2.3 Yeast one-hybrid assay of the binding of tapetum TFs to the *MS1* genomic locus

To confirm the interaction between the *in-silico* identified motifs and putative MS1 TF regulators, yeast one-hybrid assay was carried out. For yeast one-hybrid baits, the 2.5 kb upstream sequence of MS1 ATG site and 400 bp sequence downstream the ATG site was integrated into yeast strain Y187. In order to obtain valid yeast one-hybrid result, the ~3 kb DNA baits was constructed as seven ~450 bp promoter sectors, each of which has 60 bp overlapped with its neighbouring sectors on both 5' and 3' side in case important TFBSs were interrupted at the borders. For the putative MS1 TF regulators to screen against, the full length CDS of *AMS, DYT1, MS1, BES1* and *BZR1* were integrated into yeast strain AH109. The mating of yeast strain Y187 and AH109 was verified on a Petri dish plate of synthetic defined medium lacking tryptophan and leucine (SD-2) (**Figure 3.13 A**). However, no interaction occurred in yeast between the tested baits and preys as only the positive control colonies grew on the selection Petri dish plate of SD-3 (SD medium lacking tryptophan, leucine and histidine, **Figure 3.13 B**). Among the fragments of MS1 promoter used as DNA baits, the MS1 P4 possesses the BRRE site that the BES1 is binding to, which is verified by the ChIP-PCR (Ye et al., 2010), and the MS1 P5 possesses the E-box sites that DYT1 and bHLH010/089/091 binds to according to the *in vivo* dual-luciferase assay conducted by Cui et al. (2016).

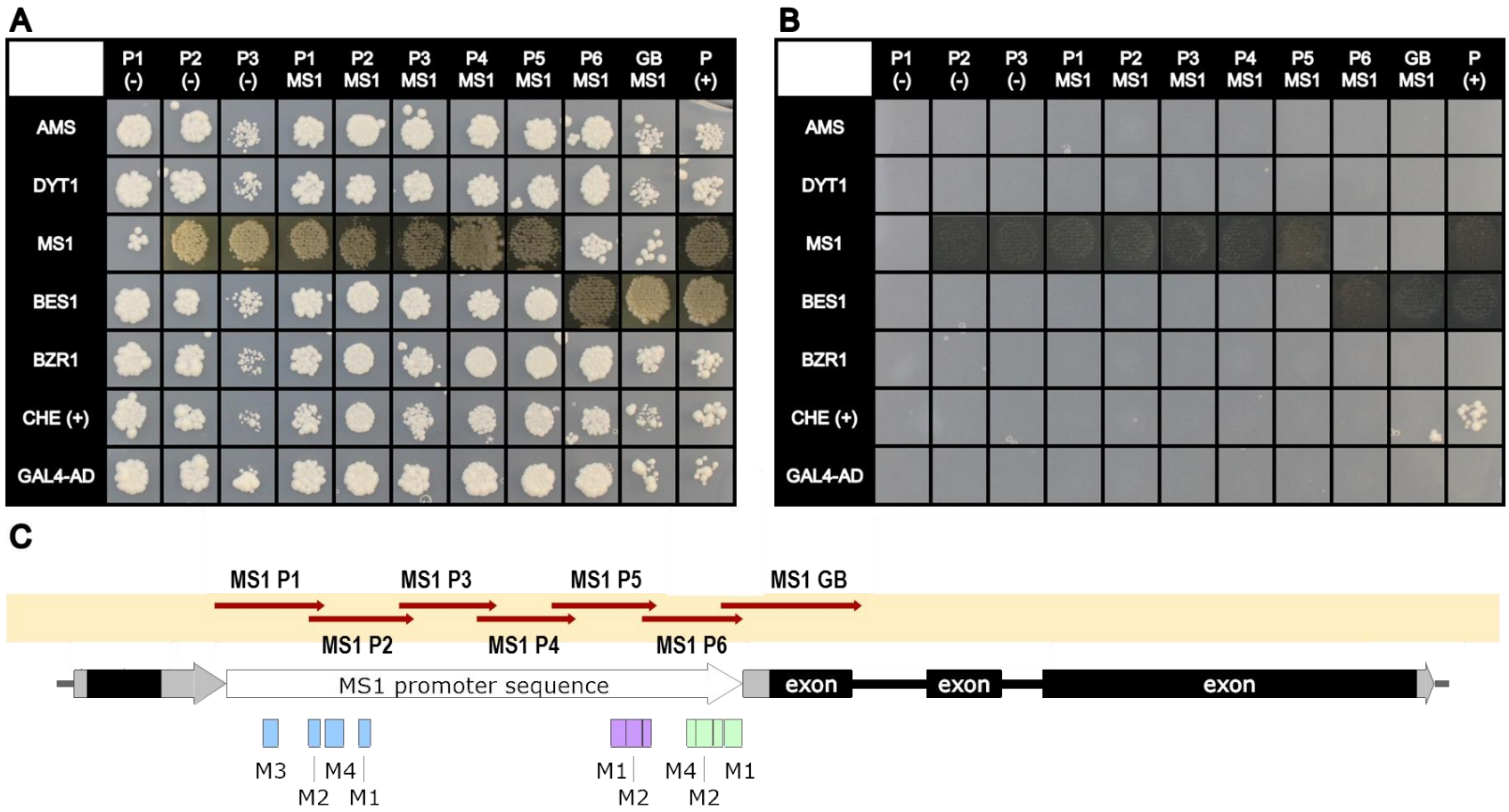


Figure 3.13 The yeast one hybrid indicated no interaction persists between selected tapetal TFs and the MS1 genomic sequence.

(A) The mating colonies of yeast strain Y187 possessing *MS1* genomic DNA baits and AH109 possessing the selected tapetum TF preys on SD-2 media. (B) The interaction assay of the mating colonies in (A) on SD-3, 5mM 3-AT media. (C) A schematic diagram showing partitioned regions of the *MS1* genomic sequence that used as bait DNA in the yeast-one-hybrid. Yeast colonies in rows from top to bottom contain AMS, DYT1, MS1, BES1, BZR1, CHE (+) and GAL4-AD (−) respectively; colonies in column from left to right contain random prom1 (−), random prom2 (−), random prom3 (−), MS1prom1, MS1prom2, MS1prom3, MS1prom4, MS1prom5, MS1prom6, MS1 gene body (GB, 5'-end 400bp), CCA1Prom (+) (217bp) respectively.

3.2.4 ChIP experiment confirmed MS1 is binding to its second exon

To verify if MS1 directly represses its own expression, a ChIP experiment was conducted on the genomic locus of *MS1* using an anti-GFP antibody and a *MS1-GFP* line inflorescence tissues as input material. As MS1 does not contain any known DNA binding protein domains, and the PHD domain is related to chromatin structure modification, the binding site of MS1 is not necessarily in the upstream region of the gene. Hence, not only the promoter region of *MS1* but also the gene body region was used for MS1 ChIP DNA enrichment detection. A few extra steps were included to optimise the ChIP output. The fixation condition is a determinant step for successful ChIP experiments. Three different fixation conditions, 1% (w/v) formaldehyde fixation in MC buffer for 15 minutes, 20 minutes and in deionized water for 20 minutes, were tested in three independent ChIP experiments. Additionally, as the MS1 protein abundance is very low, western blot was conducted as a quality control step after the chromatin isolation to make sure sufficient MS1-GFP protein was present in the immunoprecipitation reaction (**Figure 3.14**). Finally, both internal (*TA3* locus enrichment) and external (Ler tissue input material) negative controls, and percentage input analysis rather than fold enrichment analysis method were adopted to exclude diverse false positive mistakes common in ChIP experiments.

The comprehensive MS1-GFP ChIP detection at the *MS1* locus revealed quite few interesting points. First, the three batches of ChIP experiments under different fixation conditions all showed highest positive enrichments around the second exon region of *MS1* with a range of 5-14% input while the external negative control only obtained enrichment less than 1% input (**Figure 3.15**), indicating an *in vivo* binding of MS1 to the *MS1* gene body. Among the ChIP data obtained from *MS1-GFP* floral bud tissues,

the enrichment level of *MS1* upstream regions resembled that of the internal negative control *TA3*, representing a background ChIP level, which indicates there is no *in vivo* binding of MS1 to the *MS1* promoter. Secondly, the MC-buffer 10-minute batch showed the most significant enrichment at the junction of 3'-end of the first intron of *MS1* and the second exon, which coincided with the site of *ms1-1/ms1-6* SNP and *MS1 lncRNA4* (**Figure 3.15**). The flanking ChIP detected regions shared a decreased but significant enrichment. The other two batches showed similar enrichment pattern although the enrichment level was relatively low. The three ChIP enrichments covered a quite wide ~2 kb region of the *MS1* gene body, from which an approximating normal distribution trend of the enrichment was obtained (**Figure 3.15**. **red line**). Lastly, for every ChIP detected region, the enrichment level of MS1-GFP group was significantly higher than that of the L*er* group, which represents a systematic enrichment bias. This is probably because of nonspecific DNA affinity of the free GFP protein in the MS1-GFP immunoprecipitation system as abundant free GFPs were detected in western blot using *MS1-GFP* floral bud tissues (**Figure 3.14**).

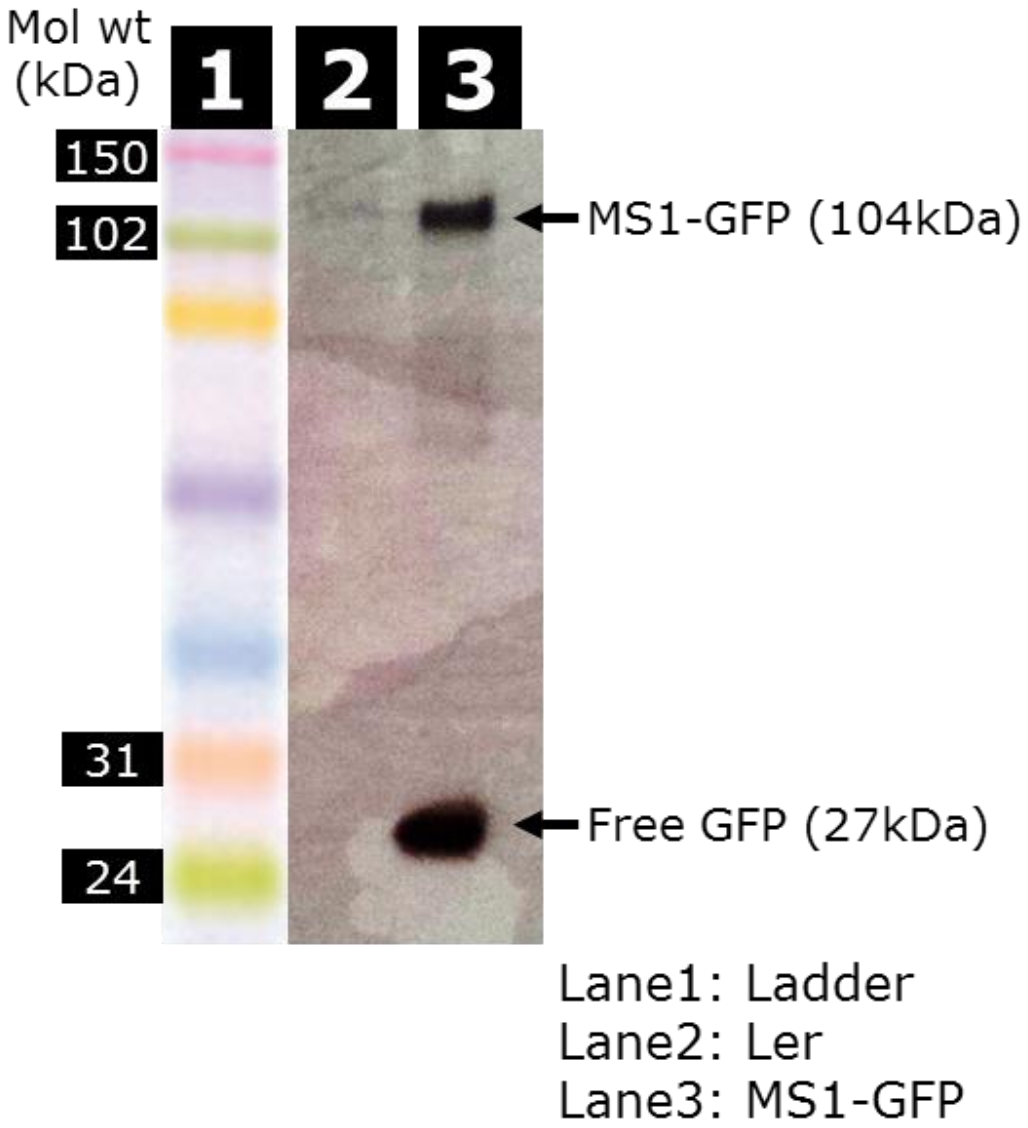


Figure 3.14 A western blot analysis of isolated chromatin for MS1-GFP ChIP.

The graph addresses the presence of MS1-GFP protein and abundant free GFP protein in the isolated chromatin extracted from MS1-GFP floral bud tissues. Lane1, ECL Rainbow™ Marker; Lane 2, the chromatin protein from the WT L*er*; Lane 3, the chromatin protein from the *MS1-GFP* line. The anti-GFP antibody and a second antibody conjugated with HRP luminophore was used to detect the GFP protein.

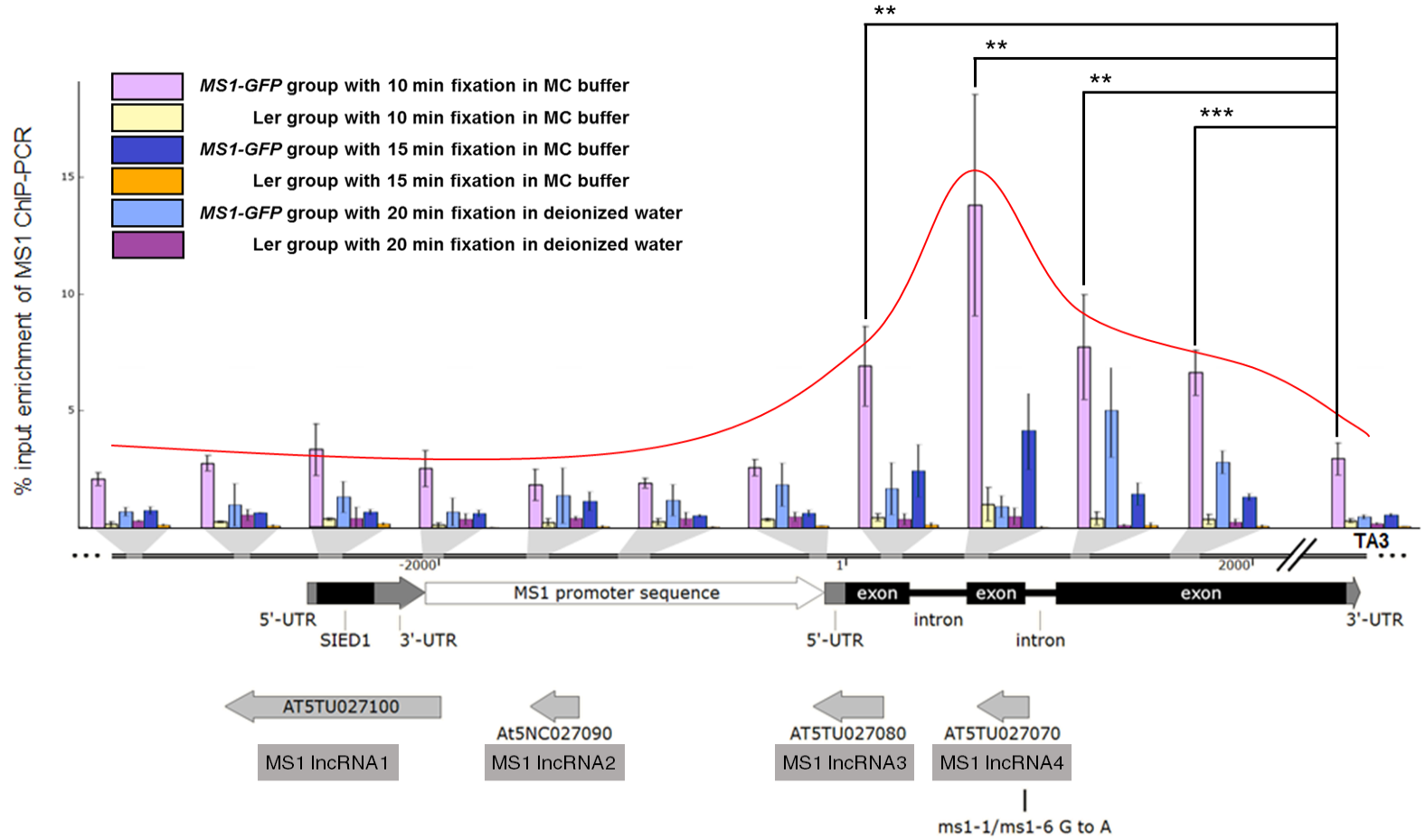


Figure 3.15 MS1 ChIP enrichment at the MS1 locus.

Three independent batches of the ChIP results are plotted in the bar chat diagram based on the genomic site of their PCR amplicon. The ChIP enrichments for each site are calculated as the percentage of input and compared with those of a negative control site, a retrotransposon gene TA3. Asterisks indicate a significant difference between the control of the tested site (paired Student's t-test; ** $p<0.01$, *** $p<0.001$, otherwise $p>0.05$). Error bar on each bar represents the 95% confidence interval of the enrichment. The red line shows a normal distribution trend of the enrichment for the region of *MS1* gene body.

3.2.5 Expression profiling of lncRNAs transcribed at the *MS1* locus

Transcription of four putative lncRNAs (*AT5TU027100*, *AT5NC027090*, *AT5TU027080* and *AT5TU027070*, termed as *MS1 lncRNA1, 2, 3, 4* respectively) have been identified at the MS1 locus by leaf transcriptome high-throughput sequencing profiling and *in-silico* prediction (**Figure 3.3** and **Figure 3.15**) (Wang et al., 2014; Jin et al., 2013), however whether they are transcribed in flowers is not verified yet. A semi-quantitative strand-specific RT-PCR, which is able to eliminate the interference of MS1 mRNA, was carried out using total RNA extracted from young floral buds (anther Stage 1-7 containing MS1 expressed anthers) and rosette leaves to evaluate the expression level of MS1 lncRNAs. For each sample type, three biological replicates were conducted. It turned out that the MS1 lncRNAs were expressed, and the level of expression differed between leaf and flower tissue. The expression of *MS1 lncRNA1* was barely detected in rosette leaves. The expression levels of *MS1 lncRNA2* and *MS1 lncRNA3* were relatively low both in rosette leaves and floral buds, whilst the *MS1 lncRNA4* showed high expression in both tissues (**Figure 3.16 A**). A semi-quantitation of the expression levels of MS1 lncRNAs was conducted based on the signal density of DNA bands on the gel image, and normalised to the Arabidopsis reference gene *YLS8*. Statistically significant differences were obtained for the expression level of MS1 lncRNA1 ($p < 0.05$) and MS1 lncRNA2 ($p < 0.0001$) between the rosette leaf and floral bud group (**Figure 3.16 B**).

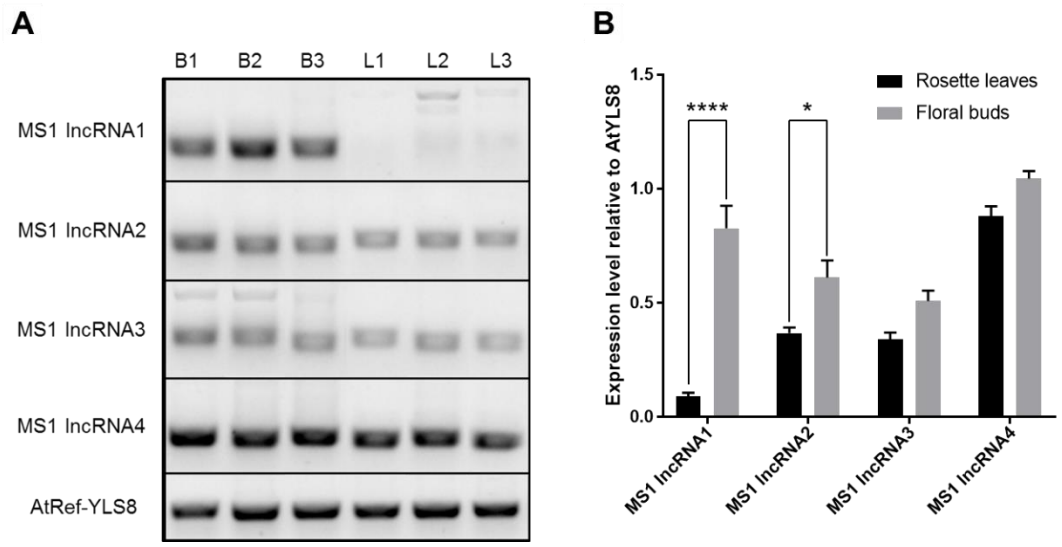


Figure 3.16 Semi-quantitative RT-PCR of four MS1 lncRNAs in flower buds and leaf tissues.

(A) Agarose gel electrophoresis of strand-specific RT-PCR products of four MS1 lncRNAs and reference gene AtYLS8. Lane B1-3 and L1-3 were loaded respectively with the PCR products of the floral buds and rosette leaves from three individual plants. (B) Semi-quantitation of the expression level of MS1 lncRNAs. All expression levels are normalised to that of the reference gene *YLS8*, which is converted to 1. Error bars represent the standard error of the calculated mean based on three biological replicates. The statistical significance is a result of t-test between rosette leaves and floral buds, and the p-value is indicated by * when $p<0.05$ and by **** when $p<0.0001$.

3.2.6 A novel tapetum secretory structure was absent in the *lrb123* mutant

In order to identify if LRB proteins are involved in the removal of MS1 protein, a series of phenotypic studies of the homozygous triple mutant *lrb123* line (a kind donation of Peter H. Quail lab) were carried out. This line was generated by rounds of crossings of three single T-DNA insertion mutants of *lrb1-1* (Salk_145146), *lrb2-1* (Salk_001013) and *lrb3-1* (Salk_082868) (Ni et al., 2014). After PCR-genotyping procedure confirming the T-DNA presence in the *lrb123* genome and RT-PCR expression assay verifying the knockout of LRBs, the homozygous triple mutants were sowed to grow in two conditions, the 16-hour light (long day) and 24-hour light.

In the long day condition, the *lrb123* plants showed a normal vegetative and reproductive growth, whereas in 24-hours light condition, the *lrb123* plants showed light stressed and stunted rosettes during two-three weeks after germination and bolted significantly later compared to the Col WT Arabidopsis (data not shown). This phenotype is consistent with the role of LRBs in quenching the active photoreceptor and photo-effector to prevent the photo-signalling pathways from overactivation under prolonged light conditions. On the other hand, when the *lrb123* plants entered reproductive growth, the light-stressed phenotype disappeared and the reproductive fertility displayed no observable change compared to that of Col WT Arabidopsis.

Given that plants produce high levels of pollen grains in each flower, being fully fertile does not necessarily mean that anther or pollen are normally developed. Alexander staining (Peterson et al., 2010) and Ethidium bromide (EB) and Acridine orange (AO) double staining (Fan et al., 2014) were carried out to detect pollen viability and anther lignin deposition respectively between the *lrb123* and the WT, however, no differences were obtained in these staining assays (data not shown). A precise phenotypic observation of the anther locule, tapetum, and pollen surface using SEM was therefore conducted. At the micrometre-scale, pollen grains of both the *lrb123* mutant and the Col WT plants exhibited the same regular patterned outer pollen wall, showing no alteration of the exine deposit on the pollen surface (**Figure 3.17 A** and **B**). However, a close observation of the tapetum revealed a remarkable difference between the mutant and WT at the inner tangential membrane of tapetal cells, which faces the anther locule (**Figure 3.17 C** to **F**). A large number of irregular globule bodies densely spread on the surface of inner tangential membrane of tapetum cells (**Figure 3.17 C** and **E**), while that of *lrb123* mutant was smooth and clear (**Figure 3.17 D** and **F**), indicating a dysfunction in *lrb123* tapetum development.

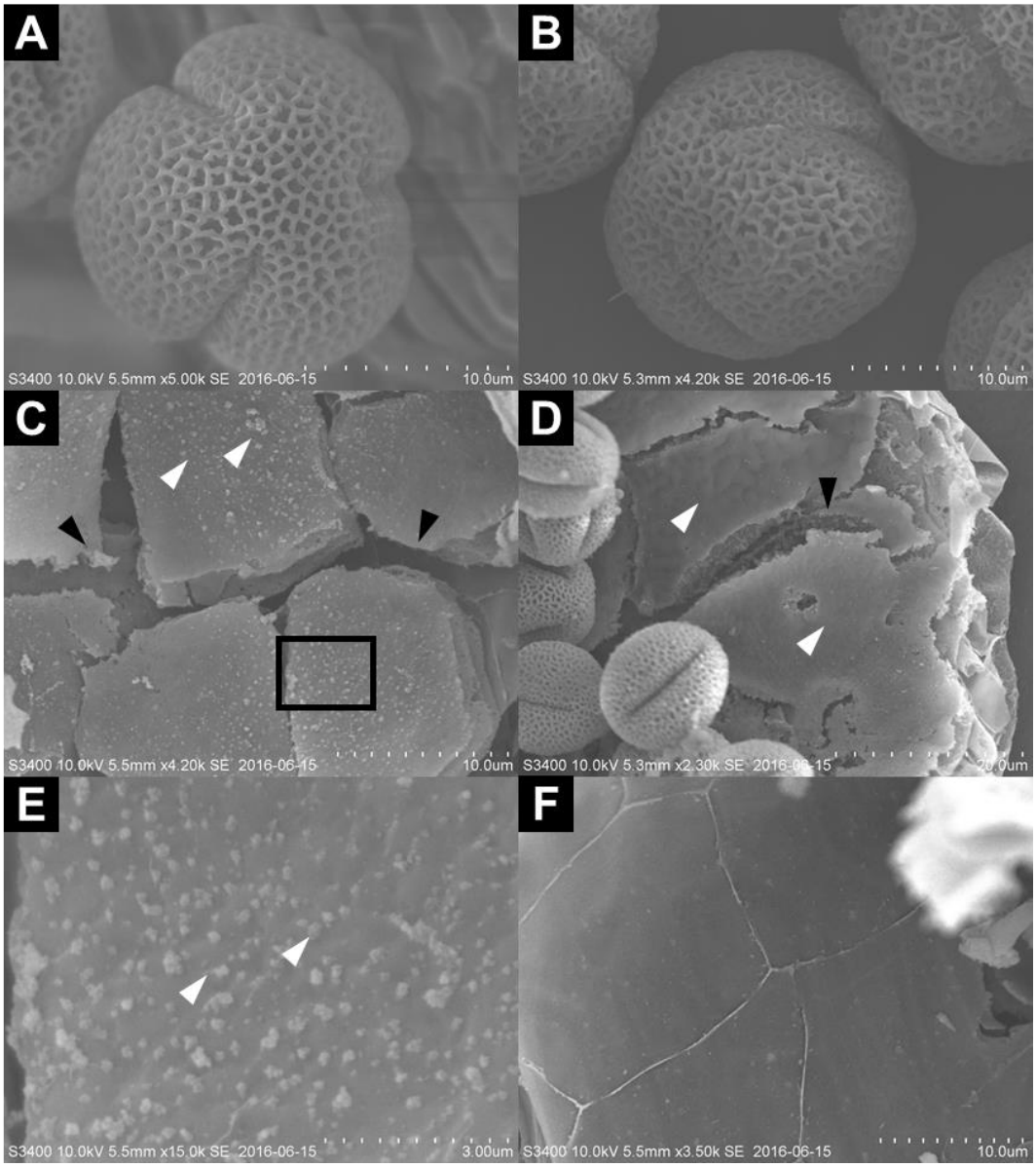


Figure 3.17 SEM images showing the pollen wall and tapetum cells of the *lrb123* mutant and the WT.

(A) and (B) The pollen exine ornament pattern showed no difference between the WT (A) and the *lrb123* mutant (B). (C-F) The inner tangential membrane of tapetal cells of the WT (C and E) and the *lrb123* mutant (D and F). (C) A type of small irregular globule bodies spread on the surface of WT tapetal cell membrane, indicated by white triangles. (D) The *lrb123* tapetal cell membrane are smooth and clear, indicated by white triangles. (E) A zoom in of the black frame area of (C). (F) The smooth inner tangential surface of the *lrb123* tapetum which is at an earlier stage with radical tapetal cell walls intact connecting the cells to form a tapetum lamella. The black triangles in (C) and (D) indicate disconnected tapetum cell margin as the tapetal cell wall was degraded at this stage.

3.3 DISCUSSION

For successful reproduction process in flowering plant, a tightly regulated development of anther is important. For instance, a correct timing of the mature pollen grain releasing is crucial for the accomplishment of cross-pollination and self-pollination (Krizek and Fletcher, 2005). An early or delayed anther dehiscence causes severe fertility reduction or even complete sterility (Zhu et al, 2004).

As in the case of MS1, the correct timing and transient duration of the MS1 expression is crucial for its function in the regulation of late tapetum phase transition and late tapetum development. Therefore, the genetic control of MS1 is key to uncover the regulation network for late tapetum development. In this chapter, three levels of the regulation of MS1 expression were studied and each level was discussed as follows.

3.3.1 Activation of MS1 expression needs a combination of CREs

The −2.9 kb upstream sequence of MS1, which possesses an adjacent gene *SIED1*, the three phylogenetic footprinting identified CNSs and two lncRNAs (**Figure 3.3** and **Figure 3.12**), is sufficient for the tapetum specific activation of MS1 expression at tetrads release stage. Promoter dissection assay uncovered that the transcriptional expression of MS1 cannot be activated without the −2900 to −2066 bp region of MS1 upstream sequence, which overlap with the gene body of *SIED1* and most part of *MS1 lncRNA1* (**Figure 3.12**). The *SIED1* is highly expressed in leaves, cotyledons and seeds (Peng et al., 2014), whilst the floral bud high expressed *MS1 lncRNA1* has an extremely low expression in rosette leaves, revealed by the strand-specific RT-PCR

(**Figure 3.16 B**). Although the *MS1 lncRNA1* is an anti-sense lncRNA of *SIED1*, their expression seems mutually exclusive in leaf and flower tissues, which indicates a possible transcriptional regulatory interaction between them. Interestingly, in a laser capture microdissection two-colour-array data of tapetum (unpublished data, see **Section 4.1.4**), *SIED1* showed tapetum specific expression at the tetrad stage, which is very similar with that of *MS1*. Hence, it is highly possible that an unknown robust tapetum specific CRE or CREs exist at the immediate upstream region of *SIED1* (−2.9 to −2.6 kb) driving the tapetum expression of both *MS1* and *SIED1* at tetrads stage, meanwhile, the flower highly-expressed *MS1 lncRNA1* represses the *SIED1* expression in other parts of the flower. The evolutionarily conserved MS1 distal CNS (−1950 to −1507 bp) approximates the transcription start site (TSS) of *MS1 lncRNA1*, and possibly promotes the characterised differentiated expression of *MS1 lncRNA1*.

The upstream region (−1507 to −692 bp) between MS1 distal CNS and core CNS2 appears to have a higher degeneracy among the Brassicaceae species used for the phylogenetic footprinting, which indicates it may play a less significant role in the regulation of MS1 expression. However, at least three BRRE boxes were found irregularly interspersed within this region of MS1 and each of its Brassicaceae orthologs, some of which even possess five ones. These BRRE boxes collectively provide a fine DNA-context for the binding of BR signalling effectors, such as BES1. Ye et al. (2010) reported *BES1* is essential for the expression of MS1 and identified direct binding of BES1 to the −1121 to −952 bp upstream region of MS1, which, however, does not contain any BRRE boxes. This may explain why in their ChIP data, this tested MS1 upstream region only got a moderate ChIP enrichment. Additionally, the pervasive BRRE boxes within a relatively short region of MS1 upstream sequence highlighted the importance of BR signalling in the activation of MS1. On the other

hand, the fact that these BRRE boxes mostly locate out of identified CNSs and have diverse positions relative to the TSS of MS1 or MS1 orthologs suggests that the successful regulation of MS1 by BES1 is not sensitive to the distance between the BES1 binding sites and the TSS.

The core CNS1 (− 606 to −454 bp) and CNS2 (−315 to 103 bp) region of MS1 upstream sequence together possesses seven DNA consensus motifs with unusually length ranging from 30 to 70 bp. Unlike TFBSs of vertebrates that are frequently more than 100 bp in length, plant TFBSs are infrequently over 30 bp, which make it less likely that the motifs within each MS1 CNS are individual plant TFBSs (Davuluri et al., 2003; Higo, 1998; Haberer et al., 2006). However, the high conservation of the DNA sequence suggests they may perform important CRE functions that are under selective constraints on their evolution. Although little is known about the exact function of these long motifs, a few relevant TFBSs was found within the motif, which highlights their CRE role in the regulation of MS1 expression.

In the core CNS1 motif 2, A closed linked pair of MYB binding sequence and bHLH E-box were identified, which might present a binding site for the AMS-MYB80 complex (Xiong et al., 2016). *AMS* and *MYB80* both are upstream in MS1 pathway and was speculated to regulate the MS1 expression. More recently, Xiong et al. (2016) demonstrated that MYB80 complexed with AMS and directly targeted to the promoter of *CYP703A2*, a MS1 co-expressed gene, and activated its expression *in vivo*. It is possible that MYB80 and AMS regulate the expression of MS1 in an analogous way.

In the core CNS2 motif 2, a highly conserved DYT1 binding sequence and bHLH E-box were found with an interval of 38bp between them. Cui et al. (2016) demonstrated that the DYT1-bHLH010/089/091 complex binds to the MS1 upstream

sequence and activated its expression. However, in their study, the exact binding site of the DYT1-bHLH010/089/091 complex in the MS1 upstream sequence was not revealed. Additionally, *TDF1*/*MYB35*, another direct target of DYT1-bHLH010/089/091 complex (Cui et al., 2016), possesses two upstream DYT1 binding E-box variants which are ~1kb apart from its TSS. The DNA context of the DYT1 binding sites of *TDF1*/*MYB35* and their distance to TSS are distinct from those of MS1, which might account for the temporal expression difference between *TDF1*/*MYB35* and *MS1*.

Two GA/TC-rich element variants, which are TFBS specific to BPC proteins, were identified within the core CNS2 motif 3 and the first intron of *MS1* respectively. The BPC proteins are also called GAGA BINDING PROTEINS owing to its specifically affinity to GA/TC-rich elements (Monfared et al., 2011). Early reports suggest this protein family can display both activator/anti-repressor and repressor activity (Berger and Dubreucq, 2012). A more recent report revealed that a member of the BPC family, BPC6, recruits POLYCOMB-REPRESSIVE COMPLEX1 (PRC1) component to GAGA DNA motif and plays a role in the concerted repression of homeotic genes (Hecker et al., 2015). The PRCs usually associate with inactive chromatin and maintain gene silencing through histone modifications, for instance, the successive recruitment of PRC1 and PRC2 to the chromatin can trimethylates lysine 27 on histone 3, generating H3K27me3, a repressive mark of gene transcription. Interestingly, the locus of *MS1* is highly modified by the H3K27me3 both in leaf and flower tissues (Zhang et al., 2007a). The recruitment of BPCs to the GAGA DNA repeats at the *MS1* locus may help maintaining the *MS1* silenced state in other plant tissues beyond the tapetum.

The identified conserved TFBSs underpin the possibility that MS1 is simultaneously directly-regulated by a series TFs or TF complexes, including BES1, AMS-MYB80 and DYT1-bHLH010/089/091. However, BES1, AMS and DYT1 cannot bind to the MS1 upstream sequence in yeast cells, and each of the TFs is not sufficient to promote the transcription of *MS1*. This suggest that the TF regulators of *MS1* might act in a cooperative manner (Spitz and Furlong, 2012), in which the protein–protein interactions between TFs and the sequential recruitments of TFs to their cognate TFBSs are integrated to produces a sharp, switch-like activation of *MS1*. In this model, the earlier TFs just progressively ready *MS1* for transcriptional expression, such as BES1, DYT1, bHLH010/089/091 and AMS, while a final deterministic factor, which most possibly is MYB80, commence the transcription of *MS1*.

3.3.2 MS1 auto-repress via binding to the gene of body of itself

MS1 is auto-repressed as the *ms1* transcripts showed upregulated and prolonged expression in the homozygous *ms1* mutant compared to the MS1 expression level in the WT. Whether this auto-repression is via direct binding of MS1 to its genomic locus is not known. Three independent replicates of ChIP experiment verified that MS1 is recruited to the gene body region of itself. The ChIP enrichment data covered a ~2 kb region, forming a normal distribution trend with a peak enrichment value (14% input) at the junction of the first intron and the second exon (**Figure 3.15**). This well explained the fact that the transcriptional GUS reporter of MS1 has a prolonged expression compared to that of the translational GFP reporter, as the former lacked the MS1 gene body sequence to recruit MS1 protein for the auto-repression.

A functional MS1 protein is the prerequisite of the MS1 auto-repression. However, among the six loss-of-function mutants of *MS1* (*ms1-1*, *2*, *3*, *4*, *5* and *6*), the upregulation of the mutant transcript varied (**Figure 3.2 B**), which indicate there might be a second factor repressing MS1 expression. The *ms1-1*, *ms1-4* and *ms1-6* mutants all have the same premature ORF (**Figure 3.2 D**), but showed differentiated upregulation of the *ms1* transcript (**Figure 3.2 B**), which implied that the partially functioning MS1 protein in the *ms1* mutants, if present, does not take part in the repression of MS1. The overlapping of the peak MS1 ChIP enrichment, the MS1 lncRNA4 and the *ms1-1/ms1-6* SNP might be not a coincidence (**Figure 3.15**). LncRNA may execute its *in cis* transcription regulation function via participating in modifying chromatin marks (Liu et al., 2015a). Given that the MS1 lncRNA4 is stably expressed in leaf and flower buds, the lncRNA4 might play a role in maintaining the repressive histone mark, such as the H3K27me3, and the *ms1-1/ms1-6* SNP not only abolishes normal splicing of the second intron of MS1, but also possibly affects the normal function of MS1 lncRNA4.

Considering that MS1 contains a PHD finger domain, which has been reported as a histone modification reader (Mouriz et al., 2015), and the unusually wide region of MS1 ChIP enrichment, it is highly possible that MS1 binds to its gene body via recognising biologically relevant histone modifications rather than any specific TFBSs or CREs. Interestingly, the MS1 ChIP enriched region coincided with a high level of Histone 3 lysine 27 trimethylation (H3K27me3) modification that were both detected in young Arabidopsis seedling leaves and flower. How the histone 3 tail in this region in the anther is modified remains unknown. However, as MS1 possesses a PHD finger, which have been reported as a histone modification reader, the

correlation between the MS1 binding and the histone modification might be of a biological significance.

3.3.3 LRBs might mediated the degradation of MS1

Given that an excess of MS1 protein is harmful to the tapetum development and the MS1 interacting protein LRB1, a E3 ubiquitin ligase, was speculated to mediate the MS1 protein degradation (Ni et al., 2014; Yu, 2015). A defect in tapetum development was expected in the triple *lrb123* mutant. Interestingly, a precise SEM phenotypic observation identified a novel secretory structure of the tapetum was absent in the *lrb123* mutant.

This novel structure assembled rice tapetum orbicules (or Ubisch bodies), which are composed of sporopollenin similar to the pollen exine and play a part in the exine growth (El-Ghazaly, 1999), though, the tapetum specific secretory orbicules have not been reported in Arabidopsis to date. As producing orbicules is a common feature of most angiosperms that has a secretory type of tapetum (Pacini and Franchi, 1993), this novel structure could be the first report of Arabidopsis tapetum orbicules. Nevertheless, it cannot be excluded for the possibility that this structure represents a nonspecific secretory vesicle which transport lipid components of pollen coat to the anther locule (Hsieh and Huang, 2004). Further studies are needed to reveal the nature of this novel structure.

Although the pollen viability is not affected, the fact that *lrb123* lacks this tapetum secretory structure implies the absence of LRBs leads to a dysfunction in tapetum

development, whether this is caused by the delayed removal of MS1 protein still needs more detailed verification.

In the photomorphogenesis progress, LRBs plays a role in removal of activated photoreceptor and photo-effector to maintain the sensitivity of photo-signalling pathways and prevent it from overactivation (Fraser et al., 2016). One possibility is that the tapetum developmental pathway adopts the same mechanism. In the tetrad release stage tapetum, an initial robust burst of MS1 activity is presumably necessary for the induction of the tapetum developmental switch. However, unfettered and prolonged MS1 activity may be harmful to the tapetum development and result in a defect that *lrb123* showed.

3.3.4 Summary and future work

MS1 is under strict regulation, which is crucial for successful late tapetum phase transition and late tapetum development. From previous studies, three levels of regulation of MS1 were inferred and closely inspected in the studies described in this Chapter. First, a simultaneous regulation model of the activation of MS1 by a few tapetum key TFs was proposed and a series of promising TFBSs were identified at the upstream sequence of MS1. Second, MS1 was confirmed to undergo a direct auto-repression, in which, histone modifications and an anti-sense lncRNA at the MS1 genomic locus might be involved. Third, a knock-out line of a MS1-interacting E3 ubiquitin ligase, LRB1 and its two paralogs, LRB2 and 3, exhibited subtle but significant phenotype in tapetum development, which underpins the speculation that a proteolysis process of MS1 protein exists to ensures the normal development of tapetum. The three levels of regulation of MS1 are summarised in **Figure 3.18**.

For the activation of MS1, a series of promising TFBSs within the upstream sequence of MS1 were identified in this work. The yeast one-hybrid assay failed to identify any *in vitro* TF-TFBS interactions; verifying the *in vivo* cognate TF-TFBS interactions by the luciferase transactivation assay in tobacco leaf would be a quick and straightforward way to evaluate the effect of each TF-TFBS interactions on the MS1 activation. On the other hand, the direct autorepression of MS1 implied a repressor role of MS1, and there are a large number of tapetum specific genes upregulated in the *ms1* mutant (see **Section 4.2.4**), ChIP verifying of the presence of MS1 protein at these genomic loci would reveal if this repressor role of MS1 is universally applied. Finally, biochemical evidence of the LRB mediated ubiquitination of MS1 protein would be the essential step to demonstrate the post-translational proteolysis regulation of MS1, and identifying the missing components of pollen wall of *lrb123* would reveal the nature of the identified novel Arabidopsis tapetum secretory structure.

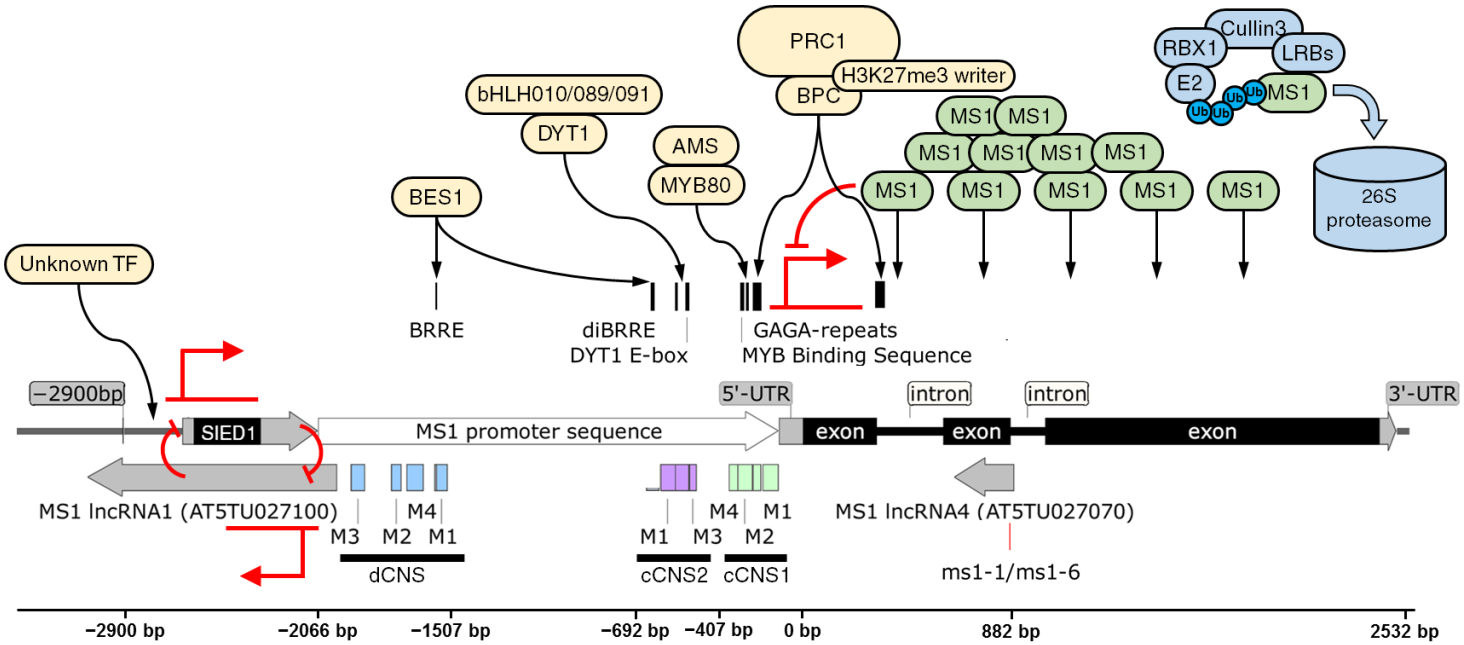


Figure 3.18 Proposed model of the three levels of regulation of MS1.

Level one, a simultaneous regulation of the activation of MS1, in which the factors involved are labelled in pale yellow ovals. The expression of *SIED1* and *MS1 lncRNA1* is mutually exclusive. Unknown TF that binds to the TFBS located in the −2900 to −2066 bp region of MS1 is essential for the tapetum specific expression of *SIED1* and *MS1* at tetrad stage. The TF of BES1, and TF complexes of DYT1-bHLH010/089/091, AMS-MYB80 and BPC-PRC1-H3K27me3 writer bind to the BRRE box, DYT1-E box, MYB binding sequence and GAGA-repeats to activate the expression of MS1 in a cooperative manner.

Level two, MS1 is directly auto-repressed, in which the factors involved are labelled in green ovals. MS1 proteins are recruited to the ~2 kb region of MS1 gene body, possibly via the interaction between the PHD finger of MS1 and the histone modifications at genomic locus MS1. The *MS1 lncRNA4* and *ms1-1/ms1-6* SNP coincide with the most enriched MS1 ChIP region, possibly indicates their role in the auto-repression.

Level three, MS1 undergoes Ubiquitin mediated proteolysis process via protein-protein interaction with LRBs, in which the factors involved in are labelled in blue ovals except MS1. The MS1 protein is ubiquitinated by a CRL3 (CUL3–RBX1–LRB) E3 Ubiquitin Ligase Complex, and then targeted for a concomitant degradation by the 26S proteasome pathway.

CHAPTER 4 PHYLOGENETIC ANALYSIS AND A PRELIMINARY FUNCTIONAL STUDY OF MS1

4.1 INTRODUCTION

MS1 directs the late tapetum transition in anther development (reviewed in **Section 1.3.1**) and therefore ensures the proper functioning of the tapetum in pollen wall and coat formation. Despite its importance, after nearly twenty years of study since the first identification of MS1 in Arabidopsis (Thorlby et al., 1997), the exact molecular function of MS1 remains uncertain, which is partly due to a lack of annotated protein domains on the relatively long protein sequence besides the PHD domain. Additionally, the function of the PHD domain in transcription regulation in plants has been unclear until recently (Mouriz et al., 2015). With more plant PHD-containing proteins characterised and more plant genome and transcriptomic data becoming available, a comprehensive meta-analysis was conducted using phylogenetic analysis, molecular modelling, and transcriptome clustering analysis, to resolve the molecular function of MS1. By using advanced super-resolution microscopy imaging techniques, a better resolution of the MS1 sub-organelle localisation was obtained, which has further helped towards identifying the functional mechanism of MS1.

4.1.1 PHD domain functions as histone readers

Chromatin remodelling plays a central role in the regulation of gene expression in eukaryotic organisms, and particularly in plants (Jarillo et al., 2009; He et al., 2011).

The concentration of different posttranslational modifications at the amino-terminal tails of histones significantly impacts upon the regulation of transcriptional programmes, which represents the main class of chromatin-mediated regulation. For instance, the modification by acetylation or methylation of specific residues in the tail of histone 3 influences many plant developmental transitions such as the initiation of flowering and seed germination (He, 2012; Jarillo and Piñeiro, 2011). These histone posttranslational modifications (PTMs) that decorate nucleosomes regulate the transcription of underlying genes in two ways. One is in reorganising the chromatin structure to change its accessibility to the transcription machinery. The other is via recruiting a variety of protein complexes that establish or maintain the gene activation or repression (Suganuma and Workman, 2011). In the latter process, specific recognition of histone marks by functional domains, so-called 'histone readers', are a prerequisite for the recruitment of chromatin remodelling complexes and transcriptional machinery, which together with the readers transform the histone PTMs to regulate gene transcription (Mouriz et al., 2015).

In recent years, the histone reader modules that can specifically recognise the covalent histone PTM have been identified (Yun et al., 2011). Among them, the Plant Homeo Domain (PHD) represents a substantial and versatile family of histone readers. The PHD domain is a Cys4-His-Cys3 double loop Zn-finger motif, which was firstly found in *Arabidopsis thaliana* HAT3.1 (Sanchez and Zhou, 2011). Since then, a variety of proteins possessing PHD domain have been found in other eukaryotic organisms (Li and Li, 2012). In the studies of human BROMODOMAIN AND PHD FINGER TRANSCRIPTION FACTOR (BPTF) and INHIBITOR OF GROWTH2 (ING2) proteins, the PHD motif was reported to function via specifically binding trimethylated Lysine 4 in histone H3 (H3K4me3), a histone mark of active gene

expression. The binding of these human PHD proteins to nucleosomes can result in either transcriptional activation, or repression depending on the factors subsequently recruited, which implies that the histone reader need additional factors, such as TFs, histone erasers and writers, to translate the information of histone PTMs into active regulators (Becker, 2006; Mellor, 2006; Wysocka et al., 2006; Shi et al., 2006; Peña et al., 2006; Li et al., 2006a). Later investigation have found a number of derived PHD domains that binds H3K4me2/3, H3K9me3, H3K36me3, acetylation of H3K14 or even unmodified H3K4 (Musselman and Kutateladze, 2011; Li and Li, 2012).

Large PHD-containing proteins often bear additional domains, which can also 'read', 'erase' or 'write' histone PTM marks. For instance, a rice Chromodomain-Helicase-DNA (CHD)-binding protein, CHR729, contains a single PHD domain and two chromodomains, which is required for many aspects of plant development. The CHR729 can interact with H3K4me2 and H3K27me3, respectively, through the chromodomains and the PHD domain of the protein. These interactions are required to maintain H3K4 and H3K27 methylation over repressed or tissue-specific genes, therefore directing the gene transcription program of plant development (Hu et al., 2012). DNA-binding domains, which can specifically recognise CREs, can sometimes be seen in PHD-domain containing proteins. For instance, the rice flowering activator protein VIN3-like2 (OsVIL2) contains a putative B3 DNA-binding domain besides the PHD domain. OsVIL2 promotes rice flowering via binding histone H3 through its PHD domain and directing the enrichment of H3K27me3 at the floral repressor loci, which requires the presence of the B3 DNA-binding domain (Yang et al., 2013). Furthermore, PHD-containing proteins have been seen to possess PPI domains, which link them to chromatin remodelling protein complexes. The Alfin1-like (AL) protein contains an H3K4me3 binding PHD

domain at the C-terminus and a family specific PHD-associated AL (PAL) domain at the N-terminus (Song et al., 2013). Functional studies have uncovered that AL6 and AL7 can interact with the Polycomb Repressive Complex 1 (PRC1) via their PAL domain, and repress seed developmental genes by switching the active histone mark of H3K4me3 to the inactive H3K27me3 during seed germination and early seedling growth (Wei et al., 2015; Molitor et al., 2014).

These enhanced features of PHD containing proteins strengthen the capacity of PHD domains to exert remarkably complex regulatory functions and to mediate the cross talk between the local histone PTM combinations and the gene transcription status.

MS1 represents a unique class of C-terminal PHD finger possessing protein family, which exclusively presents in the taxa of Viridiplantae (Ito et al., 2007). Although proteins in this family have a relatively large size of ~680 aa compared with the average Arabidopsis protein size of 403aa (Tiessen et al., 2012), there is only one PHD motif annotated on the C-terminal 60 aa region (Letunic et al., 2015; Finn et al., 2016). The function of the remaining regions of the proteins in this family remains unknown. Although MS1 has been long inferred as a TF or histone reader protein, the lack of DNA binding evidence and the undefined histone PTM type the MS1 PHD domain binds, increases the difficulties in studying the molecular function of MS1. More recently, two detailed studies on the MS1 paralogous gene, *MALE MEIOCYTE DEATH1* (*MMD1*), have uncovered that the MMD1 PHD domain has the strongest affinity to H3K4me2, and a relatively weak affinity to H3K4me3 and H3K9me1/2 *in vitro*. MMD1 is essential to the progression of male meiosis process, in which the Chromosome-Associated Protein D3 (CAP-D3) mediates the meiotic chromosome condensation and the THREE DIVISIONMUTANT 1 (TDM1) and directs the spindle

organisation. MMD1 directly activates the expression of CAP-D3 and TDM1 with the H3K4me2 binding ability of MMD1 PHD domain required (Wang et al., 2016; Andreuzza et al., 2015).

4.1.2 The evolutionary origin of the tapetum and MS1

Land plants successfully colonised terrestrial habitats with the help of many key adaptations, one of which is the development of a durable spore wall structure. The walls of spores and their more derived homologues, pollen, contain a tough and chemically resistant layer comprised of biological polymers called sporopollenin (Wellman, 2004). Early divergent land plants, like mosses, had a tapetum-derived outer peri-spore layer, composed of sporopollenin (Wallace et al., 2011). Because the simplest type of tapetum was found in Bryophytes (**Figure 4.1**), it seems plausible that the tapetum in land plants has evolved to facilitate sporopollenin production and nourish the spore (Wallace et al., 2015). This is also consistent with the fact that more recently evolved land plant groups (gymnosperms and angiosperms) exhibit more complex and stratified exines as a result of their highly developed tapetum. The number of spores that plants can produce is also enhanced with the increase of tapetum nourishing ability (Wellman, 2004). For example, in mosses, owing to the persisting cell walls of tapetal cells, the transport of nutrients and other materials from tapetum to the locular fluid is inhibited, hence in each microsporangium only three or four rows of microspores develop within the encasing tapetum cells. In species of Selaginella, the tapetum cell wall degrades after the tetrads release, nourishing the 20-60 spores within the oval locule. While in most seed plants the tapetum loses its cell walls just before they begin actively producing pollen wall materials, thus

maximising their contact with spores. Therefore, hundreds and sometimes even thousands of pollen grains are produced per locule (Pacini, 1990).

Although sporopollenin production is facilitated by the tapetum in land plants, it also occurs in a variety of diverse algal groups, which lack the structure of the tapetum, including charophytes, which is the sister group to embryophytes (Wellman, 2004). This conflict raises the question as to whether genes in land plants directing tapetal sporopollenin production are conserved in charophytes participating in the sporopollenin biogenesis pathway.

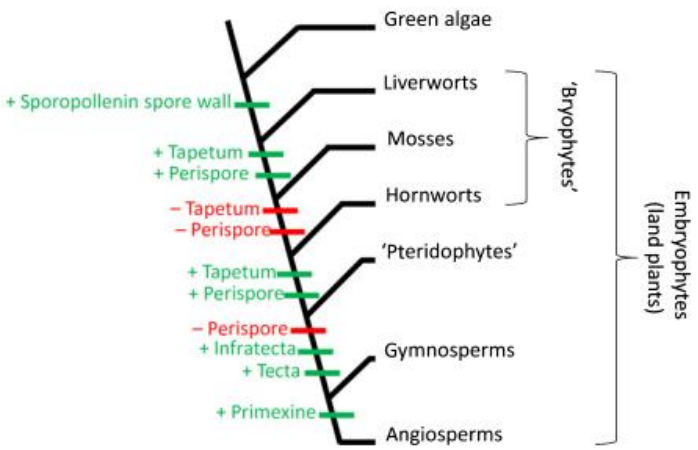


Figure 4.1 Simplified phylogeny of tapetum and spore/pollen wall evolution.

The most basal extant land plants (liverworts) lack a tapetum, which is acquired in mosses and vascular plants. These structures are also absent in hornworts. Taken from Wallace et al., (2015).

The presence of MS1 and MMD1 in Arabidopsis are associated with male reproductive development. MS1 directs late tapetum development, therefore, maintaining the production and secreting of sporopollenin while MMD1 ensures the proper progression of meiosis (Wang et al., 2016; Fernández Gómez and Wilson, 2014). Two homologue proteins of MS1 and MMD1 were reported in basal plant

species moss, which possibly indicates that MS1 and MMD1 play conserved roles in moss tapetum development and meiosis respectively. Although MS1 is proposed as a regulator directing the development of the tapetum in land plants, a single MS1/MMD1 homolog was also found in the chlorophyte species *Ostreococcus tauri* (Ito et al., 2007), which suggests that MS1/MMD1 is a type of protein that is more ancient in land plants occurring prior to the evolution of tapetum.

4.1.3 Nuclear bodies and transcription in plant nuclei

Eukaryotic cells all have defining characteristics: the chromatin is enclosed into a double-membrane organelle termed the nucleus. The nucleus is a complex but well-organised organelle, which harbours chromatin territories and numerous sub-nuclear compartments, collectively referred to as nuclear bodies (NBs) (Reddy et al., 2012). This higher-order nuclear organisation builds a unique and dynamic microenvironment to regulate the genome during transcription, replication, maintenance and the orchestration of these essential processes (Guo and Fang, 2014).

NBs are highly dynamic structures, enriched with proteins, and also RNAs in some cases, which are self-assembled within constrained nuclear foci exerting similar biological processes. NBs of humans and other mammals are well studied and characterised. Numerous mammalian NBs have been characterised, including nucleoli, promyelocytic leukaemia (PML) nuclear bodies, nuclear speckles, Cajal bodies (CBs), polycomb group (PcG) bodies, transcription factories, histone locus bodies and many more (**Figure 4.2**) (Mao et al., 2011). However, the plant NBs are relatively less studied. The characterised plant NBs so far include CBs, nuclear

speckles, nuclear dicing bodies (or D-bodies), photobodies, transcription factories, PcG bodies, cyclophilin-containing speckles, and AKIP1-containing NBs (Guo and Fang, 2014). A few plant NBs that have been linked to the regulation of transcription were decribed as follows:

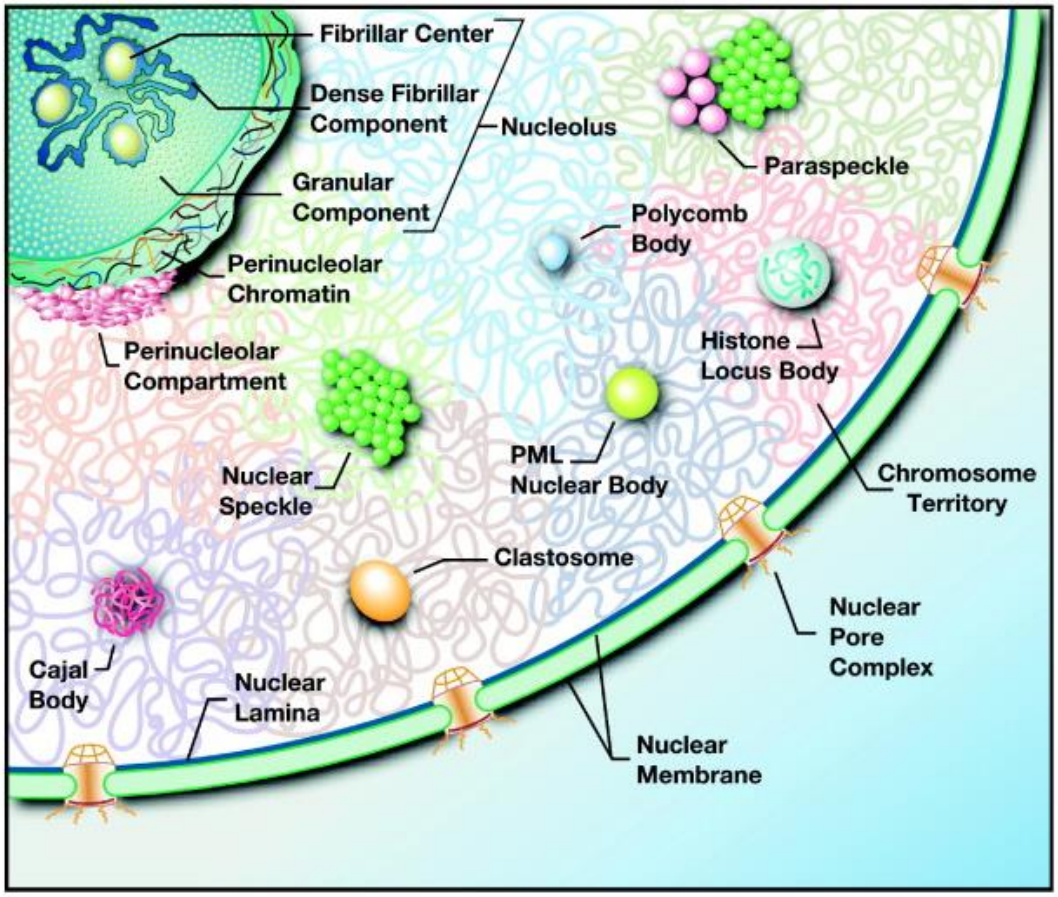


Figure 4.2 A landscape of the diversity of interphase mammalian nuclear bodies.

The interchromatin space is very organised, highly dynamic, and harbours multiple nuclear bodies, such as clastosomes, nucleoli, Cajal bodies, histone locus bodies, paraspeckles, nuclear speckles, PML-nuclear bodies, and Polycomb bodies. Taken from Mao et al. (2011).

Plant CBs have a spherical structure of 0.3-1.0 μm in diameter numbering one to five per nucleus based on cell type and developmental stage. The function of CBs is likely to be involved in snRNAs and snoRNAs assembly, modification, maturation and

trafficking (Li et al., 2008). The characterised protein components include NO CAJAL BODIES 1 (NCB1) and a distant homolog of vertebrate coilin, Atcoilin, both of which are essential for the forming of plant CBs, though the plant lacking CBs shows normal growth and reproduction (Guo and Fang, 2014).

Plant nuclear speckles, also termed as splicing factor compartments (SFCs) or interchromatin granule clusters (IGCs), distribute unevenly at 20-50 sites throughout the nucleoplasm with irregular but discrete domains, which are very dynamic in size (0.5-3.0 μm), shape and numbers depending on the nuclear transcription status and cell environmental conditions (Reddy et al., 2012). The function of nuclear speckles is considered as storage sites for splicing factors and transcription factors and coordinating sites for the pre-mRNA processing. Hence nuclear speckles are often observed closely adjacent to active transcription sites (Spector and Lamond, 2011). It is noteworthy that not all nuclear speckles are homogeneous to each other even in the same nucleus as Reddy et al. (2012) demonstrated that two plant nuclear speckle proteins, AtSR34 and RS2Z33, which belongs to the same subfamily, do not always colocalize in the same speckles.

Dicing bodies (D-bodies) are a small class of rounded NBs (0.2-0.8 μm in diameter), explicitly defined as protein assemblies containing microRNA-processing proteins DICER-LIKE1 (DCL1) and HYPONASTIC LEAVES1 (HYL1). D-bodies are usually distributed diffusely throughout the nucleoplasm, and primarily excluded from nucleoli. The function of D-bodies involves the recruitment and accurate processing of primary microRNAs (Liu et al., 2012).

The photobodies are photoreceptor-containing NBs specifically found in plant cell nuclei. The size, number, organisation and possible function of photobodies are all

directly regulated by light (Van Buskirk et al., 2012). For instance, light triggers the localisation of two classes of photoreceptors, phytochromes (PHYs) and cryptochromes (CRYs), to photobodies via inducing the conformational switch to their active states. Although localisation studies and recent genetic evidence support the theory that photobodies are sites for light signalling events, whether transcriptional regulation is directly linked to photobodies are not known (Chen and Chory, 2011).

Transcription factories and PcG bodies are two class of mutually antagonistic NBs. Transcription factories are discrete sites where transcription occurs in the nucleus. In mammalian nuclei, a transcription factory usually contains 4–30 RNA polymerases II (active or inactive), the transcriptional gene loci, the necessary transcription factors (activators and repressors) and some of many other factors facilitating transcription. The transcription facilitating factors can include but not limited to coactivators, chromatin remodelers, histone modification enzymes, Ribonucleoproteins (RNPs), RNA helicases, and splicing and processing factors depending on the cell type and surrounding environments (Rieder et al., 2012). In plant nuclei, transcription factories do not present as distinct dot-like structure as in mammals, but in a network-like manner, where the RNA polymerases II (RNAPII) are homogeneously dispersed within euchromatin absent from nucleoli and heterochromatin, forming a reticulate structure (Liu and Weigel, 2015). However, more recent studies demonstrated that Arabidopsis active RNAPII aggregate within smaller distances, functioning as mammalian transcription factories (Schubert and Weisshart, 2015). In animals, PcG bodies are hubs where PcG proteins are concentrated for gene repression. PcG proteins form at least two classes of complexes termed as Polycomb repressive complexes 1 and 2 (PRC1 and PRC2), which together or separately silence the

expression of a variety of genes via several mechanisms, including histone post-translational modifications, chromatin compaction, or formation of higher-order chromatin structures (Schubert and Weisshart, 2015). Animal PcG bodies can be clearly visualised in the cell nucleus close to pericentromeric heterochromatin (Perera et al., 2016). Whereas in plants, however, it is still a matter of debate whether PcG bodies are true nuclear entities (Del Prete et al., 2014). In Arabidopsis, LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), a major plant PRC1 component, co-localises and interacts with components of PRC2 in cell nuclei (Derkacheva et al., 2013). The sub-nuclear distribution of LHP1 is constrained in discrete nuclear foci, whose size and number dynamically changes depending on cell differentiation. Hence, LHP1 foci are reminiscent of PcG bodies in animals (Zemach et al., 2006). Nevertheless, whether others PcG proteins take part in the formation of plant PcG bodies remains unknown (Zhou et al., 2015).

Besides these defined NBs, more recent studies on plant TFs and chromatin proteins found novel NBs. The WRKY18 and WRKY40 are transcriptional regulators in ABA signal transduction that co-localize with PIF3, PIF4 and PHYB to a class of novel NBs in a dynamic and phosphorylation-dependent manner. Interestingly, while PIF3, PIF4 and PHYB physically interact, the two WRKY TFs do not interact with PHYB, PIF3 and PIF4, which indicates that the assembly of these NBs is not merely an aggregation of interacting proteins but rather a specifically regulated sub-organelle genesis (Geilen and Böhmer, 2015; Geilen et al., 2015). The MORC proteins belong to a highly conserved protein family with diverse chromatin-related functions. Arabidopsis AtMORC4 and AtMORC7 form homomeric complexes and concentrate in discrete nuclear bodies adjacent to chromocenters to suppress a broad set of non-methylated protein-coding genes required for mounting pathogen response.

Although the function of these NBs is unknown, the subnuclear distribution proximity to the boundaries of highly compacted chromatin implied a potential role of MORCs in DNA superstructure manipulations (Harris et al., 2016).

Over the last decade, advances in nuclei 3D structure studies, has led to increasing knowledge of the regulation of transcription within eukaryotic nuclei (Mao et al., 2011). A compartmentalization paradigm of nuclei functioning was proposed: in interphase cell nuclei, discrete NBs involved in transcription and/or its regulation are dynamically organised into spatially distinct nuclear domains, allowing for more efficient reactions by concentrating the molecular players, reactants, and DNA substrates within a confined nuclear volume (Van Bortle and Corces, 2012). For instance, two major biological regulations of gene expression, transcription activating and silencing, are accomplished through relocation of gene loci from chromosome territories into compartmentalised nuclear foci of transcription factories and repressive nuclear structures respectively (Van Bortle and Corces, 2012). Most transcription regulatory machinery is compartmentalised into transcription factories or PcG bodies, creating distinct sites to facilitate more efficient biological reactions (Politz et al., 2013). In other words, the high-order nuclear organisation governs and reflects the genome function of transcriptional regulation.

In the study of nuclear organisation, live cell fluorescence imaging techniques using fluorescent protein (FP)-tagged reporters provides a powerful means to directly image the protein of interest (POI) within given NBs. The recently developed super-resolution microscopy techniques such as stimulated emission depletion microscopy (STED), has improved both sensitivity and resolution of the imaging, which makes it a perfect means to investigate the nuclear organisation and precise

subnuclear localisation of POI (Trinkle-Mulcahy and Lamond, 2007). STED is a diffraction-unlimited imaging method, which can super-resolve subnuclear structures at a very fast rate. STED can also confine the fluorescence to a minuscule spot on the emission area, which eases the phototoxicity to the living sample. After the image capturing, the final image resolution can be improved by deconvolution algorithms (Komis et al., 2015).

In previous studies of the plant C-terminal PHD possessing proteins (MS1 and MMD1), their exact molecular function in transcription regulation has remained unknown, albeit their importance in Arabidopsis male reproduction development is evident. A lack of DNA binding motif of MS1 and MMD1 also makes their proposed role as TF ambiguous. By the application of the lately developed super-resolution microscopy (Leica TCS SP8 STED 3X confocal) on MS1 FP-tagged reporter, and comparison of its subnuclear distributions with a diverse variety of plant NBs recorded in the database of Brown et al. (2005), Narula et al. (2013) and Rastogi and Rost (2011), the molecular function of MS1 was analysed.

4.1.4 Transcriptomes of the Arabidopsis Anther

Since angiosperms typically have small but complex structured anthers, which accomplishes a complex developmental process in a relatively short time, one of the most challenging aspect in the anther development is differentiating the gene expression specifically of a particular cell type at a given developmental stage from that of other cell types or stages. This is particularly the case when transcriptome profiling of floral buds is used to study the anther-expressed genes, since it also

introduces interference from genes that are expressed in petals, sepals and female reproductive organs.

In Arabidopsis, due to its small flower size, the transcriptome profiling of anthers has frequently been studied using the whole inflorescence or flower, which consists of both male and female organs as well as the sporophytic organs such as petal, sepal and stalk (Huang et al., 2011). This strategy has brought obvious drawbacks. First, substantial amounts of different cells from both sporophytic and gametophytic tissues are used for the transcriptome profiling, making it impossible to distinguish the genes expressed in one cell type from that of the others. Secondly, since the flower developmental progress is not always synchronous within one flower, anther cells from several developmental stages are invariably pooled despite detailed staging and collection. This imprecision makes it difficult to identify the primary genes that cause the original phenotypes rather than the secondary changes in mutant analysis.

Although transcriptome studies of anthers (using flowers) from tapetum mutants have revealed abundant differentially expressed genes (DEGs), large numbers of pollen-expressed genes have also been identified. This is due to the absence of the normal developed microspore/pollen in tapetum mutants. The defects in the tapetum cause subsequent developmental aborted microspores/pollens, which in some mutants degrade or collapse and eventually results in empty locules. The microspore/pollen is very biological active in anther. Among the 28 000 *A. thaliana* genome encoded genes, 6 587 genes were estimated as expressed in mature pollen grains, with a proportion of 11% selectively expressed and 26% enriched than any vegetative tissue (Pina et al., 2005). Additionally, most tapetum mutants have very abnormally vacuolated tapetal cells at early to middle phases, which result in dysfunctional or

even necrotic late-phase tapetal cells, which have a heavily disordered gene expression pattern. Thus, a series of tapetum mutant share a high proportion of their DEGs, which are related to late tapetum function and microspore/pollen development, yet the direct targets of the mutant genes are not easily extracted from the transcriptomic data.

Recent advances are helping to address these issues; laser capture microdissection (LCM) of samples enables isolation of specific cell population from anthers and accurate staging with the help of microscopic visualisation when doing the laser dissection. Combined with the microarray technology, the fine isolation and characterization of staged tapetal cell transcriptome become available. A previous lab member, Dr Gema Vizcay-Barrena, carried out an LCM two-colour-array profiling on WT tapetum at the stage of tetrads release (St. 8), pollen mitosis I (PMI; St. 11) and pollen mitosis II (PMII; St.12), which provides invaluable transcriptomic data to analyse MS1 molecular function and its direct regulatory targets (unpublished data). The latest next generation sequencing (NGS; RNA-Seq) (Goodwin et al., 2016) are providing unbiased detection of novel, rare and low-abundance transcripts with a broader dynamic range (Zhao et al., 2014). The application of RNA-Seq on anther development has been recently reported (Niederhuth et al., 2013; Chen et al., 2010; Dukowic-Schulze and Chen, 2014; Bai et al., 2015; Chen et al., 2015; Zhao et al., 2014).

With these advances in post-genomic technologies, a vast number of genome-wide transcriptomic data sets have been generated, which often focused on identifying specific changes such as the DEGs in a transcription factor mutant compared with a WT. Moreover, most of these data sets are publicly accessible online. Here, a

transcriptome meta-analysis was carried out using a combination of Arabidopsis anther RNA-Seq and microarray data to predict MS1 molecular function and its direct regulatory targets.

4.2 RESULTS

4.2.1 Molecular phylogenetic study of MS1 protein family

MS1 belongs to a small class of protein family featured with the C-terminal possessed PHD finger, which were hereafter referred to as C-terminal PHD possessed proteins (CPCPs). To elucidate the evolutionary conservation and distribution of MS1, the full-length MS1 protein was used as the query to screen its homologs among the public protein databases, including Phytozome, pico-PLAZA and the National Centre for Biotechnology Information (NCBI). It appears that MS1 homologs are only ubiquitously present in green plants (Viridiplantae), while proteins from the primitive algae, such as red algae (Rhodophyta) and Glaucocystophyta, only have low homology with MS1 in the PHD region. This suggests the emergence of CPCP in green plants preceded the plant colonisation of land, which makes CPCP an ancient protein family. The CPCP sequences from a selected group of species, including two Angiosperms, *A. thaliana* and *O. sativa*, one bryophyte, *Physcomitrella patens* and seven green algae, *Chlorella* sp. NC64A, *Chlamydomonas reinhardtii*, *Ostreococcus tauri*, *Ostreococcus lucimarinus*, *Ostreococcus* sp. RCC809, *Bathycoccus prasinos* and *Micromonas pusilla* strain CCMP1545, were used to construct a phylogenetic tree of the CPCPs. The topology of the trees clearly showed that MS1 homologs appear to cluster into a single ortholog group within the land plants (Embryophyta) and green

algae (Chlorophyta) respectively (**Figure 4.3 A**). Unlike most plant TF families, the CPCP does not go through a huge expansion of the family size. Up to four CPCP family members exist in higher plant species, such as Arabidopsis, and only one to two family members exist in the primitive plant species. Notably, for the two CPCPs possessed by the green algae species, such as *Chlorella* sp. NC64A and *Chlamydomonas reinhardtii*, only one shows a similar protein length to MS1, while the other is much longer, with the MS1 homologous sequences interspersed by non-homologous sequences. These extended MS1 homologous proteins may represent gene products of species-specific new genes originated from chimeric gene fusions. However, the possibility that the higher plant MS1 homologs originate from the ancient CPCPs after several rounds of gene fission events during the plant evolution cannot be excluded. Furthermore, the trees obtained here are consistent with the phylogenetic tree of plant species as it is now known (**Figure 4.3 B**) (Sakakibara, 2016). Thus, the CCP proteins represent an evolutionarily stable lineage of proteins that have constrained function.

In the phylogenetic tree produced here, the angiosperm CPCPs were clustered into three subfamilies, which respectively contain the MS1 tapetum regulators, MMD1 meiosis regulators and a group of uncharacterised CPCPs. Arabidopsis and rice do not have any species-specific CPCP, which suggests that the CPCP family that presents in the last common ancestor of monocots and core eudicots may have obtained the function of extant CPCPs. Interestingly, the moss CPCPs neither cluster with MS1, nor with MMD1, although tapetum development and meiosis were essential for the life cycle of moss. Wallace et al. (2015) reported a moss *MS2* gene playing a conserved function in spore and pollen wall development and suggested an evolutionarily conserved early sporopollenin biosynthetic pathway. Whether the

regulatory role of MS1 in this pathway is also conserved in moss needs further investigation.

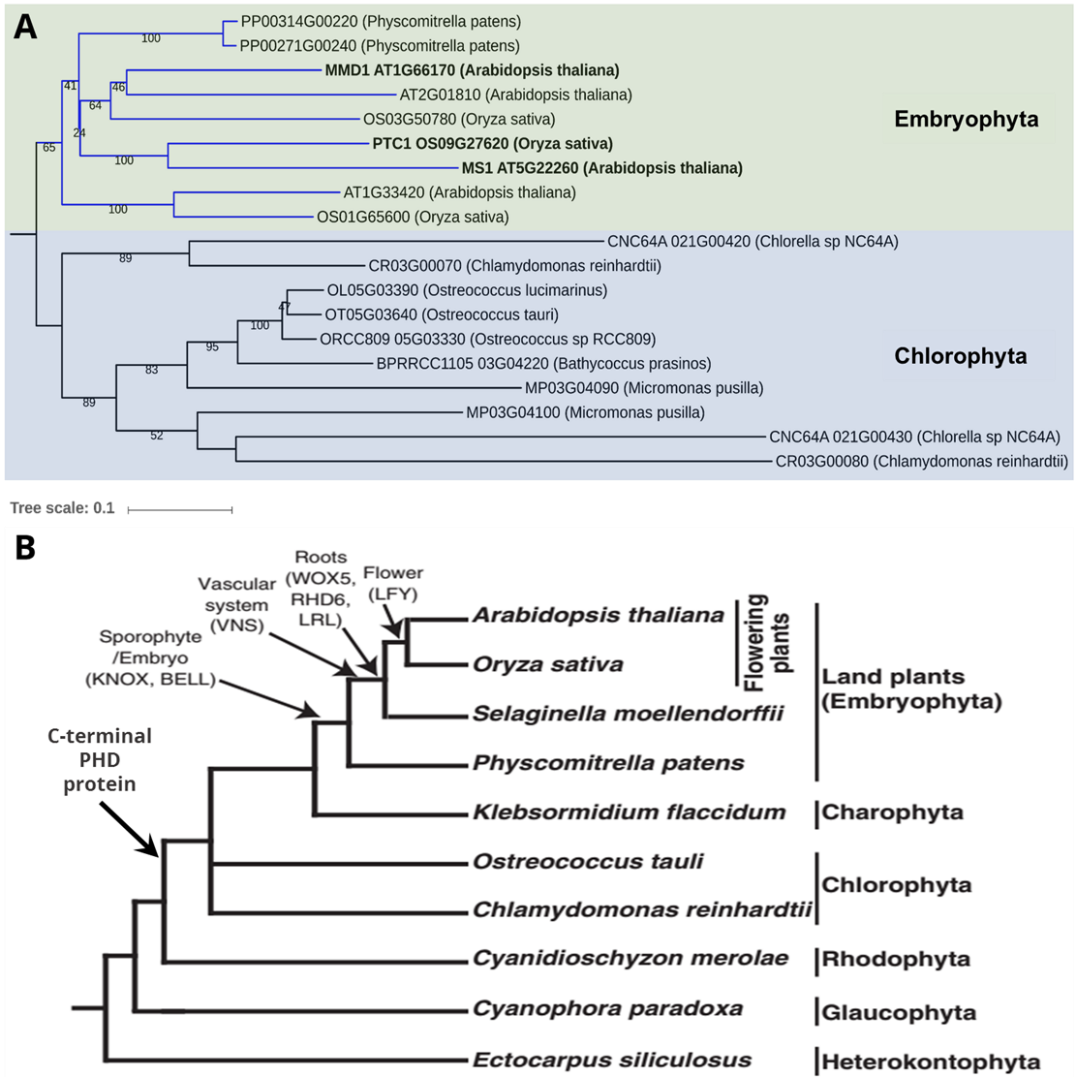


Figure 4.3 Phylogenetic tree of the C-terminal PHD proteins of selected representative species of Chlorophyta and Embryophyta.

(A) The proteins sequence used for tree construction are from Embryophyta species, including Arabidopsis (bold type), rice (*Oryza sativa*, bold type) and moss (*Physcomitrella patens*), and Chlorophyta species, including *Chlorella* sp. NC64A, *Chlamydomonas reinhardtii*, *Ostreococcus tauri*, *Ostreococcus lucimarinus*, *Ostreococcus* sp. RCC809, *Bathycoccus prasinos* and *Micromonas pusilla* strain CCMP1545. The numbers next to the branches indicate replicated values using the bootstrap test (100 iterations). The tree scale bar 0.1 represents the evolutionary distance of 0.1 amino acid substitutions per site. The tree is based on the stripped protein sequence alignment of pico-PLAZA gene family ORTHO000988, with non-homologous positions and diverged sequence removed. (B) Phylogenetic relationship among species used for comparative analysis adapted from Sakakibara (2016). Arrows indicate the timing of acquisition of morphological innovations and genes related to the morphological innovation.

4.2.2 Characterisation of The MS1 protein functional domains

MS1 encodes a CPCP which is required for normal tapetum development and male fertility in Arabidopsis, and the PHD finger is essential for the MS1 function (Yang et al., 2007; Ito et al., 2007). PHD fingers are reported to bind to the lysine residue of Histone 3 peptides (Mouriz et al., 2015), and more recently, Andreuzza et al. (2015) and Wang et al. (2016) demonstrated that the PHD domain of MS1 paralogous gene MMD1 could specifically bind to H3K4me2/3, and weakly binds to H3K9me1/2 *in vitro*. To further investigate possible binding activities of the MS1 PHD finger, a Clustal Omega alignment comparison were first made using the protein sequences of the MS1 PHD domain (MS1 612-661 aa) and other different types of known H3K4me2/3 binding PHD fingers. The alignment showed that MS1 PHD contains a Cys4-His-Cys3 conserved motif (**Figure 4.4 A**), which is delineated to form a Zn ion stabilised aromatic cage that interacts with different modified histone peptides (Mouriz et al., 2015). Additionally, key amino-acid residues that determine the binding specificity of the PHD domain are conserved among the aligned sequence,

forming a consensus (**Figure 4.4 A**). To test whether the MS1 PHD finger can bind modified histone peptides, *in-silico* molecular modelling was conducted using the MLL5 PHD domain as a template. The human MLL5 PHD domain is homologous to the MS1 PHD domain and specifically binds to histone mark of H3K4me2/3 (Lemak et al., 2013). The crystal structure of the hetero-oligomer of the MLL5-PHD-H3K4me3 complex has been resolved at the resolution of 1.48 Å using X-ray diffraction (Ali et al., 2013) (**Figure 4.4 B**), which provides a structural template for MS1 molecular modelling. The modelled MS1 PHD domain perfectly matches the H3K4me3 molecular in three-dimensional conformation, and the 625D, 629M, 638W and 653P amino acid residues in the alignment consensus distribute on the interacting interface between the PHD domain and H3K4me2/3, assembling the spatial configuration of MLL5-PHD-H3K4me3 complex, which suggests the MS1 PHD has a similar binding affinity with H3K4me3 (**Figure 4.4 A**, **B** and **C**).

As PHD fingers typically act together with other domains (reviewed in **Section 4.1.1**), the MS1 protein regions besides the C-terminal PHD finger were further investigated via a series *in-silico* analysis. Firstly, the MS1 protein sequence without PHD finger was used as the query to screen the Pfam database version 30.0 (Finn et al., 2016), however, no Pfam domains matched the query sequence. A PRALINE alignment was subsequently conducted using the MS1 protein sequence, its three paralogs and two Arabidopsis genes, *AT2G07714* and *ATMG00550*, which show homology to MS1 but lack the PHD domain. A prominent level of conservation was found at the 212-305 aa and 326-387 aa region of the MS1 protein (**Figure 4.5 A**), whereas in the other parts of the protein, the amino acid residues were quite diverse. Remarkably, these two regions were also found highly conserved in other (Viridiplantae) CPCPs whose protein sequence were used to construct the phylogenetic tree in **Section 4.2.1**.

Conserved amino acid signatures are often linked to protein functions. Hence these two evolutionarily constrained protein regions might exert important biological features that cooperate with the PHD histone reader to accomplish the regulatory role of MS1/MMD1 or other CPCPs. These protein regions, therefore, were termed as MS1/MMD1 Associated Domains (MAD) 1 and 2. Interestingly, Yang et al. (2007) reported that an internal GFP-tagged MS1 chimeric fusion protein, which had the GFP tag inserted at 366aa and broke the MAD2 (**Figure 4.5 B**), failed to rescue the phenotype of *ms1*. This may highlight the importance of the MADs for MS1 protein function.

Additionally, a nuclear localisation signal peptide and multiple ubiquitination sites with high prediction confidence were found through a series of online prediction tools (**Figure 4.5 B**). Verifying the ubiquitination of these densely-distributed sites along the MS1 protein may be important for the study of ubiquitin-mediated proteolysis of MS1 (discussed in **Chapter 3**

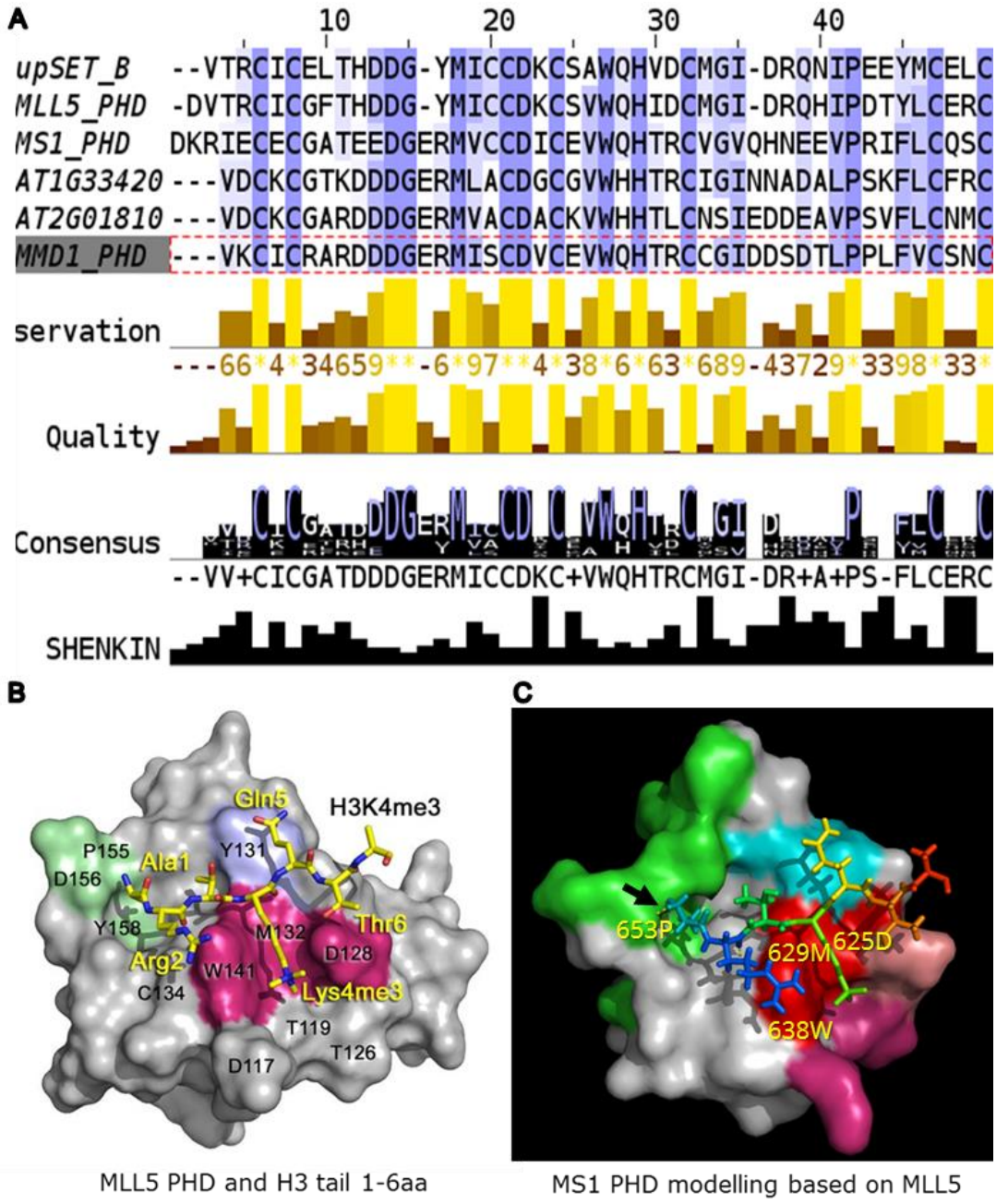


Figure 4.4 Comparative study and molecular modelling of MS1 PHD finger.

(A) ClustalOmega alignment of the MS1 PHD with known PHD fingers from different organisms. Besides the Arabidopsis CPCP PHDs, two H3K4me2/3 bind PHD domains (human MLL5, Drosophila UpSET-B) was used for the alignment. (B) Crystal structure of the hetero-oligomer of the MLL5-PHD-H3K4me3 complex. (C) SWISS-MODEL molecular structure for the hetero-oligomer of the MS1-PHD-H3K4me3 complex. The covalent bond between 1Alanine of H3K4me3 and MS1 PHD is indicated by the black arrow and yellow dashed line. The residues labelled in yellow are the conserved amino acids for recognition of H3K4me3.

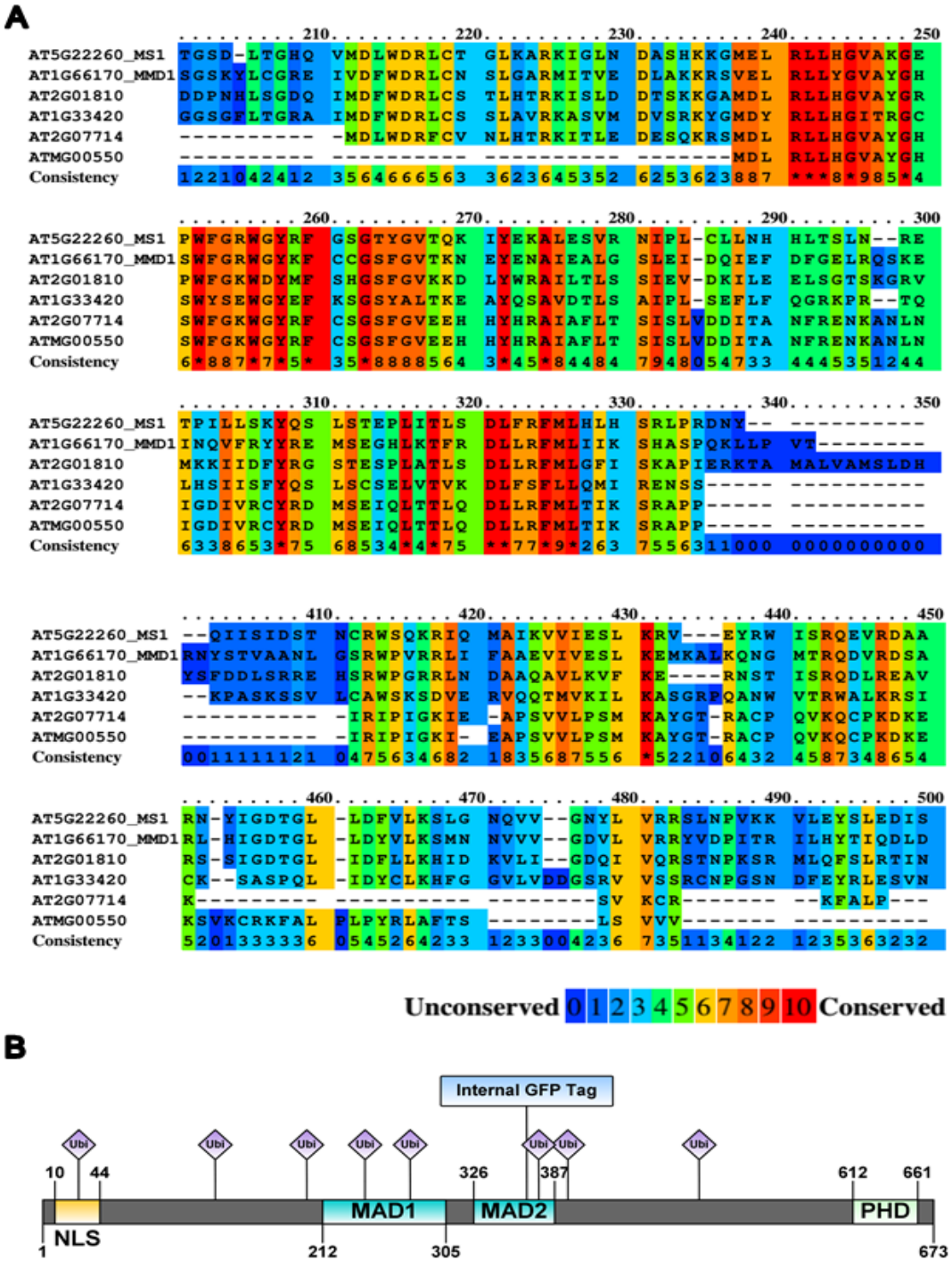


Figure 4.5 PRALINE alignment of MS1 and the related Arabidopsis proteins and the graphical representation of MS1 protein.

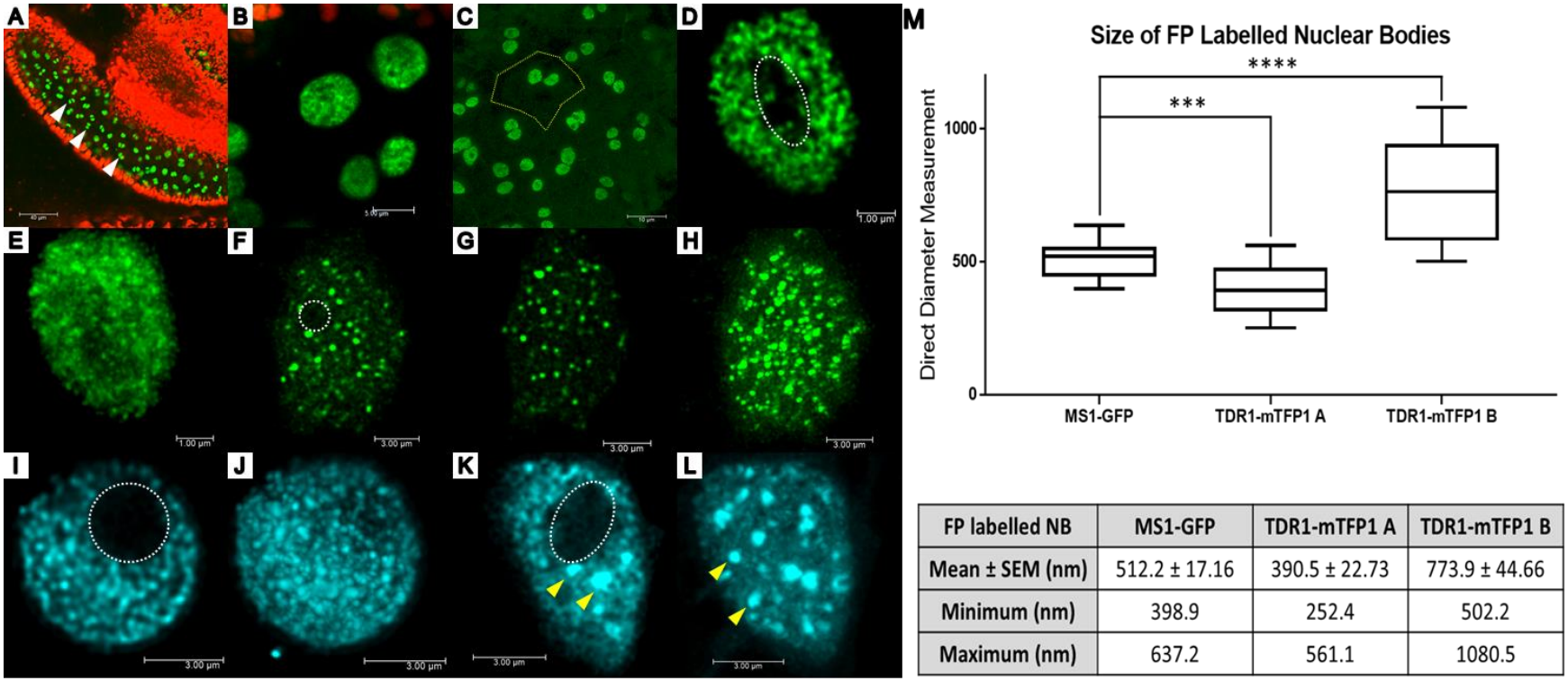
(A) The 212-305 aa and 326-387 aa regions of MS1 protein showed a prominent level of conservation with its Arabidopsis relatives. The most conserved residues are scored as 10 and indicated by red, the least conserved ones are scored as 0 and indicated by blue. (B) A schematic diagram showing the predicted domains, PTM sites and internal GFP tag location of the MS1 protein. NLS, nuclear localisation signals; Ubi, predicted ubiquitination sites.

4.2.3 The MS1 protein has a specified subnuclear localisation

MS1 and other CPCPs do not belong to canonical TF as they lack characterised DNA-binding domains. However, they appear to strongly participate in transcription regulation, possibly as a transcription coregulator that modifies chromatin structure. The subnuclear distribution of MS1 and its associated NB may provide an informative clue of the molecular mechanism of MS1 functioning. To examine the subnuclear localisation of MS1, anthers from stable MS1pro:MS1-GFP transgenic line (Yang et al., 2007) were collected for confocal microscopy. The MS1-GFP signals obtained were only observed in the nuclei of binucleate tapetal cells of anthers at tetrad release stage (**Figure 4.6 A**, **B** and **C**). The GFP signal were transiently presented in a subset of anther locules within one flower, which is consistent with the previously observed relatively short lifespan of the MS1 protein (Yang et al., 2007). By using the super-resolution STED technique and the image deconvolution processing algorithm, a zoomed inspection of the tapetal nuclei showed that the MS1-GFP was absent from the nucleolus and distributed in the nucleoplasm as a combination of punctate domains and a continuous reticulate background (**Figure 4.6 D** and **E**). Due to the small size of the tapetal nuclei and the auto-fluorescence interference from other cell types of the anther, it is hard to improve the resolution of the confocal images showing MS1-GFP subnuclear localisation. Furthermore, the thickness of anther

weakened the sensitivity of the STED technique. The transiently expressed MS1-GFP in *N. benthamiana* leaf epidermal cells, which has larger nuclei, was adopted to increase the resolution of confocal observation. A rice tapetum bHLH TF, *Tapetum Degeneration Retardation 1* (*TDR1*), was fused with mTFP1 and transiently expressed in *N. benthamiana* leaf as a control, which shows the subnuclear localisation of a canonical TF (Li et al., 2006b). In the *N. benthamiana* nuclei, MS1-GFP remained absent from the nucleoli, and localised at distinctly discrete nuclear foci of regularly round-shape speckles, while the continuous reticulate structure within the nucleoplasm was no longer obtained (**Figure 4.6 F**, **G** and **H**). Remarkably, the TDR1-mTFP1 showed two distinctly different subnuclear distribution patterns (referred to as pattern A and B). Pattern A exhibited densely small round dots connected by a continuous network manner (**Figure 4.6 I** and **J**), which is analogous to the RNAPII indicated transcription factories (Liu and Weigel, 2015). Whereas pattern B showed large irregularly shaped speckles within an unevenly smeared reticulate background labelling (**Figure 4.6 K** and **L**), which assembled the SFC subnuclear distribution (Reddy et al., 2012). Both patterns A and B exclude from the nucleoli and seem correlated with the shape of the nucleus, as pattern A was always present in globular nuclei, while pattern B mostly appeared in nuclei with irregular shapes (**Figure 4.6 I** and **K**). The different nucleus shapes might represent different transcription states of the chromatin and developmental stage of the cell, which needs the TFs allocated into the specific apartment of the nucleus. The mean speckle size of labelled by MS1-GFP and TDR1-mTFP1 was different with each other. The largest speckle appears in TDR1-mTFP1 pattern B, with a diameter of 1.08 μm, while the smallest speckle appears in TDR1-mTFP1 pattern A, with a diameter of 252.4 nm (**Figure 4.6 M**). The diameter of MS1-GFP speckles ranged from ~400 nm

to ~600 nm (**Figure 4.6 M**) and exhibited a discrete and sparse distribution in the nucleoplasm compared with that of the TDR1-mTFP1 control, which resembles the subnuclear localisation Arabidopsis LHP1 in mesophyll nuclei (Zemach et al., 2006).



FP labelled NB	MS1-GFP	TDR1-mTFP1 A	TDR1-mTFP1 B
Mean ± SEM (nm)	512.2 ± 17.16	390.5 ± 22.73	773.9 ± 44.66
Minimum (nm)	398.9	252.4	502.2
Maximum (nm)	637.2	561.1	1080.5

Figure 4.6 STED confocal images of the subnuclear distributions of the MS1-GFP and TDR1-mTFP1 fusion reporter.

(A-E) The MS1-GFP fusion reporter driven by the native promoter in MS1pro: MS1-GFP stable transgenic line. (A) Anther locule at tetrad stage encasing binucleate tapetal cells with the MS1-GFP signals indicated by the white arrowhead. (B) The MS1-GFP signal was constrained to the nuclei. (C) A single layer of tapetal cells. One of the binucleate tapetal cells was indicated by a yellow dashed frame. (D) A zoomed inspection of one tapetal nucleus. (E) A maximum projected image of Z-stack capturing for (D). (F-H) The MS1-GFP fusion reporter transiently expressed in a *N. benthamiana* leaf. (F) and (G) MS1-GFP labelled NB. (H) A maximum projected image of Z-stack capturing of (F) and (G). (I-L), The TDR1-mTFP1 fusion reporter transiently expressed in a *N. benthamiana* leaf. (I) and (K), TDR1-mTFP1 labelled nuclei showing two different TDR1 subnuclear distribution pattern. (I), Pattern A. (K), Pattern B. (J) and (L) are the maximum projections of Z-stack images of (I) and (K) respectively. (M) Box plot of measured diameters of the fluorescent protein (FP) labelled NBs and their calculated statistics.

4.2.4 Meta-analysis of the anther and MS1 microarrays

In previous work, two sets of transcriptomic profiling data were available for *MS1*. One was generated on young (stage PMC to PMI) and old floral buds (stage PMII) using Affymetrix ATH1 microarrays (Yang et al., 2007). The other is a time-series transcriptomic profiling of seven staged floral bud samples using Agilent two-colour arrays(Alves-Ferreira et al., 2007). Pearce et al. (2015) renormalized both sets of *MS1* microarrays with other Affymextrx ATH1 and Agilent two-colour arrays data respectively, making these data comparable to each other. The renormalized data was deposited in a publicly accessible database (Arabidopsis Anther/Floral Expression Plots https://www.cpib.ac.uk/anther/), which provides an immediate visual representation of how the genes are expressed in the WT and *ms1* mutants across various stages. Combining all the DEGs between *ms1* and wild-type in each stage, there are 623 up-regulated and 1,301 down-regulated genes in the *ms1* mutant compared with wild type.

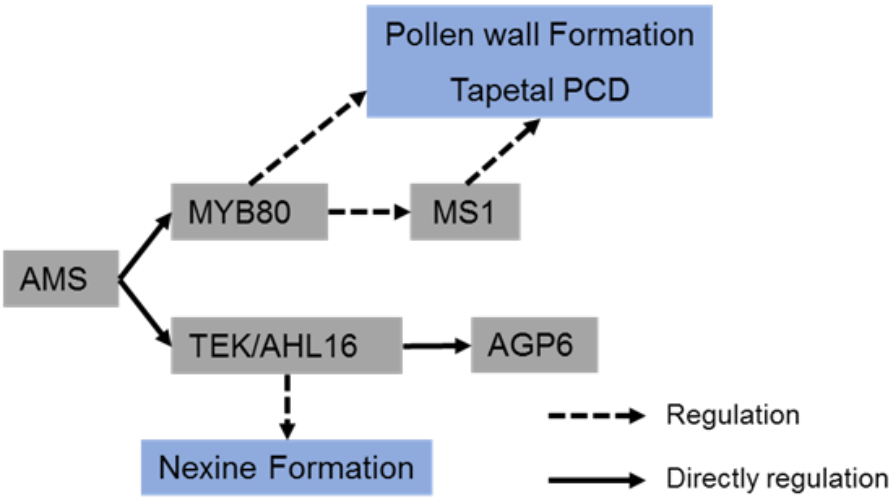


Figure 4.7 The regulatory pathway of selected Arabidopsis tapetum key TFs.

To determine whether these DEGs are *ms1* specific, they were compared with *MYB80* and *TEK* microarray data and analysed by GO enrichment. *MYB80*, which functions upstream of *MS1* in the tapetum developmental pathway, has a spatial and temporal expression pattern that resembles that of *MS1* (Xiong et al., 2016; Phan et al., 2012). Meanwhile, *TEK* expresses at a similar stage in the tapetum with *MYB80* and *MS1*, but functions in a parallel pathway regulating pollen nexine formation (**Figure 4.7**). In the *tek* mutant, the expression of both *MS1* and *MYB80* were not affected (Jia et al., 2015). The Venn diagram analysis of the DEGs of the *MS1*, *MYB80* and *TEK* microarrays showed that they all shared a significant amount of downstream targets (**Figure 4.7**, **Figure 4.8 A** and **B**). A significant number of *MS1* DEGs overlap with those of *MYB80*, which is consistent with the fact that *MS1* and *MYB80* act in the same pathway. However, a significant propotion of MS1 DEG also overlap those of *TEK*, which indicates these shared DEG are possibly not a immediate consequece of the mutant gene. The common up and down-regulated DEGs between *TEK* and *MS1* were submitted for a Biological Process GO enrichment analysis, while the total number of DEGs of *MS1*, *TEK* and *MYB80* were submited as background reference genes. A series of GO terms which are closely related to the cell stress responding and mature pollen were significantly enriched (**Figure 4.7**, **Figure 4.8 C** and **D**), which suggests that *MS1* DEGs include a considerable number of non-specifically altered genes that also change in other aberrant anther mutants, mostly due to defective anther development rather than the abcence of the funtional gene. Although *MS1* DEGs were used as a resource to screen MS1 regulatory targets (Yang et al., 2007; Alves-Ferreira et al., 2007; Ito et al., 2007), the large size and the heterogeneity of *MS1* DEGs make it defficult to identify the genuine regulatory targets from expression fluctuation noise.

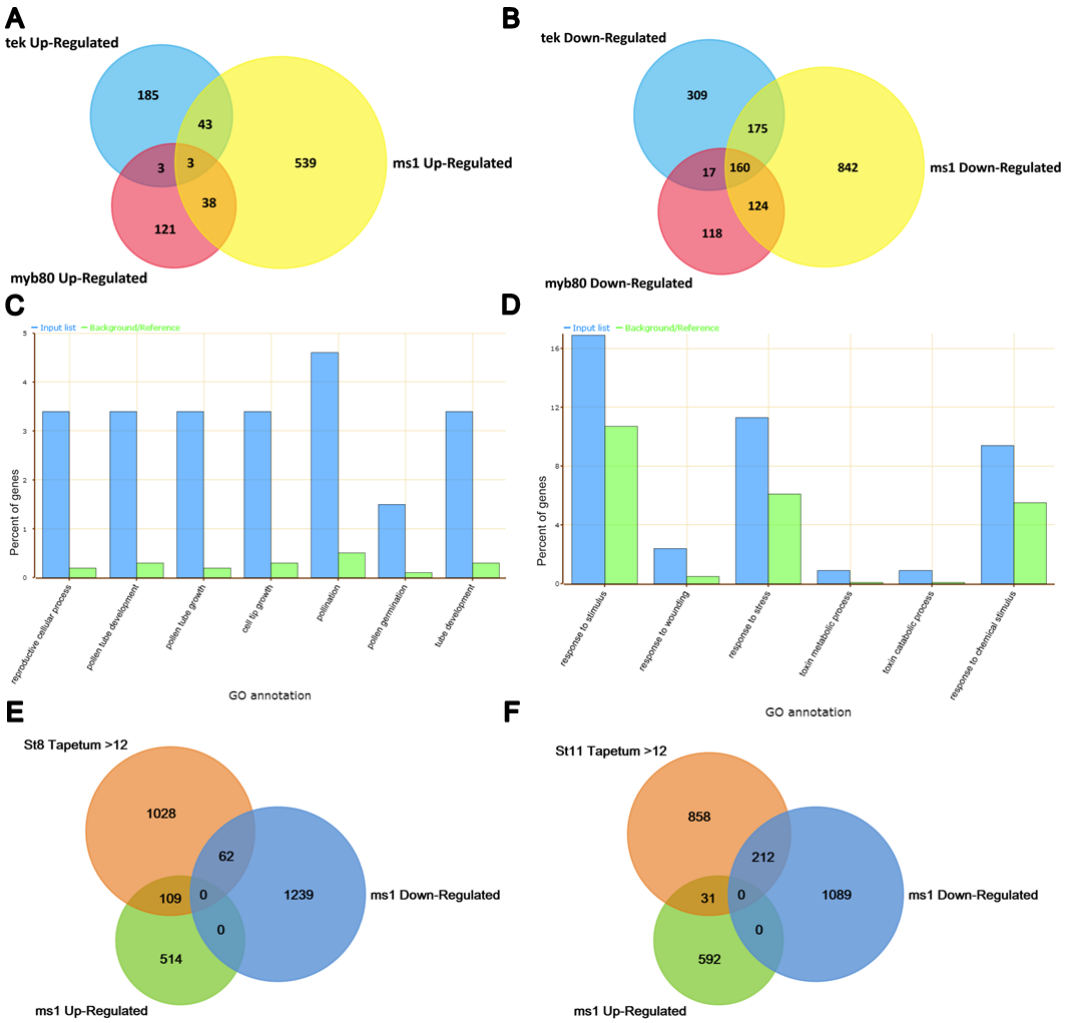


Figure 4.8 Venn diagram dissections and biological process GO analysis of the 2-fold DEGs of *MS1*, *MYB80* and *TEK*.

(A) and (B) The Venn diagram among *MS1*, *MYB80* and *TEK* for the up (A) and down-regulated (B) DEGs. (C) and (D) Presentation of the selected GO biological process terms enriched in common down (C) and up-regulated (D) DEGs shared by *MS1* and *TEK*. (E) and (F) The Venn diagram among the tapetum enriched genes and the up and down-regulated DEGs of *MS1*.

In order to avoid the interference of the non-specific DEGs, such as genes only expressed in pollen or very late tapetum stages, the immediately responding genes (IRGs) of *MS1* were isolated from the DEGs. Theoretically, the *MS1* IRGs should have expression patterns that temperally and spacially aproximate that of *MS1*. Hence,

the *MS1* DEGs were analysed by Venn intersection with the tapetum high expressed genes at the tetrads release stage and the PMI stage respectively (see **Section** 2.7.3). Four sets of putative MS1 IRGs were generated, *i.e.* 109 up-regulated and 62 down-regulated IRGs at tetrads release stage, and 31 up-regulated and 212 down-regulated IRGs at PMI stage (**Figure 4.8 E** and **F**). Interestingly, not like the fact that the down-regulated DEGs predominating in the total DEGs (down-regulated / total DEGs = 68%), the tetrads release stage down-regulated IRGs was much less than the up-regulated IRGs (down-regulated : total tetrads release stage IRGs = 36%), which suggests that the first responding genes are mostly repressed by *MS1* rather than activated.

To further resolve the putative direct targets of MS1, the dynamics of the expression patterns of MS1 IRGs were analysed by clustering and heat map analysis using the time-series microarray data (Alves-Ferreira et al., 2007) and LCM two-colour array data (**Figure 4.9**-**Figure 4.11**).

As most of the up-regulated IRGs at the tetrads release stage are included in the PMI stage, these two IRG sets were merged for the Clustering and heat map analysis. The up-regulated IRGs were clustered into three groups based on their expression level and duration in the WT (**Figure 4.9 A**). All three Groups showed higher levels and longer duration of the expression in *ms1* mutant compared to those in WT, meanwhile exhibited overall decreased expression level as the group number increase (**Figure 4.9 A**). A close inspection of the 24 genes in group I showed that a considerable proportion of these IRGs function in the sporopollenin metabolic pathway (**Figure 4.9 B and Table 4.1**). Genes in Group II exhibited a weaker and more transient expression pattern compared to those of Group I genes. A few tapetum functioning

TFs presents in group II, such as *bHLH91*, *MS1* and *TEK*, while the rest are mostly uncharacterized genes (**Figure 4.9 B** and **Table 4.1**). Group III genes have relatively low expression in WT anthers. Among them, there are a few genes showed clear up-regulation rather than prolonged expression in the *ms1* mutant compared to those in WTs. These genes have no expression in WT anthers, but show a relatively high expression level in *ms1* anthers (**Figure 4.9 B** and **Table 4.1**).

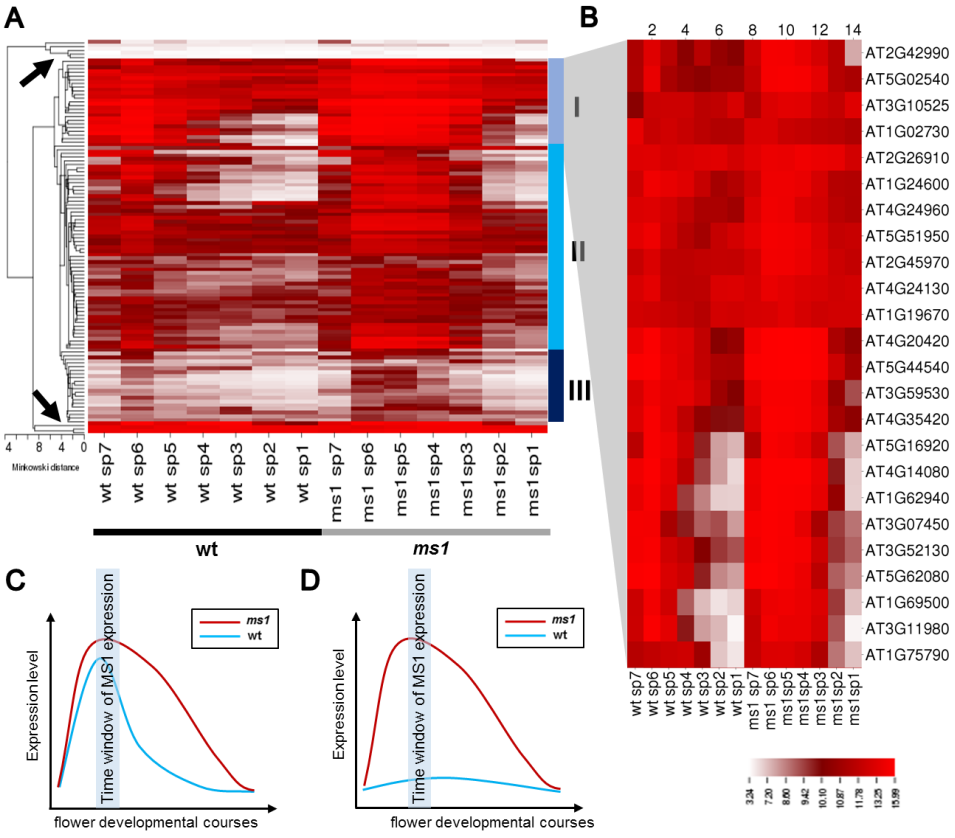


Figure 4.9 Microarray Clustering and heat map analysis of up-regulated IRGs and two proposed paradigmatic expression patterns.

(A) Heat map presenting the row clustered up-regulated IRGs, data in column 1-7 and column 8-14 are normalised time-course expression value from seven samples of wild-type and the *ms1* floral buds (Alves-Ferreira et al., 2007). Constitutively low-expressed and high expressed genes were indicated by black arrows at the top and bottom of the heat map respectively. The remaining up-regulated IRGs clustered into three groups. (B) A zoomed display of IRGs in group I. (C) and (D) Proposed paradigms showing the expression pattern of prolonged (C) and genuinely up-regulated (D) IRGs. Sp., Sample.

To summarise the expression characteristics of up-regulated IRGs, two paradigmatic expression patterns are proposed (**Figure 4.9 C and D**). The elevation of expression level of most up-regulated IRGs in *ms1* was predominantly attributed to the prolonged expression duration (**Figure 4.9 C**), while a minority of the up-regulated IRGs showed a genuine up-regulation in *ms1* with a quick prompt response and prominent level of expression elevations (**Figure 4.9 D**). Some of these genes (**Table 4.1**) are related to ethylene signalling, which might represent a stress response to the disordered internal environment of *ms1* tapetal cells. However, the possibility that they are directly regulated by MS1 cannot be excluded.

The clustering and heat map analysis of down-regulated IRGs at the tetrads release stage using data of time-series microarray (Alves-Ferreira et al., 2007) divided the genes into three groups. Genes in Group I and III are constitutively high-expressed and low-expressed respectively across stages even in both the WT and *ms1* mutant (**Figure 4.10. A**). Although these genes showed overall 2-fold differential expression between the WT and *ms1* mutant, their time-series expression pattern revealed that MS1 has little impact on their early tapetum expression, indicating genes in Group I and III are less likely to be direct regulatory targets of *MS1*. Whereas in Group II, genes have high expressions early in the WT with only a limited duration, and significantly reduced expression in the *ms1* mutant (**Figure 4.10. A**). The 38 genes in group II include two putative tapetal functioning TFs, *MYB99* and *LOB DOMAIN-CONTAINING PROTEIN 2* (*LBD2*), which have been presumed to be direct targets of *MS1* (Alves-Ferreira et al., 2007). Additionally, quite a few tapetum functioning genes were found in Group II, including *ATP-BINDING CASSETTE G29* (*ABCG29*), which is a monolignol transporter involved in lignin biosynthesis (Alejandro et al., 2012), three *UDP-glycosyltransferase* genes, which encode

modification enzymes that possibly determine the pollen-specific flavonol structure (Yonekura-Sakakibara et al., 2014), four pectin lyase-like superfamily genes, which were reported to be involved in pollen intine formation in Chinese cabbage (Jiang et al., 2014b, 2014a), and two O-glycosyl hydrolases family 17 genes (**Figure 4.10 B** and **Table 4.1**). These tapetal functional genes constitute a promising candidate set for direct regulatory targets of *MS1*.

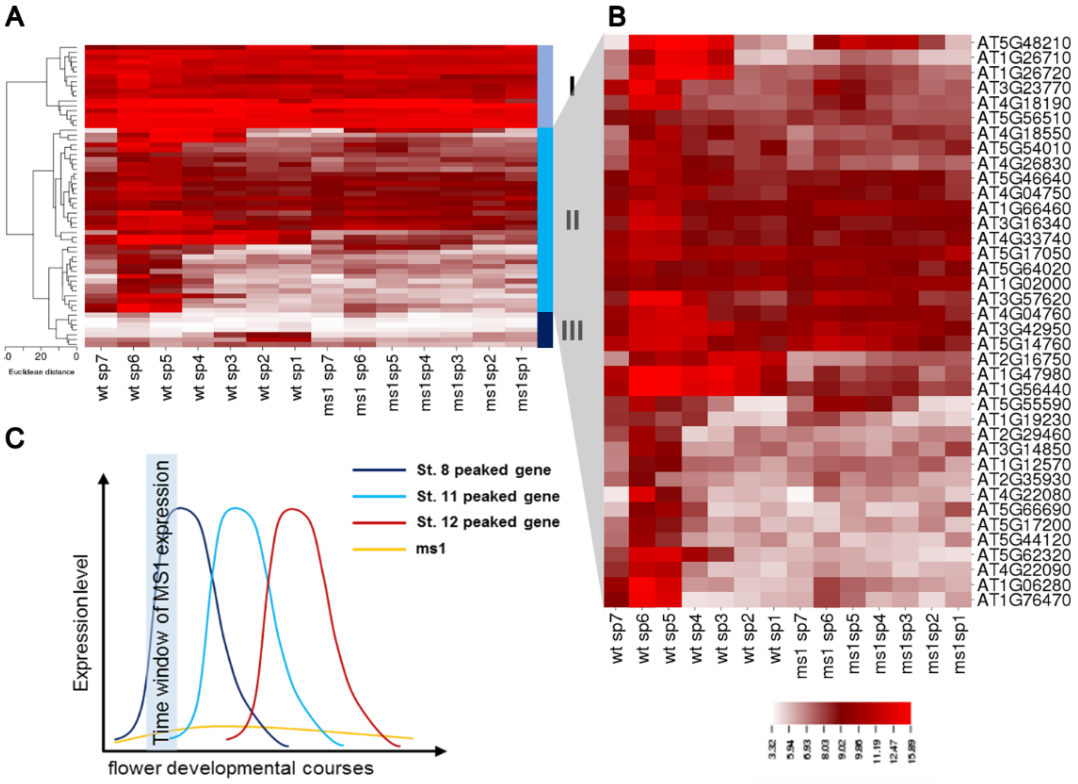


Figure 4.10 Microarray Clustering and heat map analysis of the down-regulated Stage8 IRGs and a proposed paradigmatic expression patterns.

(A) Heat map presenting the row clustered down-regulated Stage8 IRGs, data in column 1-7 and column 8-14 are normalised time-course expression values from seven samples of wild-type and the *ms1* floral buds (Alves-Ferreira et al., 2007). The genes are clustered into three group. (B) A zoomed display of IRGs in group II. (C) Proposed paradigms showing the expression pattern of group II similar genes peaked at distinct stages of tapetum development and showed significantly reduced expression in the *ms1* mutant. Sp., Sample.

Paradigmatic expression patterns are proposed to represent the Group II genes and other DEGs that shared an analogical expression trend to Group II genes but was significantly late (**Figure 4.10 C**). These genes do not belong to IRGs, as they are expressed much later than MS1. However, they appear to be quickly responsive to tapetum developmental progression and to confer diverse biological activities to the tapetum. These genes are easier to be selected as MS1 direct target candidates based on their tapetum function specificity and distinct expression pattern. However, this tendency should be avoided as the expression of these genes appears to be too late to be regulated by MS1.

The clustering and heat map analysis of down-regulated IRGs at the PMI stage were conducted differently. As the expression data of many genes within this IRG set are missing from the time-series microarray (Alves-Ferreira et al., 2007), the LCM tapetum-specific two-colour array was used instead, which only contains the data of WT. The down-regulated PMI-stage IRGs were clustered into three groups, all with expression peak at the PMI stage. Group I IRGs have the longest expression duration among the three group (**Figure 4.11 A**). Three paradigmatic expression patterns are proposed to represent the characteristic expression of IRGs of Group I, II and III (**Figure 4.11 C**). As the PMII stage is relatively far from the expression stage of MS1, and the IRGs Group II and III initiate the expression at the PMII stage, they are less likely to be direct regulatory targets of MS1. Therefore, A close inspection was only conducted on the 40 genes in group I, among which, interestingly, two serine carboxypeptidase-like proteins, *SCPL48* and *SCPL49*, and three lipid-transfer proteins (LTPs) were found. The *SCPL48* was reported to be a conserved PCD indicator gene, whose expression is closely associated with the plant tissues undergoing developmental PCD (Olvera-Carrillo et al., 2015). *SCPL49* is a close paralog of

SCPL48. Hence both genes possibly act redundantly with each other. Although the molecular function of *SCPL48/49* in PCD is unknown, their MS1 dependent tapetum-specific expression highlights the regulation role of MS1 in the tapetal PCD progress. Verifying the direct regulation of *SCPL48/49* by MS1 would provide strong molecular evidence that MS1 initiates and regulates the tapetal PCD progress. The LTPs were previously reported as constituent proteins of the pollen coat (Gong et al., 2015) and essential for pollen exine formation in rice (Zhang et al., 2010). Two variant LTPs, glycosylphosphatidylinositol (GPI)-anchored LTP3/4 were involved in the accumulation of sporopollenin in pollen (Edstam and Edqvist, 2014). The three LTPs included in Group II are close homolog to GPI-anchored LTP4, which provide promising candidates as MS1 direct regulatory target involved in pollen wall formation.

In summary, by a comprehensive analysis of datasets of the time-series, cell-type-specific and pooled tissue expression profiles of the *ms1* mutant and WT, a list of putative MS1 direct targets was obtained. Among these genes, quite a few have reported functions closely linked with late tapetum activity and pollen development, which are listed in **Table 4.1** and provided promising directions for future MS1 functional research.

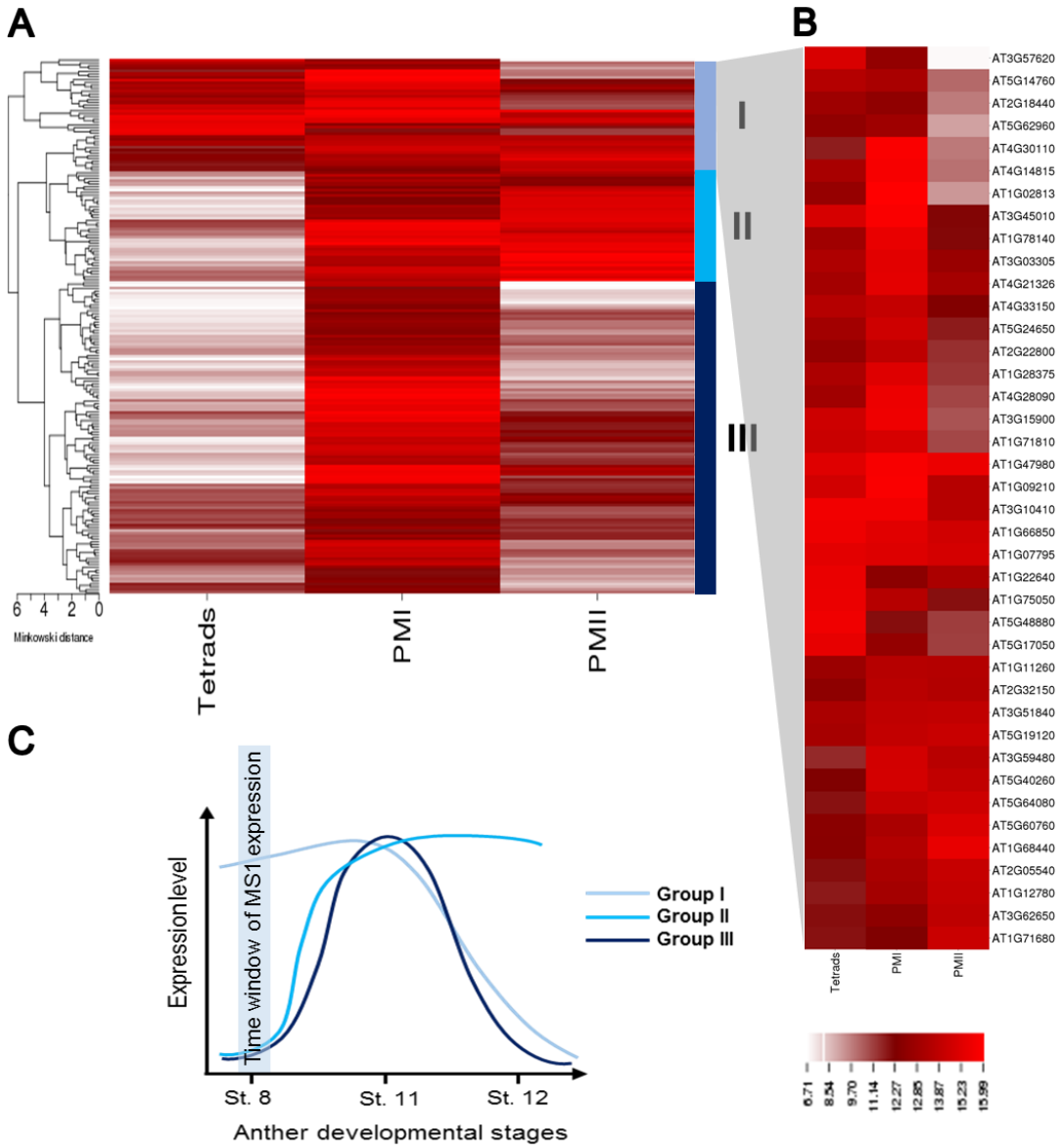


Figure 4.11 Microarray Clustering and heat map analysis of the down-regulated St.11 IRGs and a proposed paradigmatic expression patterns.

(A) Heatmap presenting the row clustered down-regulated St.11 IRGs, data in column 1-3 are normalised two-colour array expression value from LCM-collected tapetum tissues at St.8, St.11 and St.12 respectively. The genes are clustered into Group I, II and III. (B) A zoomed display of IRGs in group I. (C) Proposed paradigms showing the expression pattern of group I, II and III. PMI, pollen mitosis I. PMII, pollen mitosis II.

Table 4.1 Selected genes in clustered groups of each of the *ms1* IRG sets.

IRG sets	Clustered groups	AGI code	Gene Name	Description	Reference
Up-regulated at St.8 and St.11	Group I	AT2G26910	ABCG32	ATP-BINDING CASSETTE G32	(Quilichini et al., 2014c)
		AT2G45970	CYP86A8	CYTOCHROME P450, FAMILY 86, SUBFAMILY A, POLYPEPTIDE 8	(Dobritsa et al., 2009b)
		AT3G59530	LAP3	LESS ADHERENT POLLEN 3	(Dobritsa et al., 2009a)
		AT4G35420	TKPR1	TETRAKETIDE ALPHA-PYRONE REDUCTASE 1	(Grienenberger et al., 2010)
		AT1G62940	ACOS5	ACYL-COA SYNTHETASE 5	(Lallemand et al., 2013)
		AT1G69500	CYP704B1	CYTOCHROME P450, FAMILY 704, SUBFAMILY B, POLYPEPTIDE 1	(Dobritsa et al., 2009b)
		AT3G11980	MS2	MALE STERILITY 2	(Chen et al., 2011)
	Group II	AT2G31210	bHLH91	BASIC HELIX LOOP HELIX PROTEIN 91	(Zhu et al., 2015a)
		AT2G42940	TEK	TRANSPOSABLE ELEMENT SILENCING VIA AT-HOOK	(Jia et al., 2015)
		AT5G22260	MS1	MALE STERILITY 1	(Yang et al., 2007)
	Group III	AT3G25730	EDF3	ETHYLENE RESPONSE DNA-BINDING FACTOR 3	
		AT4G31970	CYP82C2	CYTOCHROME P450, FAMILY 82, SUBFAMILY C, POLYPEPTIDE 2	
		AT2G25230	MYB100	MYB DOMAIN PROTEIN 100	
Down-regulated at St.8	Group II	AT3G16340	ABCG29	ATP-BINDING CASSETTE G29	(Alejandro et al., 2012)
		AT5G62320	MYB99	MYB DOMAIN PROTEIN 99	

IRG sets	Clustered groups	AGI code	Gene Name	Description	Reference
Down-regulated at St.8		AT1G06280	LBD2	LOB DOMAIN-CONTAINING PROTEIN 2	
		AT5G17050	UGT78D2	UDP-GLUCOSYL TRANSFERASE 78D2	(Yonekura-Sakakibara et al., 2014)
		AT5G66690	UGT72E2	UDP-GLUCOSYL TRANSFERASE 72E2	
		AT5G54010	UGT79B6	UDP-GLUCOSYL TRANSFERASE 79B6	
		AT3G23770		O−Glycosyl hydrolases family 17	
		AT4G26830			
		AT3G42950		Pectin lyase-like superfamily protein	(Jiang et al., 2014b, 2014a)
		AT4G22090			
		AT4G22080	RHS14		
		AT5G17200			
Down-regulated at St.11	Group I	AT3G10410	SCPL49	serine carboxypeptidase-like protein, PCD related	(Olvera-Carrillo et al., 2015; Hautegem et al., 2014)
		AT3G45010	SCPL48		
		AT4G14815		Bifunctional inhibitor/lipid-transfer protein	(Edstam and Edqvist, 2014)
		AT1G66850			
		AT5G64080			

4.3 DISCUSSION

The regulatory role of MS1 in late tapetum development has been demonstrated (Yang et al., 2007; Ito et al., 2007; Yu, 2015), however, the molecular function of MS1 remains poorly understood. In this chapter, a combination of approaches, including phylogeny study, protein sequence comparison, molecular modelling, subnuclear localisation observing and meta-microarray-data analysis, were used to preliminarily probe the MS1 molecular function and regulatory targets. A few possible directions for MS1 function are indicated by the results of this chapter.

4.3.1 MS1 belongs to an ancient Viridiplantae-specific CPCP family

The transcriptional regulation role of Arabidopsis CPCP in late tapetum development and the meiosis process is essential for successful male reproductive development progress. The identified green plants CPCPs and their phylogenetic relationship suggest that these genes probably share a common ancestor, but have diverged functions between land plants and green algae. Sporopollenin synthesis is one of the main tapetum functions that is regulated by CPCP in angiosperms. The biological processes of sporopollenin synthesis and meiosis process are observed in several taxa of primitive green algae that possess CPCPs (Good and Chapman, 1978), such as *Chlorella*. Therefore verifying the effect of the CPCP knockouts on these biological process in green algae might reveal the role of CPCPs in primitive plants.

According to the endosymbiotic theory, plants originate from ancient endosymbiosis events that brought chloroplasts into primitive eukaryotic cells. Three lineages have

evolved from the initial plant endosymbiosis, *i.e.* the Viridiplantae lineage, which includes the green algae (Chlorophyta) and land plants (Embryophyta), the red algae (Rhodophyta) lineage, and the blue lineage (Glaucocystophyta) (Botté and Maréchal, 2014). Molecular phylogenetic information indicates that land plants (Embryophyta) originated from a freshwater green alga (Sakakibara, 2016). The CPCPs only present in the green algae and their descendants the land plants, but not in the red and blue lineages, which suggest CPCP is an ancient protein obtained by the last common ancestor of all higher plants. However, the family size of the CPCPs remains small. Particularly, after multiple times of whole-genome duplication over the past 200 million years, the Arabidopsis and rice genome does not possess more than four CPCPs, which suggested that the function of CPCP is unique and specific. This scenario is probably caused by the fact that organisms are sensitive to the level of CPCPs, and lost the duplicated CPCPs after whole-genome duplication events.

4.3.2 MS1 possesses two MADs and a possible H3K4me2/3 binding PHD finger

The molecular function of CPCPs remains less understood until recently when Andreuzza et al. (2015) and Wang et al. (2016) demonstrated the H3K4me2/3 binding affinity of MMD1. Using *in-silico* comparative protein sequence analysis of the MMD1, MS1 and other H3K4me2/3 binding proteins, a Cys4-His-Cys3 conserved motif and additional key amino-acid residues that determine the histone peptides binding specificity were found in the MS1 PHD domain, which suggests that MS1 might share the same H3K4me2/3 binding affinity as MMD1. This speculation is also strengthened by the result of molecular modelling of the MS1 PHD domain using the crystal structure of the human MLL5-PHD-H3K4me3 complex as a template. It was

found that the surface groove recognition site of PHD-H3K4me3 interaction is highly conserved between MS1 and MLL5 in the modelled MS1 PHD-H3K4me3 hetero-oligomer, in which the trimethylated lysine of H3 interacts with an aromatic pocket formed by the 625D 629M 638W and 653P residue of the MS1 protein.

It is worth noting that MLL5 and MMD1 PHD could bind to several other types of histone peptides besides H3K4me2/3 *in vitro*, while the former is preferable to H3K3me3 binding while the later is preferable to H3K3me2 binding (Ali et al., 2013; Lemak et al., 2013; Wang et al., 2016). Although there are high conservations of the protein sequences among PHD domains of MLL5, MMD1 and MS1, the *in-vivo* molecular function the MS1 PHD remain to be illustrated by further investigations. The *in vitro* binding affinity of MS1 PHD could be possibly different from that *in vivo* as PHD proteins affinity to histone peptides can vary depending on its PPI factors.

Conserved amino acid signatures are always an indication of functions such as PPIs or regulatory sites. The evolutionarily conserved MAD1 and MAD2 identified at the middle part of CPCPs might play important roles for functional MS1 since PHD finger always need to cooperate with other domains to exert its transcription regulatory function. However, given that CPCP is a small and unique protein family, there is barely any reported evidence regarding the molecular roles of MAD1 and MAD2. Future work is needed to verify the potential role of these two newly identified domains for MS1 function.

4.3.3 MS1 proteins localise at subnuclear discrete speckles resembling PcG bodies

To better understand the molecular function of the MS1 protein, subnuclear distribution of MS1 was investigated and compared with the recorded image gallery of various plant NBs. Initial microscopic observation using anthers from Arabidopsis stably transformed with MS1pro:MS1-GFP showed that the tagged protein was specifically localised to the tapetal nuclei, showing a subnuclear distribution of punctate domains with a reticulate background. Later observation of transiently expressed MS1-GFP and TDR1-mTFP1 showed increased sensitivity and resolution of the imaging. The images showed that MS1 localised to sparsely distributed rounded speckles reminiscent of the discrete nuclear foci of LHP1, while TDR1 localised to two different subnuclear locations in various nuclei: densely distributed dots that fused with a continuous network and large irregularly shaped speckles. TDR1 is a rice bHLH TF that activates a vast number of downstream targets in early tapetum development (Ko et al., 2014; Fu et al., 2014), which suggest TDR1 should, in theory, localise to transcription factories. However, this speculation is consistent with the pattern A of TDR1-mTFP1 localisation, but not the pattern B. Notably, nuclei showing TDR1-mTFP1 pattern B exhibited an irregular shape, which might indicate these nuclei was going through some specific developmental phase that wrongly or necessarily allocates the TF activators into SFCs or SFC-like NBs. The size, shape and number of MS1-GFP labelled speckles within each nucleus varied significantly from those of TDR1-mTFP1 labelled speckles, suggesting that MS1 functions in a distinct way with the bHLH TF activator in regulating transcription. Additionally, the reticulate background GFP signals in the Arabidopsis tapetal nuclei were no longer

present in the nuclei of *N. benthamiana* leaf epidermal cells. This inconsistency is more likely due to the improvement of the imaging using *N. benthamiana* cells rather than the nuclear differences between species, as the boundary of the MS1-GFP labelled speckles in Arabidopsis is relatively clear, unlike the fused speckles labelled by TDR1-mTFP1. This evidence comprehensively demonstrated that MS1 proteins are actively transported or aggregated to specified NBs in nuclei, which are distinct to transcription factories that have a continuously dispersed reticulate structure. Therefore, the reticulate background obtained in Arabidopsis tapetal nuclei is highly possible to be the out of focus signal.

As there are extensive reports of recruitments of PRC to genomic loci by PHD proteins (Hennig and Derkacheva, 2009; De Lucia et al., 2008; Mozgova et al., 2015), and also the Arabidopsis LHP1 is globally colocalized with the histone mark H3K27me3 (Derkacheva et al., 2013), which spreads over the second exon of MS1 genomic loci that MS1 protein binds to, it is plausible to speculate that MS1 might be involved in PRC-mediated histone H3K27me3 modification. Additionally, the analogy between MS1 and LHP1 subnuclear localisation provide a suggestion that MS1 might cooperatively function with LHP1 in PcG bodies. Although PcG proteins mainly regulate their targets by silencing them, they have also been shown to mediate gene activation in specific cases (Entrevan et al., 2016). Therefore, it is not surprising if MS1 can activate a massive number of downstream targets.

4.3.4 A significant portion of MS1 IRGs are up-regulated in the *ms1* mutant.

Time-series and cell type-specific transcriptome profiling provides valuable data sets for deriving regulatory interactions of TF and prediction of its direct targets (Swift and Coruzzi, 2017). Several sets of microarray transcriptomic data from anthers were collected and comprehensively analysed. By comparison of *MS1* two-fold DEGs with those of *MYB80* and *TEK*, it was found that a considerable portion of *MS1* DEGs are not a direct result of tapetum regulatory pathway perturbation in the *ms1* mutant, but a common consequence in other anther mutants that have aborted late tapetum development and eventually collapsed pollen grains. To avoid the interference of these indirect targets, MS1 IRGs, which temporally and spatially approximate the expression pattern of MS1, were retrieved by filtering the DEGs for genes that are highly expressed in tapetum at the St.8 and St.11. Surprisingly, the number of up-regulated IRGs is much more than that of the down-regulated IRGs, which is contrary to the situation of the DEGs.

A few non-informative genes whose expression level exceeded the dynamic range of microarray, which was ruled out for further analysis. Clustering and heat map analysis of the IRGs using time-series transcriptome profiling datasets categorised the up-regulated IRGs into two types. The first type of IRGs only consists of a few genes, which exhibited a significantly high expression in *ms1* mutant but have nearly no expression in WT anthers. These IRGs are mostly non-anther genes, which, therefore, might represent the stress-induced genes responding to the disorder within the *ms1* tapetal cells. Alternatively, it is possible that these IRGs are wrongly expressed due to a failure of MS1-mediated chromatin structure change in *ms1* tapetal cells. The

second type of IRGs showed elevated level and prolonged duration of gene expression in the ms1 mutant compared to the WT. These IRGs are mostly comprised of tapetum functional genes, whose expression seems to be switched-off by the MS1 protein when tapetum development enters the post-meiosis phase.

The regulation role of MS1 was always presumed as a transcription activator (Yang et al., 2007; Ito et al., 2007; Alves-Ferreira et al., 2007). The evidence supporting this includes massive down-regulation of anther-specific genes in the *ms1* mutant compared to WT (Yang et al., 2007) and the *ms1-like* phenotype conferred by the fusion protein of MS1 and a transcriptional repression domain (MS1-SRDX) (Ito et al., 2007). However, the meta-analysis of anther microarrays in this chapter has demonstrated that a considerable proportion of the MS1 down-regulated DEGs are probably not regulatory targets of MS1. Moreover, the fact that up-regulated genes predominated in the curated MS1 IRGs provide an indication that MS1 plays a role in repressing transcription expression. Fusion of the SRDX domain can convert a TF activator into a dominant repressor (Mitsuda and Ohme-Takagi, 2009). However, the underlying molecular mechanism was not clear until recently when Ke et al. (2015) demonstrated that the SRDX-mediated repression depends on the interaction between the SRDX motif and TOPLESS-related (TPR) proteins, which belong to a plant transcriptional corepressor family. Therefore, it can be deduced that the recruitment of TPRs to the MS1 binding genomic loci by MS1-SRDX would interrupt the function of MS1, probably via disturbing the PPI of MS1 partner proteins, which can cause MS1 loss-of-function phenotype. However, interruption of the protein function by artificial PPI of unrelated corepressors cannot necessarily demonstrate the transcription-activating function of MS1. From the above discussed, and given MS1

lacks DNA-binding domain, it is not appropriate to arbitrarily treat MS1 as an activator TF.

Among the down-regulated IRGs, quite a few genes with interesting functions related to the post-meiosis tapetum activity were found, providing promising candidates for the MS1 regulatory target. For instance, several classes of PCD linked proteins, such as cysteine proteinases, serine carboxypeptidases and aspartyl protease, show dramatic down-regulation in *ms1* mutant compared with the wild-type buds (Zhang et al., 2014; Olvera-Carrillo et al., 2015), however, which are initially responsive to MS1 regulation is not known. The *SCPL48/49*, which encodes serine carboxypeptidases, belongs to MS1 IRGs and are possible to be a direct target of MS1, which provides an indication of how MS1 may be regulating the tapetal PCD progress.

4.3.5 Summary and future work

After a series of preliminary studies described in this chapter, a few basic features were revealed regarding the MS1 function. Firstly, MS1 belongs to a green plant specific CPCP family that originates from green algae. Secondly, the MS1 PHD finger has a similar H3K4me2/3 binding affinity with the PHD of human MLL5 and the MS1 Arabidopsis paralog gene MMD1. Additionally, two conserved domains, MAD1 and MAD2, locate in the middle part of the MS1 protein, which might be necessary for the MS1 protein function. Thirdly, the MS1 protein localised to PcG body-like NBs, which are distinct from the bHLH TFs. Lastly, MS1 appears to act more like a transcription repressor or corepressor rather than an activator.

Based on these preliminary findings, it would be interesting to follow up these research directions. Firstly, by verifying the function of CPCP in the most primitive land plant that is extant, such as moss, would indicate how CPCP was adopted by the land plant to regulate tapetum development. Secondly, confirming the requirement of MAD1/2 for MS1 function and characterisation of the *in-vivo* H3K4me2/3 binding affinity of MS1 PHD would better define the molecular function of MS1. Thirdly, confocal colocalization analysis of PcG marker proteins with MS1 would verify if MS1 is a PcG body component. Also, mass spectrometry-based identification of immunoprecipitation of MS1 PPI partner would provide accurate information on whether MS1 cooperates with PRC proteins. Lastly, an induction experiment using the DEX-inducible MS1 line to examine the response time of the expression perturbation of genes in **Table 4.1** after the MS1 nuclear localisation would more precisely define whether MS1 has a repressing or activating role in regulating the transcription of late tapetum gene.

CHAPTER 5 CHARACTERISATION OF MYB99, A PRESUMED MS1 DIRECT TARGET

5.1 INTRODUCTION

The MYB superfamily is one of the largest and most diverse classes of TFs (Charoensawan et al., 2010). MYB TFs are widely distributed in all eukaryotic organisms and have remarkable expansion in the plant kingdom (Stracke et al., 2001). For instance, the Arabidopsis genome contains 339 MYB genes, which comprised a large propotion of Arabidopsis TFs (Pireyre and Burow, 2015). All MYB TFs are characterised by the MYB domian, which is comprised of approximately 50 amino acids, with up to 4 repeats (R) in tandem (Lipsick, 1996). This domain can form a helix-turn-helix structure in which three tryptophan residues are evenly sapced to form a hydrophobic core, which is responsible for the DNA sequence recognition and binding (Kelemen et al., 2015).

In plants, more and more Arabidopsis MYBs have been studied and characterised by genetic approaches over the last 15 years. They are found to be involved in the control of plant-specific processes, such as cell fate and identity, developmental processes, responses to biotic and abiotic stresses and primary and secondary metabolism (Dubos et al., 2010). In the tapetum, several MYB genes have been characterised and found to be essential for normal development, including *TDF1/MYB35, MYB80, GAMYBs*. The detailed function and regulatory targets of these tapetal MYBs are reviewed in **Section 1.2.2**.

MYB99 was first identified as having a role related to tapetum development by Yang et al. (2007), since it is one of the TFs that showed the most significant downregulation in *ms1* bud compared with the WT. The regulation of *MYB99* by *MS1* was further confirmed by Ito et al. (2007), who generated a dexamethasone (DEX) inducible MS1 line *MS1pro:MS1-GR/ms1-1* and showed that MYB99 is the most rapidly DEX induced gene among six selected targets (**Figure 5.1**), and that this induction does not require *de novo* protein synthesis. Based on this result, Ito et al. (2007) suggested that *MYB99* may be a direct target for *MS1*.

Alves-Ferreira et al. (2007) confirmed that *MYB99* expression was confined to tapetal cells and Stages 8 and 9 of anther development using *in situ* hybridization experiments. They also characterised the T-DNA insertion line for *MYB99* (SALK_003193), and claimed the mutant plant had a reduced fertility phenotype with smaller siliques, less viable seeds and thiner tapetum layer compared with the WT. However, no statistical analysis of the fertility change was shown and the extent of the reported *MYB99* phenotype is ambigous; additionally, the direct regulation of *MYB99* by *MS1* lacks molecular biological evidence.

In summary, MYB99 expression coordinates with the beginning of late tapetum phase development, therefore understanding the function of *MYB99* and its regulation by MS1 can potentially help understanding of the late tapetum regulation network. In this Chapter, a series of precise phenotyping approaches were carried out on two *MYB99* mutant alleles, which indicated that a PCD indicator gene is down-regulated in the mutants. Yeast one-hybrid and ChIP experiments were conducted to verify the direction regulation of *MYB99* by *MS1*, however, the results suggest that *MYB99* is an indirect target of MS1.

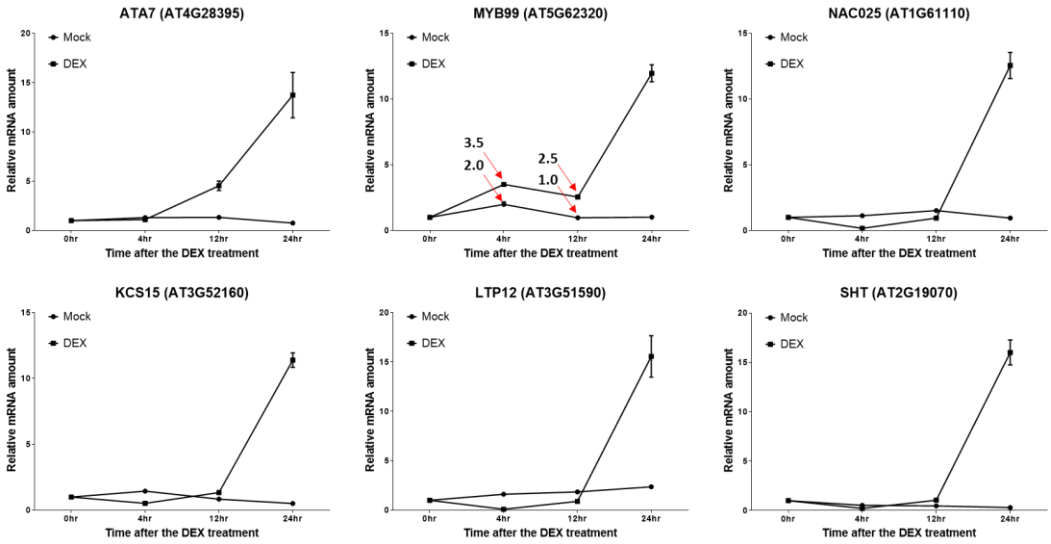


Figure 5.1 Time-course induction analysis of the six MS1 downstream targets using DEX-inducible *MS1pro:MS1-GR/ms1-1* transgenic plant.

Expression data was taken from supplementary data of Ito et al. (2007), normalised with the expression level at 0hr time point, which was set to 1 in all six genes and reploted using Prism 7. The y axis showed relative folder change of the expression compared to the 0hr time point (untreated).

5.2 RESULTS

In this chapter, the functional study of MYB99 was carried out by two approaches, one by a precise phenotyping analysis of two T-DNA knock out alleles of MYB99. The other is via analysis of the regulatory relationship between MS1 and MYB99 using both *in vivo* and *in vitro* approaches.

5.2.1 T-DNA insertion mutants of MYB99 lack phenotype

Two T-DNA knock out lines (SALK_003193C and SALK_083183C) of MYB99 were obtained from NASC (Alonso, 2003). PCR genotyping was carried out to screen for homozygous lines of T-DNA insertions. Subsequent RT-PCR verified a complete knockout of the MYB99 cDNA in both of the two lines, which suggests SALK_003193C and SALK_083183C are null alleles of MYB99, hence termed as *myb99-1* and *myb99-2* respectively.

Both of these homozygous mutant lines and the WT were sown and grown in a routine growth condition. However, the mutants showed no apparent phenotypes compared with the WT. To detect if the phenotype of *myb99* is condition dependent, the *myb99-1* mutant line was also grown with the WT in a series of suboptimal conditions (details are summarised in **Table 5.1**). However, the *myb99-1* plants grown under all these conditions showed no difference in plant appearance and fertility compared with the WT plants. To verify whether the male fertility was impaired in *myb99* mutants, the anthers of both the mutant and WT plants were collected to probe pollen viability and anther lignin synthesis using Alexander staining and ethisium

bromide/acridine orange staining. However, no difference was obtained between the mutant and the WT (data not shown).

Table 5.1 The growth conditions for the detection of *myb99* conditional phenotypes.

Suboptimal condition	Day temperature (℃)	Night temperature (℃)	Photoperiod (hour)	Light intensity ($\mu mol \cdot m^{-2} \cdot s^{-1}$)
Low light	23	18	16	37.8 ± 2.4
Low temperature	15	13	16	87.3 ± 1.6
Short day	18.5	15	12	124.6 ± 2.9
Normal condition	23	18	16	131.6 ± 4.1

5.2.2 Precise phenotyping of the *myb99* mutant using SEM

To assay the phenotype of *myb99* mutant more precisely, anthers at the tetrads release, PMI and anthesis stages were collected from the WT, *myb99-1* mutant, *ms1-1* mutant for a SEM assay. As MYB99 functions downstream of MS1, a subset of the features of the *ms1* mutant phenotype would be expected to appear in the *myb99* mutant. Consistent with previous reports, the tetrads of *ms1* mutant plants were found to have a delayed release (**Figure 5.2 A**) and the released microspores were sticky and microspores from the same tetrad appeared to be still attached to each other (**Figure 5.2 B**). Close inspection of the surface of newly released microspores found traces of substances associated with putative tetrad attachment sites, which might be the residues of callose that had not been fully degraded (**Figure 5.2 C**). This suggests that the tetrad callose degradation can be accomplished in the *ms1* mutant, but that the progress was significantly delayed. On the other hand, the tetrad release process was completely normal in the WT and *myb99* mutant. The *ms1* anther at the PMI stage had

a severe aberration under the SEM, showing dramatically collapsed and shrunken microspores/pollen grains (**Figure 5.2 D**). On the contrary, anthers from the *myb99* mutant and the WT at this stage appeared normal (**Figure 5.2 E** and **F**). The *ms1* anther at the anthesis stage lacked autonomous dehiscence. A SEM observation of the manually dissected *ms1* anther this stage showed a layer of endothecium cells that remained appeared still intact (**Figure 5.2 G**), while the dehiscent anther of the WT and *myb99* mutant showed a layer of completely dehydrated endothecium which had a knitted fabric-like look (**Figure 5.2 H** and **I**). In summary, although the SEM observation identified a series of previously uncharacterised phenotypes of the *ms1* anther, which are consistent with the reported role of MS1, no difference was obtained in these aspects between the *myb99* anther and that of WT.

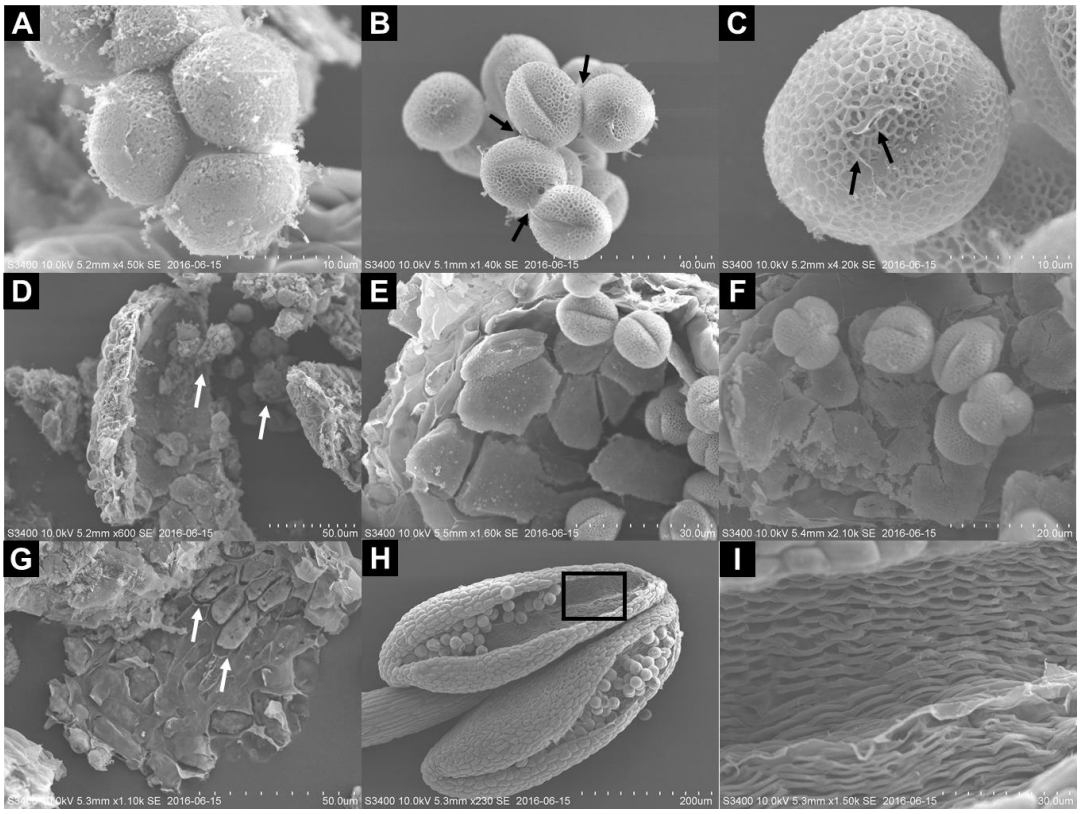


Figure 5.2 The SEM observation of the anthers of the WT Arabidopsis and the *ms1* and *myb99* mutant.

(A-E) The tetrads (A) and the newly released microspores from the *ms1* mutant anther. (B) The *ms1* microspores are attached to each other. (C) The trace remnants of possibly callose wall on the surface of an *ms1* microspore, indicated by the black arrows. (E-F) The dissected anthers at PMI stage of the *ms1* mutant (D), the *myb99* mutant (E), and the WT (F), showing the pollen grains and the locule inner cells. (G) The inner cell layer of a dissected *ms1* anther at the anthesis stage. The intact endothecium cells are indicted by the white arrows. (H) An *myb99* anther at anthesis stage showing dehiscent locules. (I) A zoomed view of the region of rectangle framed region in (H), showing the dehydrated endothecium cells.

5.2.3 Transcriptomic profiling of the *myb99* anther

After the precise phenotyping approaches conducted, the *myb99* showed no alteration in the aspects of cell wall patterning, tetrad release, tapetal cell morphology, endothecium dehydration and pollen viability, even at the micrometre scale. Considering the dramatically large size of the MYB family in the Arabidopsis genome

(Dubos et al., 2010), redundancy amongst MYB genes may explain the lack of obvious phenotype in the *myb99* mutant. The phylogenetic trees of Arabidopsis MYB family constructed by Kelemen et al. (2015), Dubos et al. (2010) and Wilkins et al. (2009) were checked, and the *MYB20*, *MYB40*, *MYB42*, *MYB43* and *MYB85* genes were found to be located within the same phylogeny clade as *MYB99*; these were thought to be the most possible genes to act redundantly alongside the MYB99 gene. However, after checking the tapetum expression level of these genes using the anther LCM two-colour-array data (**Section 4.1.4**), barely any expression was detected in the tapetum or even anthers (data not shown), suggesting they are extremely unlikely to be acting redundantly with MYB99 in pollen development.

It is hard to generate any other genetic studies of *MYB99* with the mutant alleles showing no phenotype. However, without knowing its redundant gene, no promising double mutant could be made. Considering that *MYB99* is a highly-expressed TF that belongs to MS1 IRGs. The fully knock out of *MYB99* probably could cause an expression perturbation of other *MS1* IRGs, which can be detected by qPCR. Four representive genes, *i.e. SCPL48*, *LBD2*, *UGT78D2* and *RHS14* from the MS1 IRGs (**Table 4.1**) were selected for qRT-PCR analysis as these genes each directs a different aspect of the late tapetum functions. Interestingly, the SCPL48, which is a PCD indicator gene, showed a slight but statistically significant expression reduction in the *myb99-1* mutant, while the other three genes showed unaltered expression compared with the WT (**Figure 5.3**).

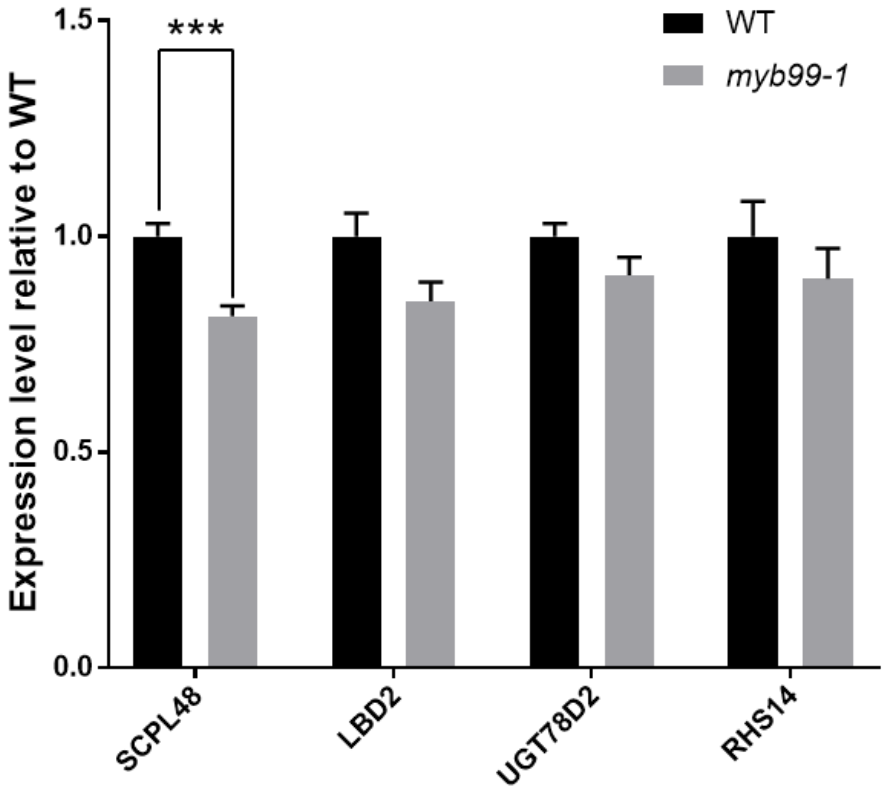


Figure 5.3 The quantitative RT-PCR of four MS1 immediately responding genes in the anther of the *myb99-1* mutant and the WT.

All expression levels are normalised to that of the reference gene *YLS8*, and the expression level of each gene in the WT anther was converted to 1. Error bars represent the standard error of the calculated mean based on three biological replicates, each biological replicate consist of three technical replicates. The statistical significance is a result of t-test between the WT and the *myb99-1* mutant, and the p-value is indicated by *** when $p<0.001$.

5.2.4 Yeast one-hybrid assay of the binding of tapetum TFs to the MYB99 locus.

Besides the functional study of the *myb99* mutant, the regulation of MYB99 expression in tapetum was also studied. Similar to the yeast one-hybrid assay conducted **in Section 3.2.3**, the −714 to +182 bp region of MYB99 genomic locus was PCR amplified and constructed as three ~450 bp promoter sections, which were integrated into yeast strain Y187. Each of the bait sector had a 60 bp overlap with its neighbouring sectors on both 5' and 3' side in case important TFBSs were interrupted at the borders. For the putative MYB99 TF regulators to screen against, the full length CDS of *MS1*, and other tapetum TFs, such as *AMS, DYT1, BES1* and *BZR1* were integrated into yeast strain AH109. The mating of yeast strain Y187 and AH109 was verified on a Petri dish plate of synthetic defined medium lacking tryptophan and leucine (SD-2) (**Figure 5.4 A**). However, no interaction occurred in yeast between the tested baits and preys, with only the positive control colonies growing on the selection Petri dish plate of SD-3 (SD medium lacking tryptophan, leucine and histidine) (**Figure 5.4 B**).

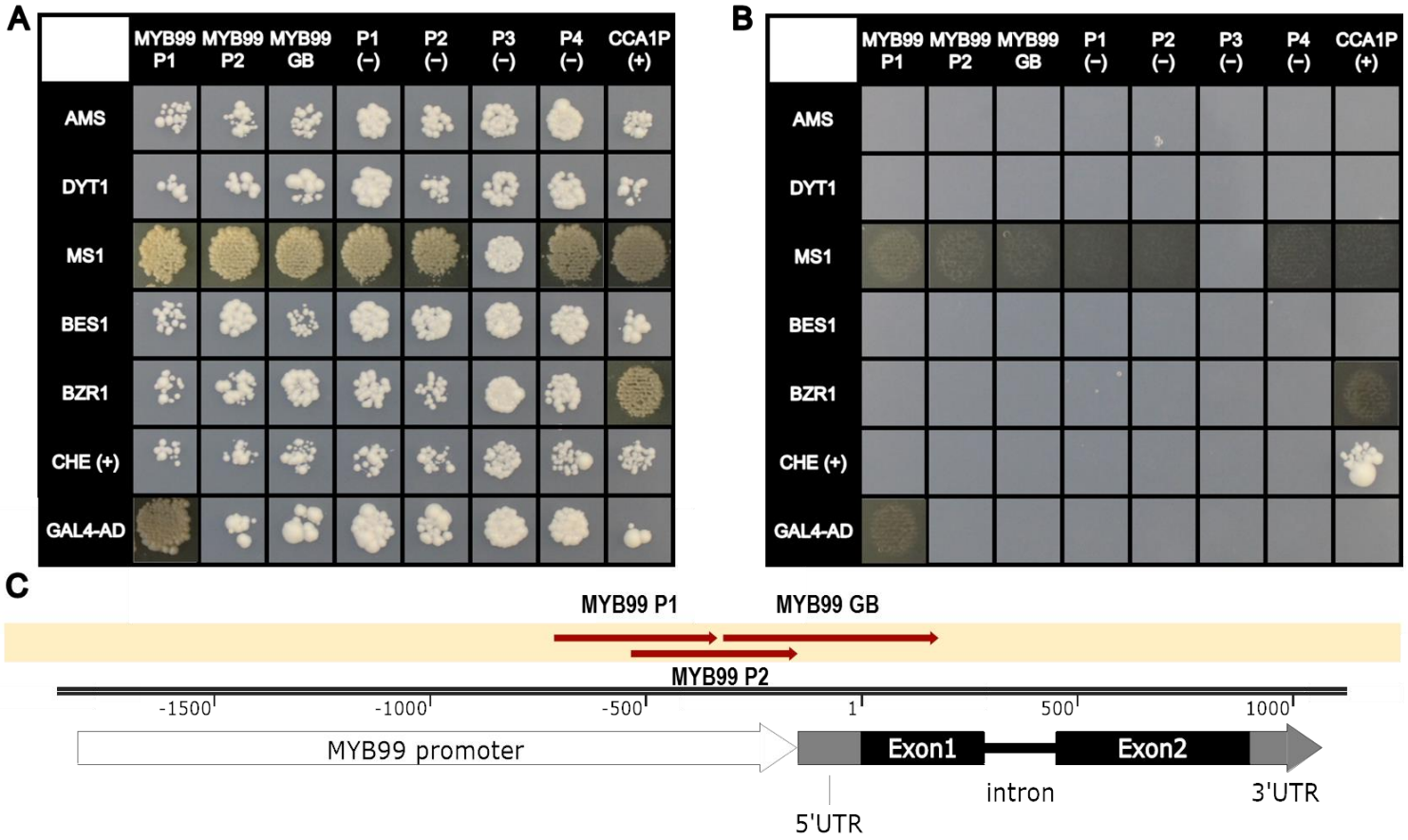


Figure 5.4 The yeast one hybrid indicated no interaction persists between selected tapetal TFs and the *MYB99* genomic sequence.

(A) The mating colonies of yeast strain Y187 possessing *MYB99* genomic DNA baits and AH109 possessing the selected tapetum TF preys on SD-2 media. (B) The interaction assay of the mating colonies in (A) on SD-3, 5mM 3-AT media. (C) A schematic diagram showing partitioned regions of the *MYB99* genomic sequence that used as bait DNA in the yeast-one-hybrid. Yeast colonies in rows from top to bottom contain AMS, DYT1, MS1, BES1, BZR1, CHE (+) and GAL4-AD (−) respectively; colonies in column from left to right contain MYB99_Prom1 (379bp), MYB99_Prom2 (385bp), MYB99 Gene Body (500bp), AMS_Prom1 (396bp), AMS_Prom2 (397bp), AMS_Prom3 (405bp), AMS_Prom4 (382bp) and CCA1_Prom (+) (217bp) respectively.

5.2.5 ChIP experiment verifying if MS1 directly binds the MYB99 genomic locus

To verify if MS1 directly regulates the expression of MYB99 a ChIP experiment was conducted on the genomic locus of *MYB99* in an analogous manner as the ChIP experiment described in **Section 3.2.4**. Specifically, an anti-GFP antibody and the inflorescence tissues of a stable *MS1-GFP* transgenic line were used. Considering the MS1 protein could potentially bind to the gene body of its regulatory target, not only the promoter region of *MYB99* but also the gene body region was used for the MS1 ChIP DNA enrichment assay. Three different fixation conditions, 1% (w/v) formaldehyde fixation in MC buffer for 15 minutes, 20 minutes and in deionized water for 20 minutes, were used in three independent ChIP experiments. The *TA3* locus was used as an internal negative control (enrichment) for ChIP enrichment and the Ler WT tissue ChIP results were used as the external negative controls. Surprisingly, all three ChIP batches did not detect any MS1 enrichment/binding to the *MYB99* genomic locus, as all the sites detected at the *MYB99* locus had lower or similar level of ChIP enrichment compared with that of the *TA3* negative control (**Figure 5.5**). Although, this result is inconsistent with the ChIP result of Yu (PhD thesis, 2015) obtained, given the comprehensive controls and quality control steps introduced in this ChIP experiment, and the consistency among the three biological replicates, it is proposed that MS1 may not directly regulate the expression of *MYB99*. The systematic enrichment bias towards the MS1-GFP group, which identified in the MS1 auto-repression ChIP, still persist here, representing the possible interference of free GFP proteins in the ChIP system. Using other epitope tags besides the GFP tag might improve this problem in further ChIP work.

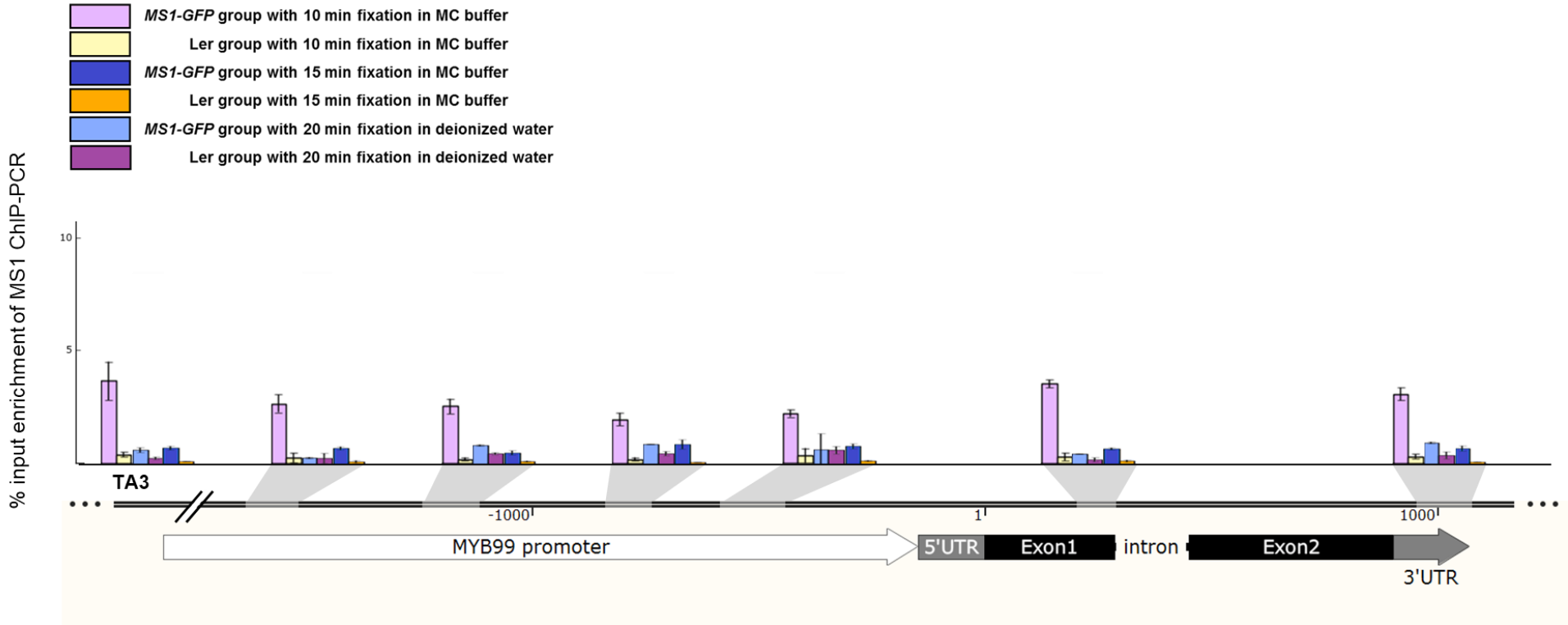


Figure 5.5 No MS1 ChIP enrichment was detected in the *MYB99* genomic locus.

Three independent batches of the ChIP results are plotted in the bar chat diagram based on the genomic site of their PCR amplicon. The ChIP enrichments for each site are calculated as the percentage of input and compared with those of a negative control site, a retrotransposon gene TA3. No significant differences between the control and the tested site are detected (all p value obtained are more than 0.05). Error bar on each bar represents the 95% confidence interval of the enrichment.

5.3 DISCUSSION

5.3.1 MYB99 may initiate the late tapetum PCD progress

MYB99 is highly expressed in the post-meiotic phase tapetum and rapidly induced by MS1, which implies it may have a vital role in the late tapetum development. However, the exact phenotype of the *myb99* mutant is poorly characterised. A series of precise phenotyping assays were conducted on two *myb99* T-DNA null alleles, however, not difference was found between the mutant and the WT in pollen viability, anther lignin accumulation, tapetal cell morphology, tetrads release and pollen wall patterning. The mutant when grown in sub-optimal conditions also did not show any conditional phenotype. Considering the enormous size of Arabidopsis MYB family, the lack of phenotype of the *myb99* mutant might be attributed to gene redundancy. By checking the phylogeny relationship and tapetum transcriptomic data of several *MYB99* close paralogs, the possibility that these genes act redundantly with MYB99 was ruled out, which indicates some evolutionarily distinct MYBs or non-MYB TFs may function as *MYB99* redundant genes. It became exceptionally difficult to characterise the function of *MYB99* without an observable phenotype in its loss-of-function mutant. However, considering *MYB99* is a high-expressed and tapetum-specific TF gene, the fully knocking out of *MYB99* might cause expression perturbations of the downstream targets even the its redundant gene is properly functioning. Interestingly, *SCPL48*, one of the four tested genes, indeed showed downregulation in the *myb99-1* mutant compared to the WT. Olvera-Carrillo et al. (2015) demonstrated *SCPL48* is a developmental PCD indicator gene, which shows

conserved expression in a series of plant cells that are undergoing differentiation-induced PCD, including tapetal cells. The late tapetum PCD progression is significantly delayed in the *ms1* tapetum, where the *SCPL48* expression is completely absent (Yang et al., 2007). Hence, it is possible that MS1 may regulate tapetal PCD progress via regulating the expression of *MYB99* and its redundant gene, which play a role in tapetal PCD initiation.

5.3.2 *MYB99* is an indirect downstream target of *MS1*

MYB99 has been long presumed as a direct regulatory target of *MS1* (Ito et al., 2007; Alves-Ferreira et al., 2007), however, no direct molecular evidence has been uncovered to fully support this assumption. The *in vitro* yeast one-hybrid assay and *in vivo* ChIP experiment conducted here clearly showed that there was no MS1 protein binding to the *MYB99* genomic locus, which therefore indicates that *MYB99* is an indirect downstream target of *MS1*. Although this conclusion is on the contrary to that presented by Ito et al. (2007), their results could be interpreted differently.

Firstly, Ito et al. (2007) claimed *MYB99* to be a direct target of MS1 based upon the quick induction of *MYB99* expression in a *MS1-GR* inducible line after DEX treatment. Nevertheless, after a close inspection of the original data of Ito et al. (2007), it is evident that the so-called "quick induction" of MYB99 was not that 'quick', but a relatively slow induction occurred 24 hours after the treatment. At the 4-hour and 12-hour time points after treatment, the DEX induction levels of *MYB99* were approximately twice that of the mock induction level, which remained at a basic background expression level as low as the 0-hour time point (untreated). And the 12-hour point even had a lower expression level compared to that at the 4-hour time

point (**Figure 5.1**). Apparently, twice of a very low value is still low, therefore it is over-fit to suggest that *MYB99* was the most rapidly induced target. However only at the 24-hour time point after the treatment, was the expression of *MYB99* was significantly induced in the DEX treated group, with levels reaching as high as eleven times of the mock group. Actually, the rest five genes that Ito et al. (2007) selected for the time-series assay of the DEX induction all showed a very similar induction pattern (**Figure 5.1**), which suggest these genes represent a later group of 24-hour responding targets downstream of *MS1*. The TF-bound target can reach its expression peak in 4 hours after the transcriptional activation by induced TF binding event (Swift and Coruzzi, 2017). Additionally, considering the relatively short life span of MS1 protein in tapetal cell nuclei, these 24-hour responding genes are less likely to be the direct targets of MS1.

Secondly, Ito et al. (2007) combined the DEX induction experiment with the addition of protein synthesis inhibitor cycloheximide (CHX) to demonstrate that the induction of *MYB99* by the DEX-inducied nuclear localization of MS1 protein does not need denovo protein synthesis, which further demostrated that *MYB99* is the direct target of *MS1*. The conclusion Ito et al. (2007) made could be true if MS1 is a cononical TF as this method is widely adopted to assay direct targets of TFs. However, there are exceptions if the assayed protein is a co-transcription factor. Given that MS1 is not demonstrated to have DNA-binding affinity, and possesses a histone reader domain, MS1 is more likely to regulate the transcription via modifying the struction of chromatin. The DEX-induced nuclear localization of MS1 protein might lead to a significant chromatin structure change, which could trigger altered transcription of *MYB99* without the aid of *de novo* synthesised proteins.

In summary, although the *MYB99* has been long presumed to be directly activated by MS1 protein, the *in vivo* and *in vitro* assay conducted here suggest that *MYB99* may in fact be an indirect regulatory target of *MS1*. The newly proposed co-repressor role of MS1 does not conflict with the previously reported results.

5.3.3 Summary and future work

A comprehensive phenotyping approach failed to identify any significant phenotypes of the *myb99* mutant. However, a qRT-PCR assay identified a statistically significant expression change of a developmental PCD indicator gene, *SCPL48*, in the mutant, though the change extent was very subtle. As the it is relatively hard to identify the genes that works redundantly with *MYB99*, a transgenic line possessing a chimeric repressor fusion MYB99-SRDX, or a chimeric activator fusion MYB99-VP16, may be generated in the future work to make dominant mutant of *MYB99*. In these mutants, the regulation of tapetal PCD by MYB99 may be further studied.

CHAPTER 6 GENERAL DISCUSSION AND FUTURE WORK

Most land plants have a layer of secretory cells developed on the innermost side of locules within the anther (tapetum cells), which play an essential role in nourishing the microspores and in synthesising the pollen wall and pollen coat components (Liu and Fan, 2013). The Arabidopsis *MS1* gene is a PHD finger-containing regulator, which is essential to the middle-late tapetum phase transition and viable pollen formation (reviewed in **Section 1.3.1**). Understanding the genetic control, the molecular function and the regulatory targets of MS1 is important to help reveal the regulation network for late tapetum development. This study has investigated the transcriptional and translational regulation of *MS1* (**Chapter 3**), the MS1 protein properties (**Chapter 4**) and the regulatory interaction between MS1 and a hypothesised target of MS1, the transcription factor *MYB99* (**Chapter 5**). This work has shown that the expression of *MS1* is under multilevel regulation. A model is proposed in which MS1 is (i) activated via interactions between a series of tapetum TFs and the evolutionarily conserved CREs on the *MS1* promoter, (ii) auto repressed in the tapetum via the interaction between the MS1 protein and part of the genomic region of *MS1*, (iii) repressed in seedlings by H3K27me3 mediated polycomb repression, possibly via PRC and BPC recruitment to the GAGA-repeats of the MS1 locus, and (iv) post-translational proteolysed possibly mediated by LRBs (**Figure 3.18**). Additionally, this work found several interesting properties of MS1: it belongs to an ancient Viridiplantae-specific CPCP family, possessing two novel MADs domains and a possible H3K4me2/3 binding PHD finger; its functional FP-translational fusion exhibited an novel subnuclear localization that is distinct from

that of an bHLH TF; and it represses most of its immediately responding genes (IRGs), which is inconsistent with the previously hypothesed role of MS1 as an activator. Lastly, this work demonstrated that *MYB99* is not a direct regulatory target of *MS1*. Althought MYB99 has a high level of tapetum-specific expression, its knock-out muant showed no fertility alteration, possibly due to the gene redundancy.

6.1 MS1 HAS A BRASSICACEAE FAMILY-SPECIFIC MULTI-LEVEL REGULATION

Previous work demonstrated that the timing and duration of the expression of MS1 are critical for its function which implied that a tight regulation of *MS1* expression exists in the Arabidopsis tapetum (Vizcay-Barrena and Wilson, 2006; Ariizumi et al., 2005; Wilson et al., 2001). Over the past decade, a number of studies have shown that the expression of MS1 is possibly under multi-levels of regulation (Yang et al., 2007; Ito et al., 2007; Alves-Ferreira et al., 2007; Ye et al., 2010). The results presented in this thesis provide direct evidence supporting this hypothesis and have addressed a few key basic questions regarding MS1 regulatory mechanisms.

Previous studies using a combination of *in-vivo* and *in-vitro* studies have demonstrated that a series of tapetum-specific TFs, including *AMS*, *MYB80*, *DYT1* and *bHLH010*/*089*/*091*, and a BR-effector *BES1*, play a role in the activation of *MS1* transcription (Xiong et al., 2016; Cui et al., 2016; Ye et al., 2010). However, these TF regulators are expressed in the tapetum successively in the early-middle phase of tapetum development, whilst MS1 is not expressed until the end of the middle phase. Why the expression of MS1 is rapidly switched on at the tetrad release stage rather

than continuously activated throughout tapetum development is unknown. Additionally, the exact binding sites of these TFs or TF complexes in the upstream sequence have not been identified. Phylogenetic footprinting analysis in this thesis and DNA motif-based analyses have been used to determine the regulation of MS1 by dissecting the MS1 regulatory region, revealing that MS1 has extraordinarily large size of evolutionarily conserved cis-regulatory element modules, which possess a combination of the transcription factor binding sites (TFBSs) of the reported MS1 activators. As the regulatory information is often hardwired in the gene regulatory regions, these finding indicate that MS1 is under tight, complex and evolutionarily conserved transcriptional regulation, and that each motif may be responsible for cooperatively executing a part of the overall regulation of MS1. This may also explain why MS1 is not transcriptional activated until a relatively late stage, when all those MS1 activators are present or have been present in the tapetal cell.

Nevertheless, a yeast-one hybrid assay failed to confirm the interaction of the MS1 activators and the TFBSs identified upstream of MS1, possibly due to the absence of co-factors of the plant TF from the yeast. Interestingly, by using an *in-vivo* luciferase assay, Cui et al. (2016) have demonstrated that DYT1 can bind to a fragment of the MS1 promoter sequence, which possessed the DYT1 binding site identified in this thesis. Hence further experiments using this *in-vivo* luciferase assay will be required to determine the interactions between the TFBSs identified and the MS1 activators. The conservation of MS1 upstream non-coding sequence was shown to only exist within the Brassicaceae family. The upstream CNS of Poaceae MS1 orthologs shows a different length and pattern of conservation (**Figure 3.4**). This is consistent with the observation of Reineke et al. (2011) that clear different evolutionary rates of the gene upstream regions are seen between the monocot and eudicot groups. The results

presented here suggest that the Brassicaceae family evolved family-specific regulation of MS1, although the protein function remained the same between the monocot and eudicot group. The biological significance of this family-specific regulation needs further investigation.

A second level of regulation is observed in that the *ms1* transcript is upregulated and prolonged in the homozygous *ms1* mutant, while in the heterozygous *MS1/ms1* line, it is restored to a normal level, which suggests that MS1 is auto-repressed (Yang et al., 2007). Additionally, the transcriptional reporter of *MS1* showed longer expression compared to that of the translational reporter of *MS1*, indicating that *MS1* auto-repression requires not only the MS1 protein but also the presence of its gene body sequence (Yang et al., 2007). The ChIP result presented here confirms the direct binding of the MS1 protein to its gene body region, which strongly suggests direct auto-repression regulation by MS1. Considering the potential histone reader role of the MS1 C-terminal PHD finger (Mouriz et al., 2015; Li and Li, 2012), the recruitments of MS1 protein to its genomic locus might be via the interaction between the PHD and covalent histone tail modifications. On the other hand, the high MS1 ChIP enrichment in the second exon of MS1 and its flanking introns indicates the requirement of this DNA region for the MS1 auto-repression. This requirement may also explain why the MS1 transcriptional reporter without the *MS1* gene body lost the auto-repression, and why the *ms1-1*/*ms1-6* SNP, which locates at the immediate 3' end of the second exon, significantly affects MS1 auto-repression efficiency. Interestingly, an anti-lncRNA *AT5TU027070* was confirmed in this work as transcribed at this MS1 ChIP most enriched region (**Figure 3.16**). Whether the effective recruitment of MS1 at this region is dependent on the local DNA sequence, the chromatin histone covalent

modifications or the cis-regulatory function of the anti-lncRNA need to be further explored.

The third level of regulation is evident by the exceptionally transient lifespan of the MS1-GFP protein in the tapetum and the fact that MS1 interacts with an E3 ubiquitin ligase LRB1 *in-vitro* (Yu, 2015); it was presumed that the rapid removal of MS1 is via ubiquitin-mediated post-translational proteolysis. Whether the protein to protein interaction between MS1 and LRB1 exists *in-vivo* and plays a biological role remains uncertain, as the *lrb1* mutant lacks a phenotype in respect to tapetum development and male fertility. However in the work presented here the triple mutant of the *LRB1* and its two paralogs, *LRB2* and *LRB3* showed a smooth inner tangential tapetum membrane phenotype (**Figure 3.17 F**), while the WT tapetum exhibited a secretory membrane with small globular bodies intensively interspersed on the inner tangential surface (**Figure 3.17 E**). This suggests that there may be *LRB* gene redundancy, which is resulting in the lack of phenotype in the LRB1 knock-out line. Currently, it is not clear whether the PPI between MS1 and LRB1 mediates the rapid turnover and removal of the MS1 protein, and whether the absence of this PPI is the reason behind the tapetum phenotype of *lrb123*. However, the subtle but tapetum-specific *lrb123* phenotype provides a promising clue as to the possible mechanism for post-translational regulation of MS1. It would be interesting in future work to assay if expressing a ubiquitination site mutated MS1 protein in the homozygous *ms1* plant can copy the *lrb123* phenotype. Additionally, there are more questions to be addressed, such as whether the PPI between MS1 and LRB1 can eventually lead to ubiquitin-mediated degradation of the MS1 protein, also what is the chemical component of the tapetum secretory structure which is missing in the *lrb123* mutant, whether these chemical components are essential for pollen development, and

whether the rapid removal of MS1 is biologically meaningful to pollen viability? Considering the *MS1* is under autorepression, the MS1 protein may be stabilised during later stages of tapetal developemtn in *lrb123* campared with the WT, which will enhance this auto-repression. Hence future work could include qRT-PCR of *MS1* mRNA in *lrb123*, which may provide a clue as to whether LRBs have effects on the degradation of the MS1 protein.

In summary, the tight regulation of MS1 relies upon multiple levels of control, including (i) transcriptional activation by a combination of TFs, (ii) direct autorepression by MS1 proteins that bind to its gene body, and (iii) post-translational proteolysis possibly mediated by the LRBs. The combinatorial control of MS1 extends the regulation specificity and regulatory repertoire of tapetum development. As gene function is often intimately linked to when and where the gene is expressed, the multilevel regulation of MS1 found in this work has highlighted the significance of the tight regulation of MS1 to tapetum development and shed light on future research directions of the regulatory network of late tapetum development. The next challenge is to understand how important this multi-level regulation of MS1 is to anther development and male fertility. It will be interesting manipulating specific aspects of the regulation and see what effect would be caused on the development of the anther.

6.2 MYB99 ACTS AS AN INDIRECT TARGET OF MS1 AND ITS ROLE IN LATE TAPETUM DEVELOPMENT REMAINS UNCLEAR

MYB99 has been long been proposed as a direct regulatory target of MS1 based on the results of an induction experiment using a DEX-inducible MS1 line (Ito et al., 2007). However, this presumed direct regulation lacks molecular biological evidence. On the other hand, based on LCM two-colour-array data (Vizcay Barrena and Wilson, unpublished), MYB99 has a tapetum-specific high expression, which initiates at the stage of tetrads and lasts until PMII. In the *ms1* mutant, the expression of MYB99 is completely abolished. Tapetum specifically expressed *MYB* genes have been shown to play important roles in tapetum development, such as the *GAMYBs*, *TDF1/MYB35* and *MS188/MYB80* (Phan et al., 2011; Zhu et al., 2008; Aya et al., 2011a). Therefore, the highly expressed late-tapetum-specific MYB protein, MYB99, may also play a major role in tapetum development.

The ChIP analysis in this work indicates that MS1 does not directly regulate MYB99. This ChIP result is contrary to the conclusion Ito et al. (2007) made that *MYB99* gene was a direct target of MS1. To interpret this contradiction, a close inspection and re-analysis of the original data of Ito et al. (2007) was carried out. It turns out that the data Ito et al. (2007) does not rule out the possibility that MYB99 is a secondary target of MS1. Hence, although MYB99 acts closely downstream of MS1 in tapetum development, it appears to not be directly regulated by MS1.

The reported phenotype of *myb99* tapetum is weak and ambiguous (Alves-Ferreira et al., 2007), which makes it hard to characterise the function of MYB99. Detailed

phenotyping of the *myb99-1* and *myb99*-2 mutants did not find any significant difference between the mutant and WT (**Section 5.2.1** and **5.2.2**). The MYB proteins comprise one of the largest families of TFs in plants, which also result in a high degree of redundancy within those MYBs sharing highly similar protein sequences (Dubos et al., 2010). The MYB gene redundancy may explain a lack of obvious phenotype in the *myb99* mutant. By a maximum likelihood phylogenetic tree generated based on the MYB protein sequence similarity (Kelemen et al., 2015), the genes *MYB20*, *MYB40*, *MYB42*, *MYB43* and *MYB85*, which locate within the same clade with *MYB99*, may have a possible redundant function with *MYB99*. However, the tapetal expression levels of these genes are extremely low based on the anther LCM two-colour-array data (Vizcay Barrena and Wilson, unpublished), which suggests that one or several evolutionarily distinct MYB paralogs, rather than the evolutionarily close ones, may act redundantly with *MYB99*. Nevertheless, the possibility cannot be excluded that the evolutionarily close MYBs exert a compensatory expression in the *myb99* mutant to replace the function of MYB99. qRT-PCR of the mRNAs of the evolutionarily close paralogs in *myb99* could be conducted to help verify this hypothesis.

It is difficult to identify the MYB99 function using the single mutants in which the possibly redundant gene or genes compensate the regulatory role of MYB99 proteins. Nevertheless overexpression of MYB99, or transformation of a chimeric activator or repressor version of MYB99 into plants may generate phenotypes, which can provide an alternative way to do the functional analysis of MYB99. The chimeric repressor gene-silencing technology (CRES-T) is widely adopted to convert a transcription factor into a strong repressor by fusion with an SRDX repression domain, which is then expressed in plants to induce a loss-of-function phenotype (Oshima et al., 2011).

Analogously, by fusion with a VP16 viral trans-activation domain and then expression in planta, a TF of unknown function can be converted to a strong activator, providing opportunities to uncover the TF function (Fujiwara et al., 2014). Future work could include the generation of a chimeric activator/repressor of MYB99 driven by a constitutive or native promoter and transformation into plants to help identify the function of MYB99.

In summary, MYB99 appears to act as a downstream indirect target of MS1. The gene redundancy of MYB99 makes its functional analysis difficult, a chimeric repressor fusion MYB99-SRDX or a chimeric activator fusion MYB99-VP16 may be generated in the future work to make dominant mutant of MYB99 for further function studies.

6.3 MS1 MIGHT REGULATE TAPETUM DEVELOPMENT BY REPRESSING GENE EXPRESSION

It is well established that MS1 has a transcriptional regulatory role in tapetum development, regulating a large number of genes that are involved in determining late tapetum activities, such as pollen wall component secretion and tapetal PCD (Yang et al., 2007; Alves-Ferreira et al., 2007). However, the exact molecular function of MS1 is poorly understood. In most cases, MS1 has been presumed to be a transcriptional activator (Yu, 2015; Ito et al., 2007), and to directly activate the expression of MYB99 in late tapetum (Alves-Ferreira et al., 2007; Yu, 2015). However, the ChIP result and transcriptomic meta-analysis in this work do not support this assumption. Firstly, three sets of independently conducted ChIP experiment indicated that there is no significant binding of the MS1 protein to the MYB99 gene locus, which suggests that

MYB99 is not a direct regulatory target of MS1. Secondly, among the MS1 immediately responding genes (IRGs), most tapetum functional genes are switched off by MS1, whereas the expression of a small proportion of the MS1 IRGs, including MYB99, is switched-on by MS1. This phenomenon suggests MS1 has a repressing role rather than activating one. Lastly, as discussed in **Section 4.3.4**, more recent data indicates that the evidence provided in the work of Ito et al. (2007) is not sufficient to prove the activating role of MS1.

As discussed in **Section 3.3.2**, MS1 protein can auto-repress its expression via binding to the second exon of MS1 locus. Although the molecular mechanism of the MS1 auto-repression is poorly understood, the co-localization of H3K27me3 and LHP1 protein to the MS1 binding site at MS1 locus suggests that MS1 may mediate a Polycomb-mediated repression. This hypothesis is consistent with the fact that the MS1 protein localises to numerous subnuclear rounded speckles that resemble the discrete foci of LHP1 or plant PcG body, where chromatin compaction, H3K27me3 spreading and PRC2 nucleation occur (Förderer et al., 2016). A working model was proposed here illustrating how MS1-mediated repression occurs at the MS1 locus (**Figure 3.18**). The histone mark of H3K27me3 spreads on the genomic locus of MS1, which might play a role in maintaining the MS1 silenced in plant tissues other than the tapetum. After the MS1 is transcribed, the repressive H3K27me3 mark modifying the MS1 locus is less likely to be maintained, instead activated H3K4me3 mark(s) may be placed around the transcription start site. The MS1 H3K4me2/3 binding PHD might therefore play a vital role in localising the MS1 protein at its gene locus. Moreover, possibly via the MS1, MS1/MMD1 Associated Domain1/2 (MAD1/2) domains, PRC2 is consequently recruited, and meanwhile, MS1 proteins aggregate in the PcG bodies. Subsequently the expression of MS1 may be silenced by the

H3K27me3 mark spreading again onto the MS1 genomic locus. Whether this model could be applied to other MS1 repressed targets needs further evidence, such as the MS1 binding to its target genomic loci. Further experiments will be required to identify the MS1 binding site, such as ChIP-SEQ, and determining the effect of MS1 on the overall or local histone tail modification, such as histone-ChIP and histone blots.

6.4 CONCLUDING REMARKS

Among the characterised tapetum regulators, MS1 is a protein that has many interesting characteristics. It regulates a large number of genes associated with tapetum and pollen development, but does not have any DNA binding domains. It determines the crucial phase transition of tapetum development, but the MS1 protein is only present for a very transient period. It is directly regulated by a series of tapetum TFs during early tapetum development, but only commences expression at a relatively late stage. It has been long presumed to act as a TF activator, but direct regulatory targets have not been characterised.

In this work, a combination of approaches was adopted to try to address these inconsistencies presented in MS1 function. A multilevel regulation of MS1 expression has been demonstrated and resolved in detail. The molecular function of MS1 protein was preliminary explored and a functional study of MYB99, a promising target of MS1, was conducted. Lastly, the regulatory relationship between MS1 and MYB99 was determined. Some of the results identified are quite distinct from what was previously proposed or believed.

Ultimately, an MS1-mediated PRC repression model has been proposed based on the findings of this work. How the PRCs are recruited to genomic loci are not yet fully understood and becoming an important issue in epigenetic research (Pu and Sung, 2015; Förderer et al., 2016). A detailed investigation of MS1 regulation mechanisms could provide novel insight into Polycomb-mediated gene repression and highlight the complexity of gene regulation in late tapetum development.

REFERENCES

Alejandro, S., Lee, Y., Tohge, T., Sudre, D., Osorio, S., Park, J., Bovet, L., Lee, Y., Geldner, N., Fernie, A.R., and Martinoia, E. (2012). AtABCG29 is a monolignol transporter involved in lignin biosynthesis. Curr. Biol. **22**: 1207–1212.

Ali, M., Rincón-Arano, H., Zhao, W., Rothbart, S.B., Tong, Q., Parkhurst, S.M., Strahl, B.D., Deng, L.-W., Groudine, M., and Kutateladze, T.G. (2013). Molecular basis for chromatin binding and regulation of MLL5. Proc. Natl. Acad. Sci. U. S. A. **110**: 11296–301.

Allen, R.S., Li, J., Stahle, M.I., Dubroué, A., Gubler, F., and Millar, A. a (2007). Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. Proc. Natl. Acad. Sci. U. S. A. **104**: 16371–16376.

Alonso, J.M. (2003). Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana. Science (80-.). **301**: 653–657.

Alves-Ferreira, M., Wellmer, F., Banhara, A., Kumar, V., Riechmann, J.L., and Meyerowitz, E.M. (2007). Global expression profiling applied to the analysis of Arabidopsis stamen development. Plant Physiol. **145**: 747–62.

Andreuzza, S., Nishal, B., Singh, A., and Siddiqi, I. (2015). The Chromatin Protein DUET/MMD1 Controls Expression of the Meiotic Gene TDM1 during Male Meiosis in Arabidopsis. PLoS Genet. **11**: e1005396.

Ariizumi, T., Hatakeyama, K., Hinata, K., Sato, S., Kato, T., Tabata, S., and Toriyama, K. (2005). The HKM gene, which is identical to the MS1 gene of Arabidopsis thaliana, is essential for primexine formation and exine pattern formation. Sex. Plant Reprod. **18**: 1–7.

Ariizumi, T. and Toriyama, K. (2011). Genetic regulation of sporopollenin synthesis and pollen exine development. Annu Rev Plant Biol **62**: 437–460.

Aya, K., Hiwatashi, Y., Kojima, M., Sakakibara, H., Ueguchi-Tanaka, M., Hasebe, M., and Matsuoka, M. (2011a). The Gibberellin perception system evolved to regulate a pre-existing GAMYB-mediated system during land plant evolution. Nat. Commun. **2**: 544.

Aya, K., Suzuki, G., Suwabe, K., Hobo, T., Takahashi, H., Shiono, K., Yano, K., Tsutsumi, N., Nakazono, M., Nagamura, Y., Matsuoka, M., and Watanabe, M. (2011b). Comprehensive network analysis of anther-expressed genes in rice by the combination of 33 laser microdissection and 143 spatiotemporal microarrays. PLoS One **6**: e26162.

Aya, K., Ueguchi-Tanaka, M., Kondo, M., Hamada, K., Yano, K., Nishimura, M., and Matsuoka, M. (2009). Gibberellin modulates anther development in rice via the transcriptional regulation of GAMYB. Plant Cell **21**: 1453–72.

Bai, B., Wu, J., Sheng, W.-T., Zhou, B., Zhou, L.-J., Zhuang, W., Yao, D.-P., and Deng, Q.-Y. (2015). Comparative Analysis of Anther Transcriptome Profiles of Two Different Rice Male Sterile Lines Genotypes under Cold Stress. Int. J. Mol. Sci. **16**: 11398–11416.

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res. **43**: W39–W49.

Becker, P.B. (2006). Gene regulation: a finger on the mark. Nature **442**: 31–2.

Bedinger, P. (1992). The remarkable biology of pollen. Plant Cell **4**: 879–87.

Berger, N. and Dubreucq, B. (2012). Evolution goes GAGA: GAGA binding proteins across kingdoms. Biochim. Biophys. Acta - Gene Regul. Mech. **1819**: 863–868.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., and Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. **42**.

Blackmore, S. and Knox, R.B. (2016). Microspores Evolution and Ontogeny: Evolution and Ontogeny (Elsevier Science).

Blackmore, S., Wortley, A.H., Skvarla, J.J., and Rowley, J.R. (2007). Pollen wall development in flowering plants. New Phytol. **174**: 483–98.

Böhmdorfer, G. and Wierzbicki, A.T. (2015). Control of Chromatin Structure by Long Noncoding RNA. Trends Cell Biol. **25**: 623–632.

Botté, C.Y. and Maréchal, E. (2014). Plastids with or without galactoglycerolipids. Trends Plant Sci. **19**: 71–78.

Bowman, J.L., Drews, G.N., and Meyerowitz, E.M. (1991a). Expression of the Arabidopsis Floral Homeotic Gene Agamous Is Restricted to Specific Cell-Types Late in Flower Development. Plant Cell **3**: 749–758.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1991b). Genetic interactions among floral homeotic genes of Arabidopsis. Development **112**: 1–20.

Brameier, M., Krings, A., and MacCallum, R.M. (2007). NucPred - Predicting nuclear localization of proteins. Bioinformatics **23**: 1159–1160.

Brown, J.W.S., Shaw, P.J., Shaw, P., and Marshall, D.F. (2005). Arabidopsis

nucleolar protein database (AtNoPDB). Nucleic Acids Res. **33**: D633–D636.

Van Buskirk, E.K., Decker, P. V., and Chen, M. (2012). Photobodies in Light Signaling. Plant Physiol. **158**: 52–60.

Canales, C., Bhatt, A.M., Scott, R., and Dickinson, H. (2002). EXS, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in Arabidopsis. Curr. Biol. **12**: 1718–27.

Cardarelli, M. and Cecchetti, V. (2014). Auxin polar transport in stamen formation and development: how many actors? Front. Plant Sci. **5**: 333.

Chang, F., Wang, Y., Wang, S., and Ma, H. (2011). Molecular control of microsporogenesis in Arabidopsis. Curr. Opin. Plant Biol. **14**: 66–73.

Charoensawan, V., Wilson, D., and Teichmann, S.A. (2010). Lineage-specific expansion of DNA-binding transcription factor families. Trends Genet. **26**: 388–393.

Chekanova, J.A. (2015). Long non-coding RNAs and their functions in plants. Curr. Opin. Plant Biol. **27**: 207–216.

Chen, C. et al. (2010). Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated Arabidopsis male meiocytes. BMC Plant Biol. **10**: 280.

Chen, L. and Liu, Y.-G. (2014). Male sterility and fertility restoration in crops. Annu. Rev. Plant Biol. **65**: 579–606.

Chen, M. and Chory, J. (2011). Phytochrome signaling mechanisms and the control of plant development. Trends Cell Biol. **21**: 664–671.

Chen, R. et al. (2015). A Gene Expression Profiling of Early Rice Stamen Development that Reveals Inhibition of Photosynthetic Genes by OsMADS58. Mol. Plant **8**: 1069–1089.

Chen, W., Yu, X.H., Zhang, K., Shi, J., De Oliveira, S., Schreiber, L., Shanklin, J., and Zhang, D. (2011). Male Sterile2 encodes a plastid-localized fatty acyl carrier protein reductase required for pollen exine development in Arabidopsis. Plant Physiol **157**: 842–853.

Chow, C.-N., Zheng, H.-Q., Wu, N.-Y., Chien, C.-H., Huang, H.-D., Lee, T.-Y., Chiang-Hsieh, Y.-F., Hou, P.-F., Yang, T.-Y., and Chang, W.-C. (2016). PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. Nucleic Acids Res. **44**: D1154-60.

La Cour, T., Kiemer, L., Mølgaard, A., Gupta, R., Skriver, K., and Brunak, S. (2004). Analysis and prediction of leucine-rich nuclear export signals. Protein Eng. Des. Sel. **17**: 527–536.

Cui, J., You, C., Zhu, E., Huang, Q., Ma, H., and Chang, F. (2016). Feedback Regulation of DYT1 by Interactions with Downstream bHLH Factors Promotes DYT1 Nuclear Localization and Anther Development. Plant Cell **28**: 1078–1093.

Davuluri, R. V, Sun, H., Palaniswamy, S.K., Matthews, N., Molina, C., Kurtz, M., and Grotewold, E. (2003). AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics **4**: 25.

Derkacheva, M., Steinbach, Y., Wildhaber, T., Mozgová, I., Mahrez, W., Nanni, P., Bischof, S., Gruissem, W., and Hennig, L. (2013). Arabidopsis MSI1 connects LHP1 to PRC2 complexes. EMBO J. **32**: 2073–85.

Dobritsa, A. a., Nishikawa, S.-I., Preuss, D., Urbanczyk-Wochniak, E., Sumner, L.W., Hammond, A., Carlson, A.L., and Swanson, R.J. (2009a). LAP3, a novel plant protein required for pollen development, is essential for proper exine formation. Sex. Plant Reprod. **22**: 167–77.

Dobritsa, A. a, Shrestha, J., Morant, M., Pinot, F., Matsuno, M., Swanson, R., Møller, B.L., and Preuss, D. (2009b). CYP704B1 is a long-chain fatty acid omega-hydroxylase essential for sporopollenin synthesis in pollen of Arabidopsis. Plant Physiol. **151**: 574–89.

Du, H., Zhang, L., Liu, L., Tang, X.-F., Yang, W.-J., Wu, Y.-M., Huang, Y.-B., and Tang, Y.-X. (2009). Biochemical and molecular characterization of plant MYB transcription factor family. Biochem. Biokhimiia **74**: 1–11.

Dubos, C. et al. (2014). Integrating bioinformatic resources to predict transcription factors interacting with cis-sequences conserved in co-regulated genes. BMC Genomics **15**: 317.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. Trends Plant Sci. **15**: 573–81.

Dukowic-Schulze, S. and Chen, C. (2014). The meiotic transcriptome architecture of plants. Front. Plant Sci. **5**: 220.

Echlin, P. and Godwin, H. (1969). The ultrastructure and ontogeny of pollen in Helleborus foetidus L. 3. The formation of the pollen grain wall. J. Cell Sci. **5**: 459–77.

Edstam, M.M. and Edqvist, J. (2014). Involvement of GPI-anchored lipid transfer proteins in the development of seed coats and pollen in Arabidopsis thaliana. Physiol. Plant. **152**: 32–42.

Ehlting, J., Sauveplane, V., Olry, A., Ginglinger, J.-F., Provart, N.J., and Werck-Reichhart, D. (2008). An extensive (co-)expression analysis tool for the

cytochrome P450 superfamily in Arabidopsis thaliana. BMC Plant Biol. **8**: 47.

El-Ghazaly, G. (1999). Tapetum and Orbicules (Ubisch Bodies): Development, Morphology and Role of Pollen Grains and Tapetal Orbicules in Allergenicity. In Fertilization in Higher Plants (Springer Berlin Heidelberg: Berlin, Heidelberg), pp. 157–173.

Entrevan, M., Schuettengruber, B., and Cavalli, G. (2016). Regulation of Genome Architecture and Function by Polycomb Proteins. Trends Cell Biol. **26**: 511–525.

Fan, X., Yang, C., Klisch, D., Ferguson, A., Bhaellero, R.P., Niu, X., and Wilson, Z.A. (2014). ECHIDNA protein impacts on male fertility in Arabidopsis by mediating trans-Golgi network secretory trafficking during anther and pollen development. Plant Physiol. **164**: 1338–49.

Farquharson, K.L. (2016). A Domain in the bHLH Transcription Factor DYT1 Is Critical for Anther Development. Plant Cell **28**: 997–998.

Fellenberg, C. and Vogt, T. (2015). Evolutionarily conserved phenylpropanoid pattern on angiosperm pollen. Trends Plant Sci. **20**: 212–218.

Feng, B., Lu, D., Ma, X., Peng, Y., Sun, Y., Ning, G., and Ma, H. (2012). Regulation of the Arabidopsis anther transcriptome by DYT1 for pollen development. Plant J. **72**: 612–24.

Feng, X. and Dickinson, H.G. (2010). Tapetal cell fate, lineage and proliferation in the Arabidopsis anther. Development **137**: 2409–16.

Feng, X., Zilberman, D., and Dickinson, H. (2013). A conversation across generations: soma-germ cell crosstalk in plants. Dev. Cell **24**: 215–25.

Ferguson, A.C., Pearce, S., Band, L.R., Yang, C., Ferjentsikova, I., King, J., Yuan, Z., Zhang, D., and Wilson, Z.A. (2016). Biphasic regulation of the transcription factor ABORTED MICROSPORES (AMS) is essential for tapetum and pollen development in Arabidopsis. New Phytol. **213**: 778–790.

Fernández Gómez, J. and Wilson, Z. a. (2014). A barley PHD finger transcription factor that confers male sterility by affecting tapetal development. Plant Biotechnol. J.: 1–13.

Finn, R.D. et al. (2016). The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res. **44**: D279–D285.

Förderer, A., Zhou, Y., and Turck, F. (2016). The age of multiplexity: recruitment and interactions of Polycomb complexes in plants. Curr. Opin. Plant Biol. **29**: 169–178.

Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P.,

and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. Proc. Natl. Acad. Sci. U. S. A. **111**: 2367–72.

Fraser, D.P., Hayes, S., and Franklin, K.A. (2016). Photoreceptor crosstalk in shade avoidance. Curr. Opin. Plant Biol. **33**: 1–7.

Fu, Z., Yu, J., Cheng, X., Zong, X., Xu, J., Chen, M., Li, Z., Zhang, D., and Liang, W. (2014). The Rice Basic Helix-Loop-Helix Transcription Factor TDR INTERACTING PROTEIN2 Is a Central Switch in Early Anther Development. Plant Cell **1**.

Fujiwara, S., Sakamoto, S., Kigoshi, K., Suzuki, K., and Ohme-Takagi, M. (2014). VP16 fusion induces the multiple-knockout phenotype of redundant transcriptional repressors partly by Med25-independent mechanisms in Arabidopsis. FEBS Lett. **588**: 3665–3672.

Geilen, K. et al. (2015). Dynamic subnuclear relocalisation of WRKY40 in response to Abscisic acid in Arabidopsis thaliana. Sci. Rep. **5**: 13369.

Geilen, K. and Böhmer, M. (2015). Dynamic subnuclear relocalization of WRKY40, a potential new mechanism of ABA-dependent transcription factor regulation. Plant Signal. Behav. **10**: e1106659.

Gietz, R.D. and Schiestl, R.H. (2007). Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. **2**: 35–37.

Gocal, G.F., Sheldon, C.C., Gubler, F., Moritz, T., Bagnall, D.J., MacMillan, C.P., Li, S.F., Parish, R.W., Dennis, E.S., Weigel, D., and King, R.W. (2001). GAMYB-like genes, flowering, and gibberellin signaling in Arabidopsis. Plant Physiol. **127**: 1682–93.

Gómez, J.F., Talle, B., and Wilson, Z.A. (2015). Anther and pollen development: A conserved developmental pathway. J. Integr. Plant Biol. **57**: 876–91.

Gong, F., Wu, X., and Wang, W. (2015). Diversity and function of maize pollen coat proteins: from biochemistry to proteomics. Front. Plant Sci. **6**.

Good, B.H. and Chapman, R.L. (1978). The Ultrastructure of Phycopeltis (Chroolepidaceae: Chlorophyta). I. Sporopollenin in the Cell Walls. Am. J. Bot. **65**: 27.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. **40**: D1178–D1186.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten

years of next-generation sequencing technologies. Nat Rev Genet **17**: 333–351.

Grienenberger, E., Kim, S.S., Lallemand, B., Geoffroy, P., Heintz, D., Souza, C.D.A., Heitz, T., Douglas, C.J., and Legrand, M. (2010). Analysis of TETRAKETIDE α-PYRONE REDUCTASE function in Arabidopsis thaliana reveals a previously unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. Plant Cell **22**: 4067–83.

Gu, J., Zhu, J., Yu, Y., Teng, X., Lou, Y., Xu, X., Liu, J., and Yang, Z. (2014). DYT1 directly regulates the expression of TDF1 for tapetum development and pollen wall formation in Arabidopsis. Plant J. **80**: 1005–13.

Guo, T. and Fang, Y. (2014). Functional organization and dynamics of the cell nucleus. Front. Plant Sci. **5**: 378.

Haberer, G., Mader, M.T., Kosarev, P., Spannagl, M., Yang, L., and Mayer, K.F.X. (2006). Large-scale cis-element detection by analysis of correlated expression and sequence conservation between Arabidopsis and Brassica oleracea. Plant Physiol. **142**: 1589–602.

Harris, C.J., Husmann, D., Liu, W., Kasmi, F. El, Wang, H., Papikian, A., Pastor, W. a, Moissiard, G., Vashisht, A.A., Dangl, J.L., Wohlschlegel, J.A., and Jacobsen, S.E. (2016). Arabidopsis AtMORC4 and AtMORC7 Form Nuclear Bodies and Repress a Large Number of Protein-Coding Genes. PLoS Genet. **12**: e1005998.

Hautegem, T. Van, Durme, M. Van, Olvera-carrillo, Y., Huysmans, M., Karimi, M., Lippens, S., Gue, C.J., Krebs, M., and Schumacher, K. (2014). Article Programmed Cell Death Controlled by ANAC033 / SOMBRERO Determines Root Cap Organ Size in Arabidopsis. Curr. Biol. **24**: 931–940.

He, G., Elling, A. a, and Deng, X.W. (2011). The epigenome and plant development. Annu. Rev. Plant Biol. **62**: 411–435.

He, Y. (2012). Chromatin regulation of flowering. Trends Plant Sci. **17**: 556–62.

Hecker, A., Brand, L.H., Peter, S., Simoncello, N., Kilian, J., Harter, K., Gaudin, V., and Wanke, D. (2015). The Arabidopsis GAGA-Binding Factor BASIC PENTACYSTEINE6 Recruits the POLYCOMB-REPRESSIVE COMPLEX1 Component LIKE HETEROCHROMATIN PROTEIN1 to GAGA DNA Motifs. Plant Physiol. **168**: 1013–24.

Hennig, L. and Derkacheva, M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? Trends Genet. **25**: 414–423.

Higginson, T., Li, S.F., and Parish, R.W. (2003). AtMYB103 regulates tapetum and trichome development in Arabidopsis thaliana. Plant J. **35**: 177–192.

Higo, K. (1998). PLACE: a database of plant cis-acting regulatory DNA elements. Nucleic Acids Res. **26**: 358–359.

Hird, D.L., Worrall, D., Hodge, R., Smartt, S., Paul, W., and Scott, R. (1993). The anther-specific protein encoded by the Brassica napus and Arabidopsis thaliana A6 gene displays similarity to beta-1,3-glucanases. Plant J. **4**: 1023–33.

Hong, S.M., Bahn, S.C., Lyu, A., Jung, H.S., and Ahn, J.H. (2010). Identification and Testing of Superior Reference Genes for a Starting Pool of Transcript Normalization in Arabidopsis. Plant Cell Physiol. **51**: 1694–1706.

Hsieh, K. and Huang, A.H.C. (2004). Endoplasmic reticulum, oleosins, and oils in seeds and tapetum cells. Plant Physiol. **136**: 3427–34.

Hu, J., Wang, Z., Zhang, L., and Sun, M.X. (2014). The Arabidopsis Exine Formation Defect (EFD) gene is required for primexine patterning and is critical for pollen fertility. New Phytol. **203**: 140–154.

Hu, Y., Liu, D., Zhong, X., Zhang, C., Zhang, Q., and Zhou, D.-X. (2012). CHD3 protein recognizes and regulates methylated histone H3 lysines 4 and 27 over a subset of targets in the rice genome. Proc. Natl. Acad. Sci. **109**: 5773–5778.

Huang, C.-H., Su, M.-G., Kao, H.-J., Jhong, J.-H., Weng, S.-L., and Lee, T.-Y. (2016). UbiSite: incorporating two-layered machine learning method with substrate motifs to predict ubiquitin-conjugation site on lysines. BMC Syst. Biol. **10**: 6.

Huang, M.-D., Hsing, Y.-I.C., and Huang, A.H.C. (2011). Transcriptomes of the anther sporophyte: availability and uses. Plant Cell Physiol. **52**: 1459–66.

Ischebeck, T. (2016). Lipids in pollen — They are different. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids **1861**: 1315–1328.

Ito, T., Nagata, N., Yoshiba, Y., Ohme-Takagi, M., Ma, H., and Shinozaki, K. (2007). Arabidopsis MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development. Plant Cell **19**: 3549–3562.

Ito, T. and Shinozaki, K. (2002). The MALE STERILITY1 gene of Arabidopsis, encoding a nuclear protein with a PHD-finger motif, is expressed in tapetal cells and is required for pollen maturation. Plant Cell Physiol. **43**: 1285–92.

Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J.L., and Meyerowitz, E.M. (2004). The homeotic protein AGAMOUS controls microsporogenesis by regulation of SPOROCYTELESS. Nature **430**: 356–360.

Jarillo, J.A. and Piñeiro, M. (2011). Timing is everything in plant development. The central role of floral repressors. Plant Sci. **181**: 364–78.

Jarillo, J.A., Piñeiro, M., Cubas, P., and Martínez-Zapater, J.M. (2009). Chromatin remodeling in plant development. Int. J. Dev. Biol. **53**: 1581–96.

Jia, Q., Zhu, J., Xu, X., Lou, Y., Zhang, Z., Zhang, Z., and Yang, Z. (2015). Arabidopsis AT-hook Protein TEK Positively Regulates the Expression of Arabinogalactan Proteins for Nexine Formation. Mol. Plant: 1–10.

Jiang, J., Yao, L., Yu, Y., Liang, Y., Jiang, J., Ye, N., Miao, Y., and Cao, J. (2014a). PECTATE LYASE-LIKE 9 from Brassica campestris is associated with intine formation. Plant Sci. **229**: 66–75.

Jiang, J., Yao, L., Yu, Y., Lv, M., Miao, Y., and Cao, J. (2014b). PECTATE LYASE-LIKE10 is associated with pollen wall development in Brassica campestris. J. Integr. Plant Biol. **56**: 1095–105.

Jin, J., Liu, J., Wang, H., Wong, L., and Chua, N.-H. (2013). PLncDB: plant long non-coding RNA database. Bioinformatics **29**: 1068–71.

Kaneko, M., Inukai, Y., Ueguchi-Tanaka, M., Itoh, H., Izawa, T., Kobayashi, Y., Hattori, T., Miyao, A., Hirochika, H., Ashikari, M., and Matsuoka, M. (2004). Loss-of-function mutations of the rice GAMYB gene impair alpha-amylase expression in aleurone and flower development. Plant Cell **16**: 33–44.

Kaufmann, K., Muiño, J.M., Østerås, M., Farinelli, L., Krajewski, P., and Angenent, G.C. (2010). Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nat. Protoc. **5**: 457–72.

Ke, J., Ma, H., Gu, X., Thelen, A., Brunzelle, J.S., Li, J., Xu, H.E., and Melcher, K. (2015). Structural basis for recognition of diverse transcriptional repressors by the TOPLESS family of corepressors. Sci. Adv. **1**: e1500107.

Kelemen, Z. et al. (2015). Analysis of the DNA-Binding Activities of the Arabidopsis R2R3-MYB Transcription Factor Family by One-Hybrid Experiments in Yeast. PLoS One **10**: e0141044.

Ko, S., Li, M., Sun-Ben Ku, M., Ho, Y., Lin, Y., Chuang, M., Hsing, H., Lien, Y., Yang, H., Chang, H., and Chan, M. (2014). The bHLH142 Transcription Factor Coordinates with TDR1 to Modulate the Expression of EAT1 and Regulate Pollen Development in Rice. Plant Cell.

Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc. Natl. Acad. Sci. U. S. A. **106**: 10171–10176.

Krizek, B. a and Fletcher, J.C. (2005). Molecular mechanisms of flower

development: an armchair guide. Nat. Rev. Genet. **6**: 688–698.

Kung, J.T.Y., Colognori, D., and Lee, J.T. (2013). Long noncoding RNAs: past, present, and future. Genetics **193**: 651–69.

Lallemand, B., Erhardt, M., Heitz, T., and Legrand, M. (2013). Sporopollenin biosynthetic enzymes interact and constitute a metabolon localized to the endoplasmic reticulum of tapetum cells. Plant Physiol. **162**: 616–625.

Lemak, A., Yee, A., Wu, H., Yap, D., Zeng, H., Dombrovski, L., Houliston, S., Aparicio, S., and Arrowsmith, C.H. (2013). Solution NMR Structure and Histone Binding of the PHD Domain of Human MLL5. PLoS One **8**: e77020.

Letunic, I., Doerks, T., and Bork, P. (2015). SMART: Recent updates, new developments and status in 2015. Nucleic Acids Res. **43**: D257–D260.

Li, C.F., Henderson, I.R., Song, L., Fedoroff, N., Lagrange, T., and Jacobsen, S.E. (2008). Dynamic regulation of ARGONAUTE4 within multiple nuclear bodies in Arabidopsis thaliana. PLoS Genet. **4**: e27.

Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006a). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature **442**: 91–5.

Li, H., Yuan, Z., Vizcay-Barrena, G., Yang, C., Liang, W., Zong, J., Wilson, Z.A., and Zhang, D. (2011). PERSISTENT TAPETAL CELL1 encodes a PHD-finger protein that is required for tapetal cell death and pollen development in rice. Plant Physiol **156**: 615–630.

Li, N. et al. (2006b). The rice tapetum degeneration retardation gene is required for tapetum degradation and anther development. Plant Cell **18**: 2999–3014.

Li, N., Xu, C., Li-Beisson, Y., and Philippar, K. (2015). Fatty Acid and Lipid Transport in Plant Cells. Trends Plant Sci. **xx**: 1–14.

Li, S.F., Higginson, T., and Parish, R.W. (1999). A novel MYB-related gene from Arabidopsis thaliana expressed in developing anthers. Plant Cell Physiol. **40**: 343–7.

Li, S.F., Iacuone, S., and Parish, R.W. (2007). Suppression and restoration of male fertility using a transcription factor. Plant Biotechnol. J. **5**: 297–312.

Li, Y. and Li, H. (2012). Many keys to push: diversifying the “readership” of plant homeodomain fingers. Acta Biochim. Biophys. Sin. (Shanghai). **44**: 28–39.

Lipsick, J.S. (1996). One billion years of Myb. Oncogene **13**: 223–235.

Liu, J., Wang, H., and Chua, N.-H. (2015a). Long noncoding RNA transcriptome of

plants. Plant Biotechnol. J.: n/a-n/a.

Liu, L. and Fan, X. (2013). Tapetum: regulation and role in sporopollenin biosynthesis in Arabidopsis. Plant Mol. Biol. **83**: 165–75.

Liu, Q., Shi, L., and Fang, Y. (2012). Dicing Bodies. PLANT Physiol. **158**: 61–66.

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., Xue, Y., and Ren, J. (2015b). IBS: An illustrator for the presentation and visualization of biological sequences. Bioinformatics **31**: 3359–3361.

Liu, X., Shangguan, Y., Zhu, J., Lu, Y., and Han, B. (2013). The rice OsLTP6 gene promoter directs anther-specific expression by a combination of positive and negative regulatory elements. Planta **238**: 845–857.

Livak, K.J. and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods **25**: 402–8.

Lou, Y., Xu, X., Zhu, J., Gu, J., Blackmore, S., and Yang, Z. (2014). The tapetal AHL family protein TEK determines nexine formation in the pollen wall. Nat. Commun. **5**: 3855.

Lu, P., Chai, M., Yang, J., Ning, G., Wang, G., and Ma, H. (2014). The Arabidopsis CALLOSE DEFECTIVE MICROSPORE1 gene is required for male fertility through regulating callose metabolism during microsporogenesis. Plant Physiol. **164**: 1893–904.

De Lucia, F., Crevillen, P., Jones, A.M., Greb, T., and Dean, C. (2008). A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. Proc Natl Acad Sci U S A **105**: 16831–16836.

Ma, H. (2005). Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. Annu Rev Plant Biol **56**: 393–434.

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

Mao, Y.S., Zhang, B., and Spector, D.L. (2011). Biogenesis and function of nuclear bodies. Trends Genet. **27**: 295–306.

Mar??chal, A. and Brisson, N. (2010). Recombination and the maintenance of plant organelle genome stability. New Phytol. **186**: 299–317.

Mayfield, J.A. (2001). Gene Families from the Arabidopsis thaliana Pollen Coat Proteome. Science (80-.). **292**: 2482–2485.

Mayr, E. (1986). Joseph Gottlieb Kolreuter's Contributions to Biology. Osiris **2**: 135–176.

Mellor, J. (2006). It takes a PHD to read the histone code. Cell **126**: 22–4.

Millar, A. a and Gubler, F. (2005). The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. Plant Cell **17**: 705–21.

Mitsuda, N. and Ohme-Takagi, M. (2009). Functional analysis of transcription factors in Arabidopsis. Plant Cell Physiol. **50**: 1232–48.

Molitor, A.M., Bu, Z., Yu, Y., and Shen, W.-H. (2014). Arabidopsis AL PHD-PRC1 Complexes Promote Seed Germination through H3K4me3-to-H3K27me3 Chromatin State Switch in Repression of Seed Developmental Genes. PLoS Genet. **10**: e1004091.

Monfared, M.M., Simon, M.K., Meister, R.J., Roig-Villanova, I., Kooiker, M., Colombo, L., Fletcher, J.C., and Gasser, C.S. (2011). Overlapping and antagonistic activities of BASIC PENTACYSTEINE genes affect a range of developmental processes in Arabidopsis. Plant J. **66**: 1020–1021.

Morant, M., Jørgensen, K., Schaller, H., Pinot, F., Møller, B.L., Werck-Reichhart, D., and Bak, S. (2007). CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. Plant Cell **19**: 1473–87.

Mouriz, A., López-González, L., Jarillo, J.A., and Piñeiro, M. (2015). PHDs govern plant development. Plant Signal. Behav. **10**: e993253.

Mozgova, I., Köhler, C., and Hennig, L. (2015). Keeping the gate closed: Functions of the Polycomb repressive complex PRC2 in development. Plant J. **83**: n/a-n/a.

Musselman, C. a and Kutateladze, T.G. (2011). Handpicking epigenetic marks with PHD fingers. Nucleic Acids Res. **39**: 9061–71.

Narula, K., Datta, A., Chakraborty, N., and Chakraborty, S. (2013). Comparative analyses of nuclear proteome: extending its function. Front. Plant Sci. **4**: 100.

Ni, W., Xu, S.-L., Tepperman, J.M., Stanley, D.J., Maltby, D. a, Gross, J.D., Burlingame, A.L., Wang, Z.-Y., and Quail, P.H. (2014). A mutually assured destruction mechanism attenuates light signaling in Arabidopsis. Science **344**: 1160–4.

Niederhuth, C.E., Patharkar, O.R., and Walker, J.C. (2013). Transcriptional profiling of the Arabidopsis abscission mutant hae hsl2 by RNA-Seq. BMC Genomics **14**: 37.

Nishikawa, S., Zinkl, G.M., Swanson, R.J., Maruyama, D., and Preuss, D. (2005). Callose (beta-1,3 glucan) is essential for Arabidopsis pollen wall patterning, but not tube growth. BMC Plant Biol. **5**: 22.

Olvera-Carrillo, Y. et al. (2015). A conserved core of PCD indicator genes discriminates developmentally and environmentally induced programmed cell death in plants.

Oshima, Y., Mitsuda, N., Nakata, M., Nakagawa, T., Nagaya, S., Kato, K., and Ohme-Takagi, M. (2011). Novel vector systems to accelerate functional analysis of transcription factors using chimeric repressor gene-silencing technology (CRES-T). Plant Biotechnol. **28**: 201–210.

Owen, H. a. and Makaroff, C. a. (1995). Ultrastructure of microsporogenesis and microgametogenesis inArabidopsis thaliana (L.) Heynh. ecotype Wassilewskija (Brassicaceae). Protoplasma **185**: 7–21.

Pacini, E. (1990). 11 – Tapetum and microspore function. In Microspores Evolution and Ontogeny, pp. 213–237.

Pacini, E. and Franchi, G.G. (1993). Role of the tapetum in pollen and spore dispersal. In The Tapetum Cytology, Function, Biochemistry and Evolution, pp. 1–11.

Papini, A., Mosti, S., and Brighigna, L. (1999). Programmed-cell-death events during tapetum development of angiosperms. Protoplasma **207**: 213–221.

Parish, R.W. and Li, S.F. (2010). Death of a tapetum: A programme of developmental altruism. Plant Sci. **178**: 73–89.

Pathan, M. et al. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics **15**: 2597–2601.

Pearce, S., Ferguson, A., King, J., and Wilson, Z.A. (2015). FlowerNet: a gene expression correlation network for anther and pollen development. Plant Physiol. **167**: 1717–30.

Peña, P. V, Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V. V, Gozani, O., Zhao, R., and Kutateladze, T.G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature **442**: 100–3.

Peng, J., Li, Z., Wen, X., Li, W., Shi, H., Yang, L., Zhu, H., and Guo, H. (2014). Salt-induced stabilization of EIN3/EIL1 confers salinity tolerance by deterring ROS accumulation in Arabidopsis. PLoS Genet. **10**: e1004664.

Peterson, R., Slovin, J.P., and Chen, C. (2010). A simplified method for differential staining of aborted and non-aborted pollen grains. Int. J. Plant Biol. **1**.

Phan, H.A., Iacuone, S., Li, S.F., and Parish, R.W. (2011). The MYB80 transcription factor is required for pollen development and the regulation of tapetal programmed cell death in Arabidopsis thaliana. Plant Cell **23**: 2209–24.

Phan, H. a, Li, S.F., and Parish, R.W. (2012). MYB80, a regulator of tapetal and pollen development, is functionally conserved in crops. Plant Mol. Biol. **78**: 171–83.

Picot, E., Krusche, P., Tiskin, A., Carré, I., and Ott, S. (2010). Evolutionary analysis of regulatory sequences (EARS) in plants. Plant J. **64**: 165–76.

Pina, C., Pinto, F., Feijo, J.A., and Becker, J.D. (2005). Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol. **138**: 744–756.

Pires, N. and Dolan, L. (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. Mol. Biol. Evol. **27**: 862–74.

Pireyre, M. and Burow, M. (2015). Regulation of MYB and bHLH transcription factors: a glance at the protein level. Mol. Plant **8**: 378–88.

Proost, S., Van Bel, M., Vaneechoutte, D., Van de Peer, Y., Inze, D., Mueller-Roeber, B., and Vandepoele, K. (2015). PLAZA 3.0: an access point for plant comparative genomics. Nucleic Acids Res. **43**: D974–D981.

Pruneda-Paz, J.L., Breton, G., Para, A., and Kay, S.A. (2009). A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock. Science **323**: 1481–5.

Pu, L. and Sung, Z.R. (2015). PcG and trxG in plants - friends or foes. Trends Genet. **31**: 252–262.

Quilichini, T.D., Douglas, C.J., and Samuels, a L. (2014a). New views of tapetum ultrastructure and pollen exine development in Arabidopsis thaliana. Ann. Bot.: mcu042.

Quilichini, T.D., Grienenberger, E., and Douglas, C.J. (2014b). The biosynthesis, composition and assembly of the outer pollen wall: A tough case to crack. Phytochemistry.

Quilichini, T.D., Samuels, A.L., and Douglas, C.J. (2014c). ABCG26-mediated polyketide trafficking and hydroxycinnamoyl spermidines contribute to pollen wall exine formation in Arabidopsis. Plant Cell **26**: 4483–98.

Raghavan, V. (1988). Anther and pollen development in rice (Oryza sativa). Am. J. Bot.: 183–196.

Rastogi, S. and Rost, B. (2011). LocDB: Experimental annotations of localization for

homo sapiens and arabidopsis thaliana. Nucleic Acids Res. **39**: D230-4.

Reddy, A.S.N., Day, I.S., Gohring, J., and Barta, A. (2012). Localization and Dynamics of Nuclear Speckles in Plants. Plant Physiol. **158**: 67–77.

Regan, S.M. and Moffatt, B.A. (1990). Cytochemical Analysis of Pollen Development in Wild-Type Arabidopsis and a Male-Sterile Mutant. Plant Cell **2**: 877–889.

Reineke, A.R., Bornberg-Bauer, E., and Gu, J. (2011). Evolutionary divergence and limits of conserved non-coding sequence detection in plant genomes. Nucleic Acids Res. **39**: 6029–6043.

Rejón, J., Delalande, F., Schaeffer-Reiss, C., Alché, J., Rodríguez-García, M., Van Dorsselaer, A., and Castro, A. (2016). The Pollen Coat Proteome: At the Cutting Edge of Plant Reproduction. Proteomes **4**: 5.

Renzaglia, K.S., Lopez, R. a., and Johnson, E.E. (2014). Callose is integral to the development of permanent tetrads in the liverwort Sphaerocarpos. Planta **241**: 615–627.

Sakakibara, K. (2016). Technological Innovations Give Rise to a New Era of Plant Evolutionary Developmental Biology. In Advances in Botanical Research, pp. 3–35.

Sanchez, R. and Zhou, M.-M. (2011). The PHD finger: a versatile epigenome reader. Trends Biochem. Sci. **36**: 364–72.

Sanders, P.M., Bui, A.Q., Weterings, K., McIntire, K.N., Hsu, Y.-C., Lee, P.Y., Truong, M.T., Beals, T.P., and Goldberg, R.B. (1999). Anther developmental defects in Arabidopsis thaliana male-sterile mutants. Sex. Plant Reprod. **11**: 297–322.

Schiefthaler, U., Balasubramanian, S., Sieber, P., Chevalier, D., Wisman, E., and Schneitz, K. (1999). Molecular analysis of NOZZLE, a gene involved in pattern formation and early sporogenesis during sex organ development in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. **96**: 11664–9.

Scott, R.J., Spielman, M., and Dickinson, H.G. (2004). Stamen structure and function. Plant Cell **16 Suppl**: S46-60.

Shi, J., Cui, M., Yang, L., Kim, Y.J., and Zhang, D. (2015). Genetic and Biochemical Mechanisms of Pollen Wall Development. Trends Plant Sci. **20**: 741–753.

Shi, X. et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature **442**: 96–9.

Shukla, A.K., Vijayaraghavan, M.R., and Chaudhry, B. (1998). Biology of Pollen 1st Ed. (APH Publishing Corporation: New Delhi).

Shumin, Z., Yan, C., Hulin, S., Bang, Z., Licheng, S., and Wei, Z. (2015). One novel cis-element is essential for correct DYSFUNCTIONAL TAPETUM 1 (DYT1) expression in Arabidopsis thaliana. Plant Cell Rep. **34**: 1773–1780.

Sieburth, L.E. (1997). Molecular Dissection of the AGAMOUS Control Region Shows That cis Elements for Spatial Regulation Are Located Intragenically. Plant Cell Online **9**: 355–365.

Sievers, F. and Higgins, D.G. (2014). Clustal omega, accurate alignment of very large numbers of sequences. Methods Mol. Biol. **1079**: 105–116.

Simossis, V.A. and Heringa, J. (2005). PRALINE: A multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. Nucleic Acids Res. **33**: W289-94.

Smyth, D.R. (2001). Stamens. In eLS (John Wiley & Sons, Ltd).

Smyth, D.R., Bowman, J.L., and Meyerowitz, E.M. (1990). Early flower development in Arabidopsis. Plant Cell **2**: 755–67.

Song, Y., Gao, J., Yang, F., Kua, C.-S., Liu, J., and Cannon, C.H. (2013). Molecular evolutionary analysis of the Alfin-like protein family in Arabidopsis lyrata, Arabidopsis thaliana, and Thellungiella halophila. PLoS One **8**: e66838.

Sorensen, A.-M., Kröber, S., Unte, U.S., Huijser, P., Dekker, K., and Saedler, H. (2003). The Arabidopsis ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. Plant J. **33**: 413–23.

Spector, D.L. and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harb. Perspect. Biol. **3**: 1–12.

Spitz, F. and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. **13**: 613–626.

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in Arabidopsis thaliana. Curr. Opin. Plant Biol. **4**: 447–56.

Suganuma, T. and Workman, J.L. (2011). Signals and combinatorial functions of histone modifications. Annu. Rev. Biochem. **80**: 473–99.

Sun, Y.-J., Hord, C.L., Chen, C.-B., and Ma, H. (2007). Regulation of Arabidopsis Early Anther Development by Putative Cell-Cell Signaling Molecules and Transcriptional Regulators. J. Integr. Plant Biol. **49**: 60–68.

Swift, J. and Coruzzi, G.M. (2017). A matter of time — How transient transcription

factor interactions create dynamic gene regulatory networks. Biochim. Biophys. Acta - Gene Regul. Mech. **1860**: 75–83.

Talbot, M.J. and White, R.G. (2013). Methanol fixation of plant tissue for Scanning Electron Microscopy improves preservation of tissue morphology and dimensions. Plant Methods **9**: 36.

Tester, M. and Langridge, P. (2010). Breeding Technologies to Increase Crop Production in a Changing World. Science (80-.). **327**: 818–822.

Thorlby, G.J., Shlumukov, L., Vizir, I.Y., Yang, C.Y., Mulligan, B.J., and Wilson, Z.A. (1997). Fine-scale molecular genetic (RFLP) and physical mapping of a 8.9 cM region on the top arm of Arabidopsis chromosome 5 encompassing the male sterility gene, ms1. Plant J. **12**: 471–9.

Tiessen, A., Pérez-Rodríguez, P., and Delaye-Arredondo, L.J. (2012). Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes. BMC Res. Notes **5**: 85.

Tsuji, H., Aya, K., Ueguchi-Tanaka, M., Shimada, Y., Nakazono, M., Watanabe, R., Nishizawa, N.K., Gomi, K., Shimada, A., Kitano, H., Ashikari, M., and Matsuoka, M. (2006). GAMYB controls different sets of genes and is differentially regulated by microRNA in aleurone cells and anthers. Plant J. **47**: 427–444.

Ünal, M., Vardar, F., and Aytürk, Ö. (2013). Callose in Plant Sexual Reproduction. In Current Progress in Biological Research, M. Silva-Opps, ed (InTech), p. 385 pages.

Van de Velde, J., Van Bel, M., Van Eechoutte, D., and Vandepoele, K. (2016). A Collection of Conserved Non-Coding Sequences to Study Gene Regulation in Flowering Plants. Plant Physiol. **171**: 1–48.

Van de Velde, J., Heyndrickx, K.S., and Vandepoele, K. (2014). Inference of Transcriptional Networks in Arabidopsis through Conserved Noncoding Sequence Analysis. Plant Cell: 1–18.

Vizcay-Barrena, G. and Wilson, Z.A. (2006). Altered tapetal PCD and pollen wall development in the Arabidopsis ms1 mutant. J Exp Bot **57**: 2709–2717.

Walbot, V. and Egger, R.L. (2016). Pre-Meiotic Anther Development: Cell Fate Specification and Differentiation. Annu. Rev. Plant Biol.: 1–31.

Wallace, S., Chater, C.C., Kamisugi, Y., Cuming, A.C., Wellman, C.H., Beerling, D.J., and Fleming, A.J. (2015). Conservation of Male Sterility 2 function during spore and pollen wall development supports an evolutionarily early recruitment

of a core component in the sporopollenin biosynthetic pathway. New Phytol. **205**: 390–401.

Wallace, S., Fleming, A., Wellman, C.H., and Beerling, D.J. (2011). Evolutionary development of the plant spore and pollen wall. AoB Plants **11**.

Wang, C., Xu, J., Zhang, D., Wilson, Z. a, and Zhang, D. (2010). An effective approach for identification of in vivo protein-DNA binding sites from paired-end ChIP-Seq data. BMC Bioinformatics **11**: 81.

Wang, H., Chung, P.J., Liu, J., Jang, I.-C., Kean, M.J., Xu, J., and Chua, N.-H. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. Genome Res. **24**: 444–53.

Wang, J., Niu, B., Huang, J., Wang, H., Yang, X., Dong, A., Makaroff, C.A., Ma, H., and Wang, Y. (2016). The PHD Finger Protein MMD1/DUET Ensures the Progression of Male Meiotic Chromosome Condensation and Directly Regulates the Expression of the Condensin Gene CAP-D3. Plant Cell **28**: tpc.00040.2016.

Wei, W., Zhang, Y.-Q., Tao, J.-J., Chen, H.-W., Li, Q.-T., Zhang, W.-K., Ma, B., Lin, Q., Zhang, J.-S., and Chen, S.-Y. (2015). The Alfin-like homeodomain finger protein AL5 suppresses multiple negative factors to confer abiotic stress tolerance in Arabidopsis. Plant J. **4**: n/a-n/a.

Wellman, C.H. (2004). Origin, function and development of the spore wall in early land plants. In The Evolution of Plant Physiology, pp. 43–63.

Wijeratne, A.J., Zhang, W., Sun, Y., Liu, W., Albert, R., Zheng, Z., Oppenheimer, D.G., Zhao, D., and Ma, H. (2007). Differential gene expression in Arabidopsis wild-type and mutant anthers: insights into anther cell differentiation and regulatory networks. Plant J. **52**: 14–29.

Wilkins, O., Nahal, H., Foong, J., Provart, N.J., and Campbell, M.M. (2009). Expansion and diversification of the Populus R2R3-MYB family of transcription factors. Plant Physiol. **149**: 981–93.

Willemse, M.T.M. (2008). History and prospects of plant sexual reproduction congresses, the IASPRR and sexual plant reproduction. Sex. Plant Reprod. **21**: 89–97.

Wilson, Z.A. and Zhang, D.-B.B. (2009). From Arabidopsis to rice: pathways in pollen development. J Exp Bot **60**: 1479–1492.

Wilson, Z. a, Morroll, S.M., Dawson, J., Swarup, R., and Tighe, P.J. (2001). The Arabidopsis MALE STERILITY1 (MS1) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. Plant J. **28**: 27–39.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., Wu, C., and Allis, C.D. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature **442**: 86–90.

Xiong, S.-X., Lu, J.-Y., Lou, Y., Teng, X.-D., Gu, J.-N., Zhang, C., Shi, Q.-S., Yang, Z.-N., and Zhu, J. (2016). The transcription factor MS188 and AMS form a complex to activate the expression of CYP703A2 for sporopollenin biosynthesis in Arabidopsis thaliana. Plant J.: 1–11.

Xu, J., Ding, Z., Vizcay-Barrena, G., Shi, J., Liang, W., Yuan, Z., Werck-Reichhart, D., Schreiber, L., Wilson, Z. a, and Zhang, D. (2014a). ABORTED MICROSPORES Acts as a Master Regulator of Pollen Wall Formation in Arabidopsis. Plant Cell **2**: 1–14.

Xu, J., Yang, C., Yuan, Z., Zhang, D., Gondwe, M.Y., Ding, Z., Liang, W., Zhang, D., and Wilson, Z. a (2010). The ABORTED MICROSPORES regulatory network is required for postmeiotic male reproductive development in Arabidopsis thaliana. Plant Cell **22**: 91–107.

Xu, Y., Gan, E.-S., He, Y., and Ito, T. (2013a). Flowering and genome integrity control by a nuclear matrix protein in Arabidopsis. Nucleus **4**: 274–6.

Xu, Y., Iacuone, S., Li, S., and Parish, R.W. (2014b). MYB80 homologues in Arabidopsis, cotton and Brassica: regulation and functional conservation in tapetal and pollen development. BMC Plant Biol. **14**: 278.

Xu, Y., Wang, Y., Stroud, H., Gu, X., Sun, B., Gan, E.-S., Ng, K.-H., Jacobsen, S.E., He, Y., and Ito, T. (2013b). A matrix protein silences transposons and repeats through interaction with retinoblastoma-associated proteins. Curr. Biol. **23**: 345–50.

Yamaguchi, N., Winter, C.M., Wu, M.-F., Kwon, C.S., William, D.A., and Wagner, D. (2014). PROTOCOL: Chromatin Immunoprecipitation from Arabidopsis Tissues. Arab. B. **12**: e0170.

Yang, C., Vizcay-Barrena, G., Conner, K., and Wilson, Z.A. (2007). MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. Plant Cell **19**: 3530–48.

Yang, J., Lee, S., Hang, R., Kim, S.-R., Lee, Y.-S., Cao, X., Amasino, R., and An, G. (2013). OsVIL2 functions with PRC2 to induce flowering by repressing OsLFL1 in rice. Plant J. **73**: 566–78.

Yang, S.-L., Xie, L.-F., Mao, H.-Z., Puah, C.S., Yang, W.-C., Jiang, L., Sundaresan, V., and Ye, D. (2003). Tapetum determinant1 is required for cell specialization in the Arabidopsis anther. Plant Cell **15**: 2792–804.

Yang, W.C., Ye, D., Xu, J., and Sundaresan, V. (1999). The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes Dev. **13**: 2108–17.

Ye, Q., Zhu, W., Li, L., Zhang, S., Yin, Y., Ma, H., and Wang, X. (2010). Brassinosteroids control male fertility by regulating the expression of key genes involved in Arabidopsis anther and pollen development. Proc. Natl. Acad. Sci. U. S. A. **107**: 6100–5.

Yonekura-Sakakibara, K., Nakabayashi, R., Sugawara, S., Tohge, T., Ito, T., Koyanagi, M., Kitajima, M., Takayama, H., and Saito, K. (2014). A flavonoid 3-O-glucoside:2''-O-glucosyltransferase responsible for terminal modification of pollen-specific flavonols in Arabidopsis thaliana. Plant J. **79**: 769–782.

Yu, S. (2015). Identification and characterisation of MS1 putative interacting proteins and regulatory targets in Arabidopsis.

Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell Res. **21**: 564–78.

Zemach, A., Li, Y., Ben-Meir, H., Oliva, M., Mosquna, A., Kiss, V., Avivi, Y., Ohad, N., and Grafi, G. (2006). Different domains control the localization and mobility of LIKE HETEROCHROMATIN PROTEIN1 in Arabidopsis nuclei. Plant Cell **18**: 133–45.

Zhang, D., Liang, W., Yin, C., Zong, J., Gu, F., and Zhang, D. (2010). OsC6, encoding a lipid transfer protein, is required for postmeiotic anther development in rice. Plant Physiol. **154**: 149–162.

Zhang, D., Liu, D., Lv, X., Wang, Y., Xun, Z., Liu, Z., Li, F., and Lu, H. (2014). The Cysteine Protease CEP1, a Key Executor Involved in Tapetal Programmed Cell Death, Regulates Pollen Development in Arabidopsis. Plant Cell **12**: 1–24.

Zhang, D. and Yang, L. (2014). Specification of tapetum and microsporocyte cells within the anther. Curr. Opin. Plant Biol. **17**: 49–55.

Zhang, D.B. and Wilson, Z.A. (2009). Stamen specification and anther development in rice. Chinese Sci. Bull. **54**: 2342–2353.

Zhang, W., Sun, Y., Timofejeva, L., Chen, C., Grossniklaus, U., and Ma, H. (2006a). Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. Development **133**: 3085–95.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V, Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007a). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol. **5**: e129.

Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., and Chua, N.-H. (2006b). Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat. Protoc. **1**: 641–646.

Zhang, Z.B., Zhu, J., Gao, J.F., Wang, C., Li, H., Zhang, H.Q., Zhang, S., Wang, D.M., Wang, Q.X., Huang, H., Xia, H.J., and Yang, Z.N. (2007b). Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis. Plant J **52**: 528–538.

Zhao, S., Fung-Leung, W.P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One **9**: e78644.

Zhu, E., You, C., Wang, S., Cui, J., Niu, B., Wang, Y., Qi, J., Ma, H., and Chang, F. (2015a). The DYT1-interacting proteins bHLH010, bHLH089 and bHLH091 are redundantly required for Arabidopsis anther development and transcriptome. Plant J. **83**: 976–990.

Zhu, J., Chen, H., Li, H., Gao, J.-F., Jiang, H., Wang, C., Guan, Y.-F., and Yang, Z.-N. (2008). Defective in Tapetal development and function 1 is essential for anther development and tapetal function for microspore maturation in Arabidopsis. Plant J. **55**: 266–77.

Zhu, J., Lou, Y., Xu, X., and Yang, Z.-N. (2011). A genetic pathway for tapetum development and function in Arabidopsis. J. Integr. Plant Biol. **53**: 892–900.

Zhu, Q., Ramm, K., Shivakkumar, R., Dennis, E.S., and Upadhyaya, N.M. (2004). The ANTHER INDEHISCENCE1 gene encoding a single MYB domain protein is involved in anther development in rice. Plant Physiol. **135**: 1514–25.

Zhu, X., Liang, W., Cui, X., Chen, M., Yin, C., Luo, Z., Zhu, J., Lucas, W.J., Wang, Z., and Zhang, D. (2015b). Brassinosteroids Promote Development of Rice Pollen Grains and Seeds by Triggering Expression of CSA, a MYB Domain Protein. Plant J.: n/a-n/a.