IDENTIFICATION AND CHARACTERISATION OF MS1 PUTATIVE INTERACTING PROTEINS AND REGULATORY TARGETS IN *ARABIDOPSIS*

Suyang Yu, BSc

Thesis submitted to the
University of Nottingham
for the degree of Doctor of Philosophy

DECEMBER 2014

ABSTRACT

The *Arabidopsis thaliana MALE STERILITY1 (MS1*) gene, encodes a plant homeodomain (PHD) transcription factor critical for viable pollen formation (Wilson et al., 2001). In the *ms1* mutant, there are alterations in the production of pollen wall materials, as well as a failure of tapetal programmed cell death (PCD) (Vizcay-Barrena and Wilson, 2006). This ultimately results in the failure to produce viable pollen. Large numbers of genes are down-regulated in the *ms1* mutant indicating that MS1 plays a key role in regulating late tapetal expression and pollen wall deposition (Ito et al., 2007; Yang et al., 2007).

Two putative MS1 interacting proteins At1g58210/NET2A (termed as Y2H54) and AT2G46260/ LRB1 (termed as POB2) were identified from a previous *Arabidopsis* stamen–specific yeast-2-hybrid screen, using a truncated version of the MS1 protein without the PHD motif. POB2 and MS1 were found co-localised in the nucleus, while Y2H54 was specifically located at the plasma membrane. Further confirmation of the interaction using Förster resonance energy transfer (FRET) assay methods showed that POB2 failed to interact with MS1 *in planta*, however, the association between the two proteins occurred *in vitro*, as confirmed by protein pull-down assays. Additionally, enhanced general plant growth and floral development were seen in the overexpression lines of Y2H54. However, no significant phenotypes were observed in the RNAi silencing lines.

Chromatin Immunoprecipitation (ChIP) analysis uncovered that MS1 directly regulated the expression of MYB DOMAIN PROTEIN 99 (MYB99) by binding to its promoter. Other putative MS1 direct targets identified by ChIP include 3-KETOACYL-COA SYNTHASE 7 (KCS7), 3-KETOACYL-COA SYNTHASE 15 (KCS15), SPERMIDINE HYDROXYCINNAMOYL TRANSFERASE (SHT) and TAPETUM-SPECIFIC METHYLTRANSFERASE 1 (TSM1). Histone extraction and western blotting assays suggest a role for MS1 in facilitating detrimethylation of H3 marks. H3K36me3 deposition was enhanced at *MYB99* in *ms1* compared with the wild type, suggesting that MS1 may regulate MYB99 via H3K36me3. A new model for the MS1 regulatory network in pollen wall formation has therefore been proposed.

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my supervisor Prof. Zoe A. Wilson for the continually guidance and support throughout my PhD study. Her endless knowledge, dedication and patience have been greatly inspiring me during the years.

I would like to thank my colleagues Dr. C. Yang, Dr. A. Ferguson, Dr. S. Pierce, D. Klisch, and K. Simpson and others for their help in the project. Also thanks to all the members of Wilson's group who contribute to create a wonderful working environment. I'm grateful to my assessor Prof. M. Holdsworth for giving me valuable advices on the research. Also thanks to China Scholarship Council for the financial supports.

I specially would like to thank my boyfriend Di, in recognition of all his patience and persistence of waiting through the entire duration of my study, though we have been far apart in two countries. I am deeply grateful to my parents, and all my friends in the UK and elsewhere around the world. 我爱你们!

ABBREVIATIONS

ABC: ATP Binding Cassette

AD: Activation Domain

AG: Affymetrix Gene chip

ATP: Adenosine Tri-Phosphate

bHLH-ZIP: basic-Helix–Loop–Helix Leucine -Zipper domain

BLAST: Basic Local Alignment Search

bp: base pair

BTB: BR-C, ttk and bab

bZIP: basic Zipper

$CaCl_2$: Calcium Chloride

cDNA: complementary Deoxyribonucleic Acid

ChIP: Chromatin Immunoprecipitation

cm: centimetre

DB: DNA-Binding

DEPC: diethylpyrocarbonate

dH_2O: distilled water

DNA: Deoxyribonucleic Acid

DNase: deoxyribonuclease

dNTP: Deoxynucleotide Triphosphate

dsDNA : double-stranded DNA

DTT: Dithiothreitol

dUTP: Deoxyuridine Triphosphate

EDTA: Ethylenediaminetetraacetic Acid

FRET: Förster Resonance Energy Transfer (FRET)

g: gram

GFP: Green Fluorescent Protein

g/l: gram per litre

GST: Glutathione-S-transferase

GUS: β-Glucuronidase

h: hour

HAT: Histone Acetyl Transferase

HCl: Hydrochloric Acid

HM: Homozygous

HT: Heterozygous

H3K4: Histone H3 Lysine 4

kb: kilo base pair

kDa: kilo Dalton

l: litre

LB: Luria Broth

me2: dimethylation

me3: trimethylation

mg/l: miligram per litre

min: minute

ml: millilitre

mM: milimolar

mRNA: messenger Ribonucleic Acid

MS: Murashige and Skoog Basal Medium

NaCl: Sodium Chloride

NASC: Nottingham *Arabidopsis* Stock Centre

ng: nanogram

ng/µl: nanogram per microlitre

PBS: Phosphate Buffered Saline

PCD: Programmed Cell Death

PCR: Polymerase Chain Reaction

PHD: Plant Homeodomain

POZ: Poxvirus and Zinc finger

RNA: Ribonucleic Acid

RNAi: Ribonucleic Acid Interference

rpm: revolutions per minute

PCR: Polymerase Chain Reaction

RT: Room Temperature

RT-PCR: Reverse Transcriptase-Polymerase Chain Reaction

SAIL: Syngenta *Arabidopsis* Insertion Library

sec: second

SDS: Sodium Dodecyl Sulfate

SDS-PAGE: Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

TAIR: The *Arabidopsis* Information Resource

T-DNA: Transferred-Deoxyribonucleic Acid

µg: microgram

µg/ml: microgram per millilitre

µl: microlitre

µM: micromolar

UTR: Untranslated Regions

UV: Ultraviolet

v/v: volume to volume ratio

Wt: Wild type

w/v: weight to volume radio

Y2H: Yeast-2-Hybrid

YFP: Yellow Fluorescent Protein

TABLE OF CONTENT

CHAPTER 1 INTRODUCTION

Arabidopsis thaliana is a flowering plant with a rapid life cycle, small morphological size, and prolific seed production, which make it widely used as a model species for genetics, and cellular, molecular biology in higher plant (Meinke et al., 1998). It contains a genome organized into five chromosomes comprising an estimated 20,000 genes (Meinke et al., 1998). In 2000, *Arabidopsis thaliana* became the first plant species to be genome-wide sequenced, making possible the identification of more than 30,000 genes and providing a powerful tool for gene mapping and functional analysis (*Arabidopsis* Genome Initiative, 2000).

Male sterility in flowering plants, a phenomenon first observed by Kölreuter in 1763, has valuable significance in selective breeding, by greatly facilitating the production of hybrids via cross-pollination (Kaul, 1988). In *Arabidopsis thaliana,* molecular and genetic studies have identified numerous genes regulating stamen and pollen development, which are crucial for normal male reproductive development (Ma, 2005). One example is the *Arabidopsis MALE STERILITY1 (MS1)* gene (Wilson et al., 2001), which is a PHD finger motif transcriptional factor that plays a key role in tapetum development and pollen wall formation; it is also conserved in rice and barley (Li et al., 2011; Gómez and Wilson, 2012).

By investigating putative MS1 interacting proteins and regulatory targets, this PhD study is focused on the molecular mechanism of how MS1 acts as a key regulator in the anther and pollen development regulatory network.

1.1 MALE MICROSPOROGENESIS AND MICROGAMETOGENESIS IN *ARABIDOPSIS*

1.1.1 Stamen Development

Arabidopsis flowers consist of four types of organ, namely sepals, petals, stamens, and the pistil (Figure 1.1) (Ma, 2005). There are six stamens in each flower, which are the male reproductive organs of plants. Each of them consists of an anther with four lobes, the site where pollen develops, as well as a stalk-like filament, functioning in supporting and transporting water and nutrients to the anther.

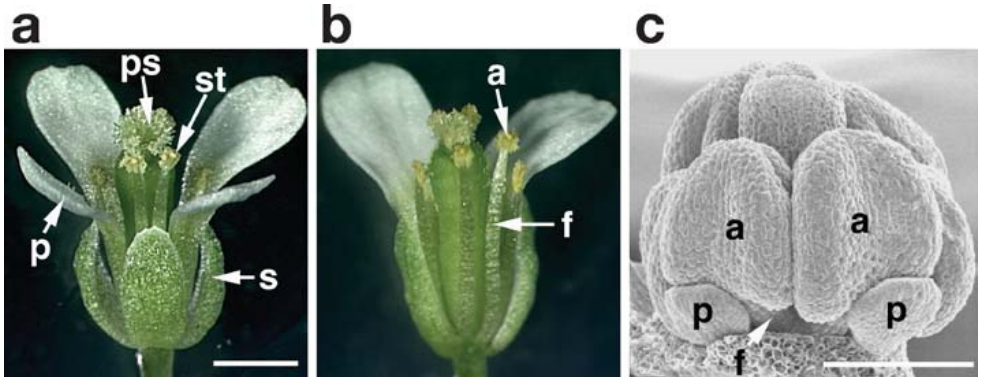


Figure 1.1. Structure of *Arabidopsis thaliana* flower (Ma, 2005). (a) An intact mature (stage 13) *Arabidopsis* flower with four types of organs, sepals (s), petals (p), stamens (st), and the pistil. (b) A mature flower with one sepal and two petals removed to reveal some of the stamens, which have an anther (a) and a filament (f). (a) and (b) are of the same magnification; bar = 1.0 nm. (c) A scanning electron micrograph of a stage 9 flower, with its sepals removed to show the inner organs; two petal primordia (p) are round and small; two of the four long stamens can be easily seen with their anthers (a) having attained the characteristic lobed shape and the filaments (f) still very short. Bar = 100 μm.

Male reproductive development in *Arabidopsis thaliana* begins in the sporophytic generation with the initiation and formation of the stamens.

Flower development has been divided into 12 stages based on observations from scanning electron microscopy (Table 1.1) (Smyth et al., 1990). Six stamen primordia are initiated at flower stage 5, while cell division and differentiation occur from stages 6 to 9. By stage 7, the four long stamens form an anther at the top and a stalk at the base, which are slightly more advanced than the two short ones. At flower stage 8, the anthers of the long stamens develop four lobes where meiosis will take place, followed by pollen development at stages 10 through 12. Subsequently, anther dehiscence occurs as the flower opens.

Table 1.1 Summary of Landmark Events of Flower Development in *Arabidopsis thaliania* (Smyth et al., 1990).

Stage	Landmark Event	Duration (h)[a]	Age of Flower at End of Stage (days)
1	Flower buttress arises	24	1
2	Flower primordium forms	30	2.25
3	Sepal primordial arise	18	3
4	Sepals overlie flower meristem	18	3.75
5	Petal and stamen primordial arise	6	4
6	Sepals enclose bud	30	5.25
7	Long stamen primordial stalked at base	24	6.25
8	Locules appear in long stamens	24	7.25
9	Petal primordial stalked at base	60	9.75
10	Petals level with short stamens	12	10.25
11	Stigmatic papillae appear	30	11.5
12	Petal level with long stamens	42	13.25

[a] Estimated to nearest 6 hr.

1.1.2 Anther Development

Plant anther development involves two phases of morphological, cellular,

and molecular events (Figure 1.2) (Koltunow et al., 1990; Goldberg et al., 1993; Ma, 2005). During phase 1, the morphology of the anther is established as the result of cell division and differentiation. This process starts with initiation of the stamen primordia on the floral meristem, followed by cell specification and tissue differentiation into the different anther layers, and ends with the meiosis of pollen mother cells (PMC), resulting in tetrads of microspores. During phase 2, the filament elongates greatly, accompanied by the process of the anther further enlarging and the microspores developing into pollen grains. Subsequently, anther dehiscence and pollen release are completed as the degeneration of anther tissues occurs at the end of this phase.

In *Arabidopsis thaliana*, the floral meristem comprise of three cell layers with separate lineages: L1 (epidermis), L2 (subepidermis) and L3 (core). All male reproductive cells of *Arabidopsis thaliana* develop within concentrically organized microsporangia derived from archesporial cells in the L2 of stamen primordial. These archesporial cells divide periclinally giving rise to an inner primary sporogenous cell and an outer primary parietal cell (PP). Primary sporogenous cell division occurs to form a central sporogenous mass, meanwhile, consecutive divisions of neighbouring cells give rise to the formation of three concentric parietal layers– the tapetum adjacent to the sporogenous cells, the middle layer, and the endothecium subjacent to the epidermis (Figure 1.3) (Canales et al., 2002; Sorensen et al., 2003; Scott et al., 2004; Wilson and Zhang, 2009; Feng and Dickinson, 2010; Wilson et al., 2011).

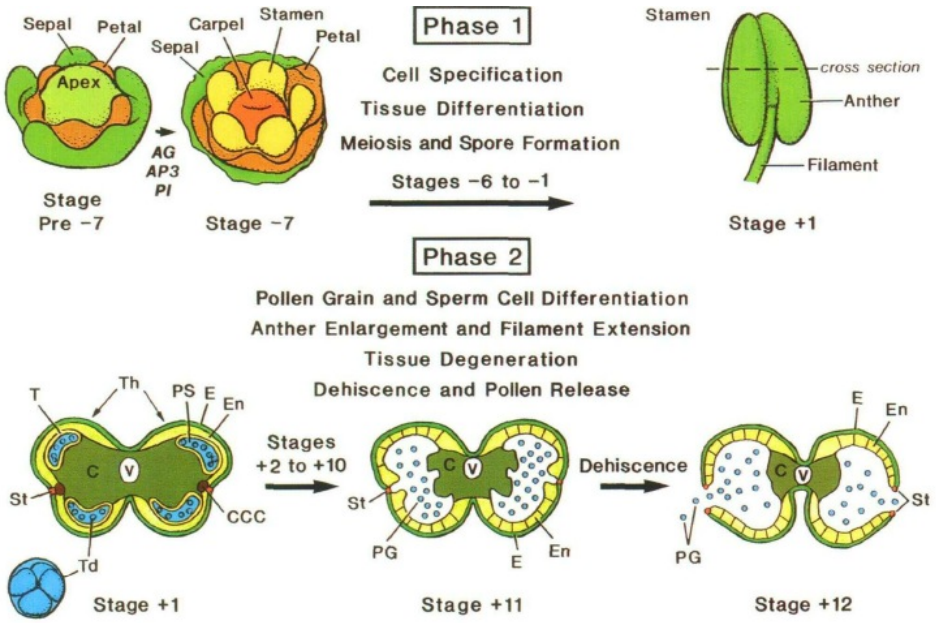


Figure 1.2 A generalized overview of anther development (Goldberg et al., 1993). Schematic representations of anther developmental stages and cross-sections based on scanning electron and light microscopy studies of tobacco anther development (Koltunow et al., 1990; Drews et al., 1992). The dashed line through the stage 1 anther drawn in the phase 1 portion of the figure represents the cross-section plane for anthers drawn schematically in phase 2. The vertical lines drawn through the endothecium at stages 11 and 12 represent fibrous cell wall bands. C, connective; CCC, circular cell cluster; E, epidermis; En, endothecium; PG, pollen grain; PS, pollen sac; St, stomium; T, tapetum; Td, tetrads; Th, theca; V, vascular bundle.

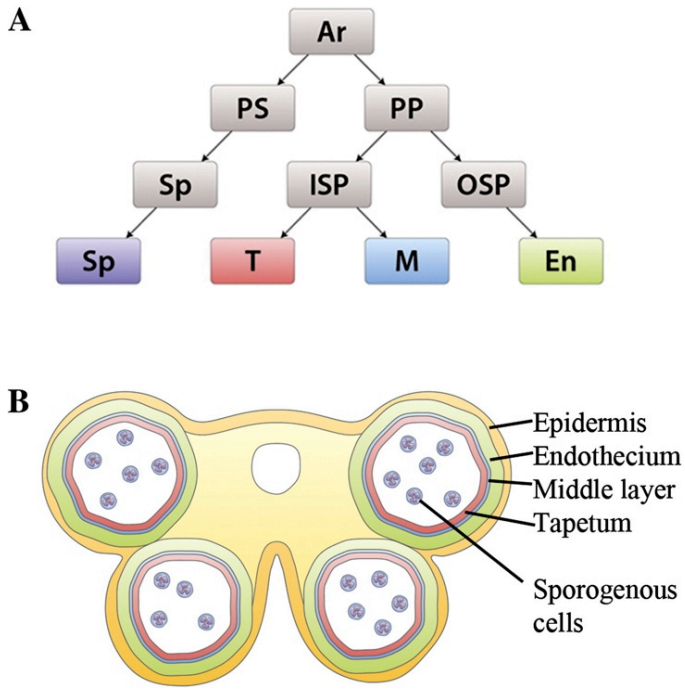


Figure 1.3 Diagram of anther lineage and structure (Wilson et al., 2011). (A) A widely accepted cell lineage model has been suggested for the origin of the cell layers in the anther (Scott et al., 2004). Four clusters of archesporial cells (Ar) in the anthers divide to form the primary parietal layer (PP) and the primary sporogenous layer. The PP layer then goes through a further division to form two secondary parietal layers, the inner secondary parietal layer (ISP) and the outer secondary parietal layer (OSP). The OSP then divides again and differentiates to form the endothecium layer (En), whereas the ISP divides and develops to form the tapetum (T) and middle cell layer (M). (B) This results in the four cell layers of the anther: E, the outer epidermis (yellow); En, endothecium (green); M, middle cell layer (blue); T, tapetum (red); Sp, the inner sporogenous cells (purple).

Arabidopsis thaliana anther development has been divided into 14 stages using morphological and cellular features from light microscopy (Sanders et al., 1999). Stages 1 to 8 constituting the aforementioned phase 1, involve the events leading to unicellular microspores (Figure 1.4). And phase 2, comprises stages 9 to 14, during which unicellular microspores differentiate into three-celled pollen grains, accompanied by the process of tapetum generation (Figure 1.5). Eventually, anther development is complete with

the anther dehiscence to release mature pollen grains (Sanders et al., 1999).

The anther stages 1 through 4 correspond approximately to flower stages 5 through 8 as defined by Smyth et al. (1990). During these stages, cell division events occur within the developing anther primordia to establish a bilateral structure with locule, wall, connective, and vascular region characteristics of the mature anther. By stage 5, the anther morphogenesis has been completed with a characteristic four-lobed morphology. In each lobe, division of the primary sporogenous layer gives rise to the pollen mother cells (PMCs), also known as microspore mother cells (MMCs), which are surrounded by four non-reproductive cell layers, respectively, the epidermis, the endothecium, the middle layer, and the tapetum in closest contact to the developing pollen (Figure 1.4). This has been considered as the start of pollen development (Smyth et al., 1990).

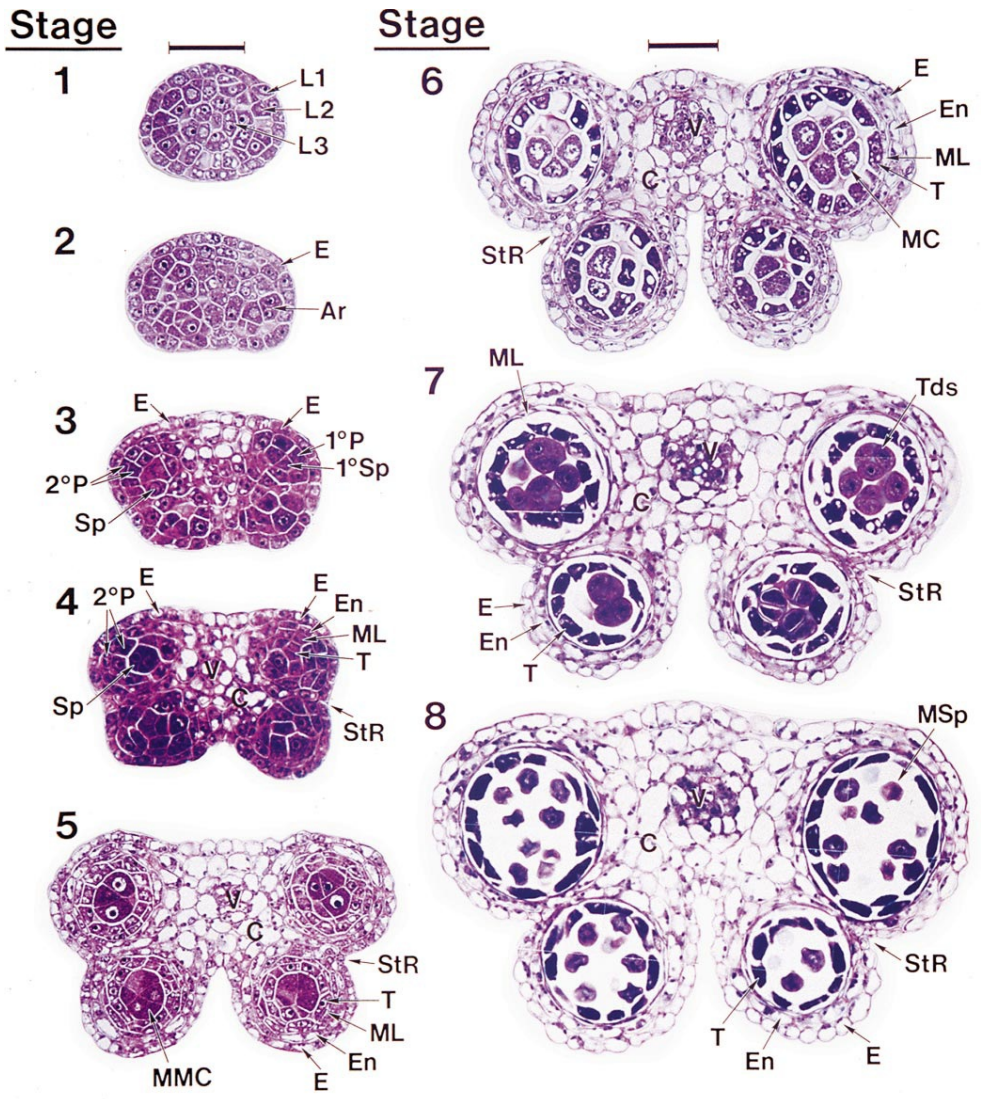


Figure 1.4 Phase one of wild-type *Arabidopsis thaliana* anther development (Sanders et al., 1999). Anther sections stained with toluidine blue. Ar, archesporial cell; C, connective; E, epidermis; En, endothecium; L1, L2, and L3, the three cell-layers in stamen primordia; MC, meiotic cell; ML, middle 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. Bar over stage 1=25 µm and this is the scale for stages 1 to 4. Bar over stage 6=25 µm and this is the scale for stages 5 to 8.

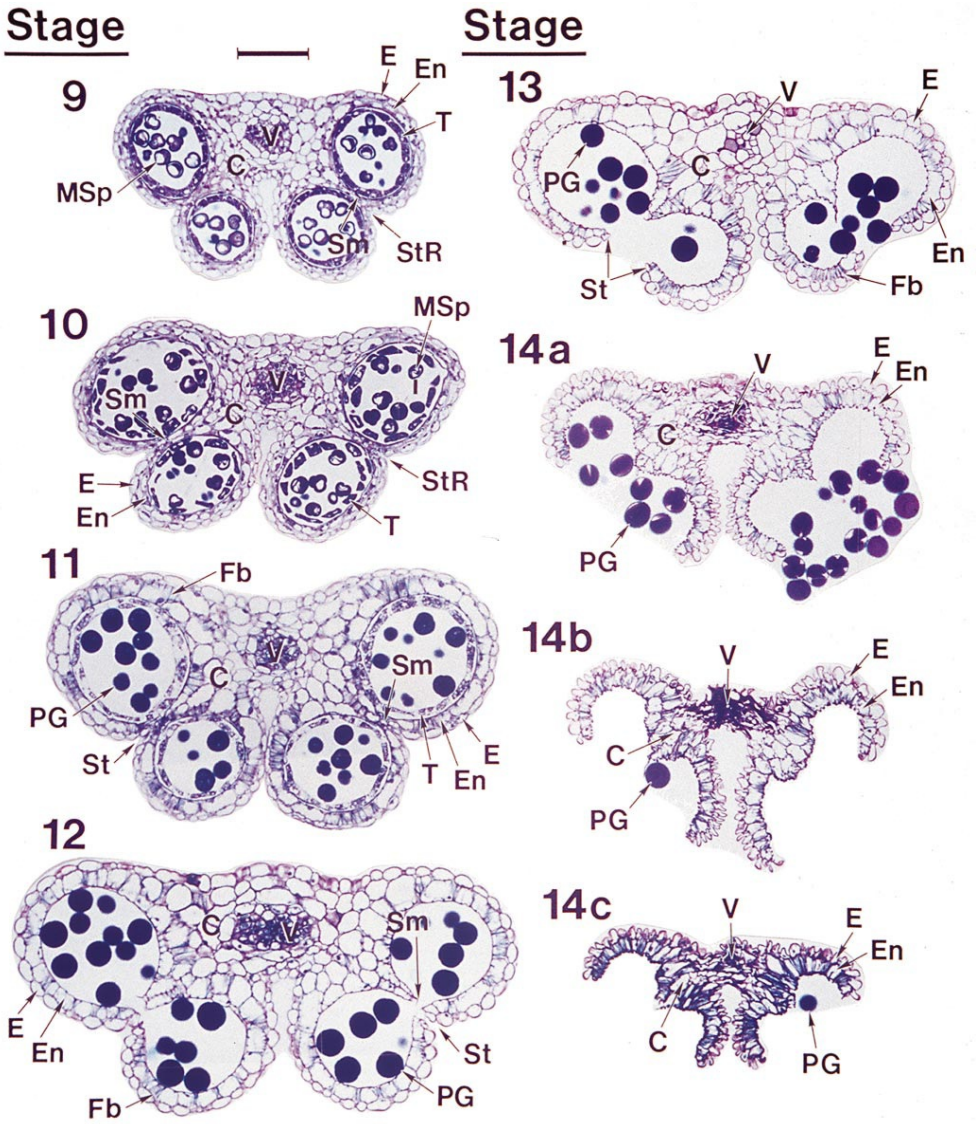


Figure 1.5 Phase two of wild-type *Arabidopsis thaliana* anther development (Sanders et al., 1999). Anther sections stained with toluidine blue. Stages 9 to 11, 12 to 13, and 14a to 14c represent anther late development, dehiscence, and senescence, respectively. C, connective; E, epidermis; En, endothecium; Fb, fibrous bands; MSp, microspores; PG, pollen grains; Sm, septum; St, stomium; StR, stomium region; T, tapetum; V, vascular region. Bar=50 μm and applies to stages 9–14c.

PMCs in each anther lobe are formed by stage 5 then proceed through meiotic divisions until anther stage 7, generating tetrads of haploid microspores. Prior to the start of meiosis, the microsporocytes synthesize a transient callose (β-1,3-glucan) wall that continues to develop during

meiosis to give the classic 'tetrad' appearance and then entirely covers and compartmentalizes individual microspores. Microsporogenesis is complete with the formation of individual microspores, which are released from the tetrads as callose breakdown occurs at anther stage 8 (Figure 1.4). During the following stages, each unicellular microspore is polarized and differentiated into a tricellular pollen grain through two rounds of mitotic division. The first round gives rise to the vegetative cell and a smaller generative cell, the latter of which undergoes cell division to generate two sperm cells (Figure 1.6) (Borg et al., 2009).

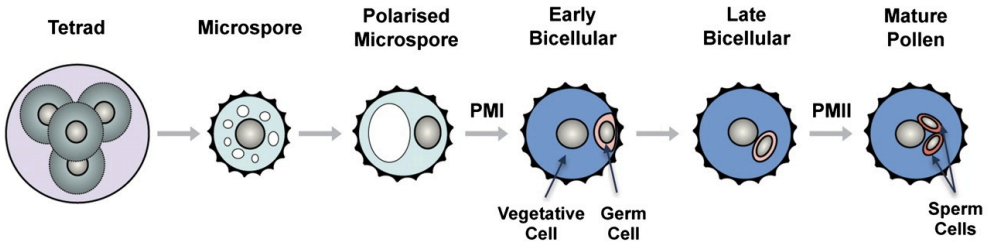


Figure 1.6 Male gametophyte development in *Arabidopsis thaliana* (Borg et al., 2009). Schematic diagram representing the distinct morphological stages of male gametophyte development in *Arabidopsis thaliana*. During microsporogenesis, microsporocytes undergo a meiotic division to produce a tetrad of four haploid microspores. During microgametogenesis, the released microspores undergo a highly asymmetric division, called Pollen Mitosis I (PMI), to produce a bicellular pollen grain with a small germ cell engulfed within the cytoplasm of a large vegetative cell. Whilst the vegetative cell exits the cell cycle, the germ cell undergoes a further mitotic division at Pollen Mitosis II (PMII) to produce twin sperm cells.

A summary of the key events occurring at each stage are listed (Table 1.2), a cross-reference between the anther stages and those described for *Arabidopsis thaliana* flower and pollen development (Regan and Moffatt 1990; Smyth et al. 1990; Bowman et al. 1991).

T**able 1.2 Summary of the Anther Development Stages of *Arabidopsis thaliana* (Sanders et al., 1999).**

Stage	Major events and morphological changes	Flower stage	Pollen stage
1	Rounded stamen primordial emerges.	5	
2	Archesporial cells arise in the four corners of the anther primordial. Primordial become oval shaped.		
3	Mitotic activity in archesporial cells generating primary parietal and sporogenous layer that will divide further on.	7	1 and 2
4	Four-lobed anther pattern with two developing stomium regions. Vascular region initiated.	8	
5	Four defined locule established. All anther cell types present and pattern of anther defined. Pollen mother cells (PMC) appear.	9	3
6	PMCs enter meiosis. Middle layer is crushed and degenerates. Tapetum becomes vacuolated and anther undergoes a general size increase.		
7	Meiosis completed. Tetrads of microspore free within each locule. Remnants of middle layer present.		4
8	Callose wall surrounding tetrads degenerates and individual microspore released.	10	5
9	Growth and expansion of the anther continue. Microspores generate an extine wall and become vacuolated.		6 and 7
10	Tapetum degeneration initiated	11	
11	Pollen mitotic divisions occur. Tapetum degenerates. Expansion of the endothecial layer and secondary thickenings appear in endothecium and connective cells. Septum degeneration starts. Stomium differentiation begins.	11 and 12	8 and 9
12	Tricellular pollen. Anther becomes binocular after breakage of septum. Stomium differentiates.		10
13	Dehiscence. Breakage along stomium and pollen release.	13 and 14	
14	Senescence of stamen. Shrinkage of cells and anther structure.	15 and 16	
15	Stamen falls off senescing flower.	17	

1.1.3 Tapetal Cell Development

The *Arabidopsis thaliana* tapetum, the innermost layer of the four anther layers, adjacent to the sporogenous cells, plays a significant role in normal pollen development (Pacini et al., 1985; Piffanelli et al., 1998). Normal tapetal cell fate can be described as three phases: tapetal cell differentiation, tapetum secretory cell formation and tapetal programed cell death (PCD).

Formation of the tapetum, occurs together with other sporogenous cells, initials within the L2 (subepidermis) layer of stamen primordial. These archesporial cells go through rounds of further cell division to successively produce primary parietal layerx, inner primary parietal layer and finally the tapetum. By anther stage 5, distinctive tapetum structures have been formed with a number of changes taking place afterwards: general shrinkage of the whole cell and the nuclei; condensation of main components of cytoplasm into densely staining spherical bodies; nucleus mitosis to form binuclear secretory cells; abundance of ribosomes, Golgi bodies and endoplasmic reticulum, all of which indicate an active cell metabolism (Figure 1.7 B) (Stevens and Murray, 1981; Papini et al., 1999). These changes provide the foundation of its secretory role during pollen development.

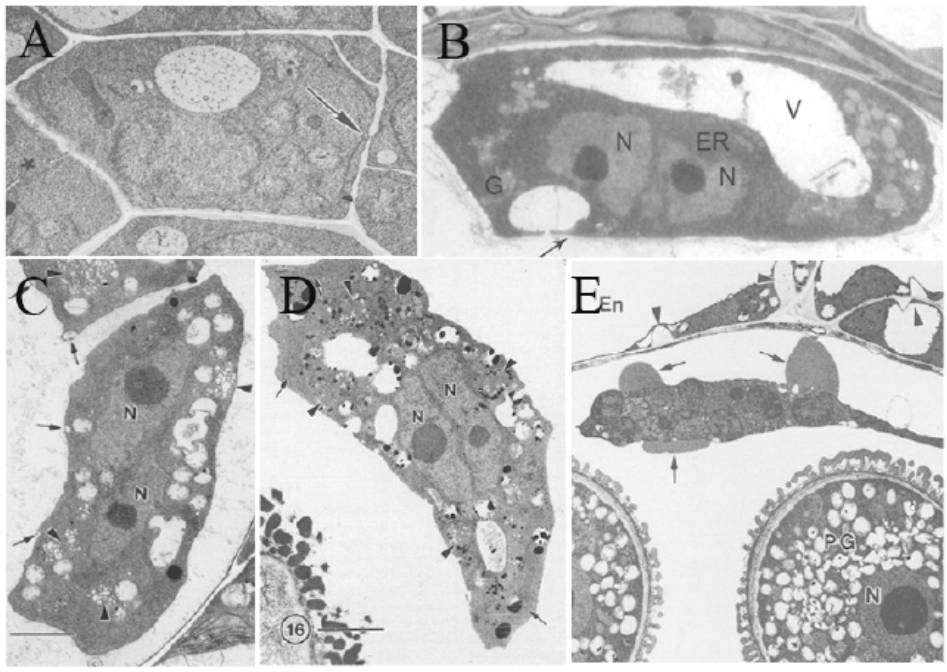


Figure 1.7 Ultrastructure of tapetum development in *Arabidopsis thaliana* using TEM (Owen and Makaroff, 1995). (A) Tapetum at anther stage 5, the arrows indicate plasmodesma. (B) Tapetum at anther stage 7. (C) Tapetum at anther stage 9, the arrows indicate small vacuoles appearing on the tapetum surface. (D) Tapetum at anther stage 10. (E) Tapetum at anther stage 11. N, nucleus; En, endothecial cells; V, vacuole; ER, endoplasmic reticulum; G, Golgi bodies; PG, pollen grains.

As microspore release from the tetrad, the tapetal cells syntheses and secrete callose degrading glucanases to catalyse the breakdown of callose wall that surrounds the tetrad (Stieglitz, 1977). The timing of callose breakdown also appears to be critical, as pollen viability is significantly affected by premature dissolution of the callose wall (Frankel et al., 1969; Worrall et al., 1992; Scott et al., 2004; Wilson and Zhang, 2009) Afterwards, vesicles containing the precursors of sporopollenin, the main component of pollen exine, fuse to the plasma membrane to release their contents into the anther locule for construction of the pollen wall (Figure 1.7 C, D) (Wilson and

Zhang, 2009) .

Immediately after microspore release from the tetrad, tapetal cells start to go through degeneration, which has been considered as programmed cell death (PCD) (Parish and Li, 2010). During this process, the tapetal cells produce tryphine and pollen kit, which are subsequently released into the locule for completion of coating the mature pollen grains (Wilson and Zhang, 2009).

Normal tapetum function is critical for viable pollen production. Male sterility is often observed when tapetal cell function is altered. Selectively destroying the tapetal cell layer by expression the chimaeric ribonuclease gene within the anther could result in aborted pollen formation, and eventually lead to male sterility (Mariani et al., 1990).

1.1.4 Pollen Wall Formation

As one of the most distinctive features of the pollen grain, the pollen wall consists of two different layers: the exine towards the exterior and the intine towards the interior (Figure 1.8) (Owen and Makaroff, 1995). Two sublayers constitute the exine, namely the innermost featureless bilayer nexine comprising nexine I (foot layer) and nexine II (endexine); and the outer sculpted sexine which displays multiple pores and grooves, comprising an outer roof-like tectum; and a central column-like segment formed by baculae (Vizcay-Barrena and Wilson, 2006). The exine wall is ornamented in a highly species specific pattern, providing a diagnostic tool for taxonomists and paleobotanists (Shukla et al., 1998).

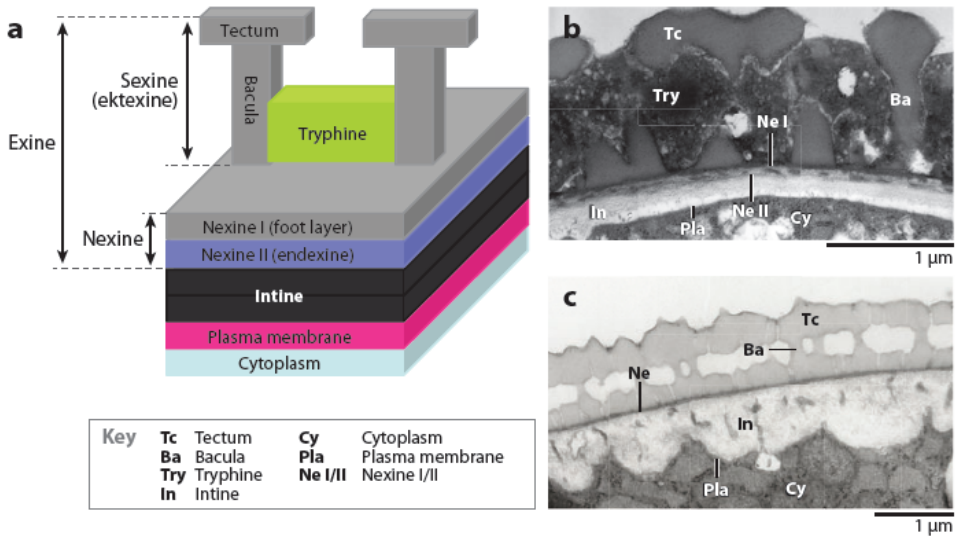


Figure 1.8 Structure of a typical angiosperm pollen grain (Ariizumi and Toriyama, 2011)**.** (a) Schematic of pollen wall. (b) Transmission electron micrographs of a cross-section of exine architecture in *Arabidopsis thaliana*. (c) Mature rice pollen grains. This diagram represents a cross-section from a nonaperture area of pollen grains.

Composition of the intine secreted by the microspore (gametophytic origin) is similar to the known primary walls of common plant cells, including hydrolytic enzymes, hydrophobic proteins, cellulose, hemicellulose, and pectic polymers (Knox and Heslop-Harrison, 1971; Owen and Makaroff, 1995).

Initiation of exine formation and the time point at which normal exine structure appear are often illustrated at the tetrad stage, when the microsporocyte finishes meiosis. Meanwhile, deposition of the primexine as an electron-dense layer composed of a microfibrillar material occurs between the callose wall and the undulating plasma membrane during the tetrad stage. Evidence indicates that the primexine acts as a scaffold, matrix, or template for initial sporopollenin accumulation by its chemical and

selective binding capacities (Scott, 1994; Piffanelli et al., 1998; Scott et al., 2004; Blackmore et al., 2007; Gabarayeva et al., 2009; Wilson and Zhang, 2009; Feng and Dickinson, 2010). Whereas, the callose layer is thought to provide a surface against which the tectum can form, considering correct columellae positioning still occurs in the absence of the callose wall in *Brassica napus* (Scott et al., 2004).

However, the exact components of the exine sporopollenin secreted by the tapetum (sporophytic origin) are still not fully known. Evidence suggests that it is not a homogeneous macromolecule, but rather made up of complex biopolymers derived mainly from saturated lipid precursors such as long-chain fatty acids or long aliphatic chains (Scott, 1994; Bubert et al., 2002). Phenol metabolism has also been found to be involved in the synthesis of these biopolymers (Osthoff and Wiermann, 1987; Koltunow et al., 1990; Ahlers et al., 1999). Sporopollenin then polymerizes onto anchoring points provided by the primexine, a microfibillar polysaccharide matrix, which establishes the basis of the divergent morphological structure of pollen wall (Piffanelli et al., 1998; Scott et al., 2004). As a robust material widely present in algae, fungi, moss, and fern spore walls, sporopollenin is extremely biochemically resistant and considered to be one of the toughest and most durable materials in nature, serving as a protective barrier against excessive dehydration, and fungal and bacterial attack (Ma, 2005).

The gaps between baculae are filled with the pollen coat (sporophytic

origin), which consists of two different coat materials originating from tapetal cell lipids, the pollenkitt and the tryphine (Owen and Makaroff, 1995). The pollen coat allows pollen grains to stick to pollinator vectors or to the dry surface of the stigmas. It also carries proteins involved in self-incompatibility responses (Piffanelli et al., 1998).

1.1.5 Pollen development and male sterility

Pollen wall development is a tightly regulated process that is critical for viable pollen development (Figure 1.9) (Jiang et al., 2013). Controlled pollen wall formation involves callose wall development, lipid metabolism, and sporopollenin deposition. Alterations in these processes have been frequently linked to reduced fertility (Ma, 2005; Blackmore et al., 2007; Jiang et al., 2013).

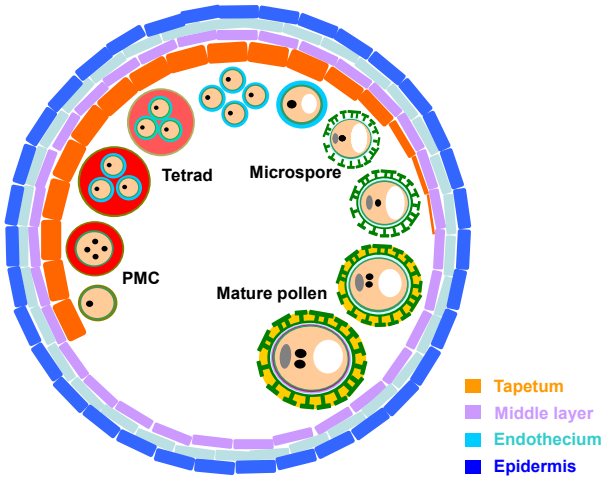


Figure 1.9 Schematic model of pollen wall development in *Arabidopsis thaliana* (Jiang et al., 2013). The anther consists of four distinct cell layers: epidermis, endothecium, middle layer and tapetum. Microspores are produced in the locule surrounded by the tapetum. Pollen wall development is initiated at the tetrad stage when the microsporocyte undergoes meiosis to produce four microspores, which are entirely covered with the microsporocyte-produced callose wall, and the pollen mother cell primary cell wall. The surface of the microspore plasma membrane is smooth. Wavy undulations, which are necessary for exine formation, are subsequently observed on the plasma membrane surface. The primexine, which acts as the sporopollenin receptor, is produced around each microspore. Deposition of sporopollenin precursors builds up the proexine that contains the basis of the bacula (probacula) and tectum (protectum) on the primexine. Then, at released microspore I or early unicleate microspore stage, the callose wall becomes soluble and the thickness of the exine increases along with deposition of polymerized sporopollenin derived from the tapetum. The probacula grows and elongates. Undulations of the plasma membrane gradually disappear. With additional polymerization of sporopollenin, the mature exine structure is visually complete by the bicellular pollen stage. The tryphine, a tapetum remnant, fills the inter space between the tectum and the foot layers after the second round of mitosis, which generates the tricelluar pollen. The intine begins to develop and is completed at the late binucleate stage. The integral pollen wall subsequently forms.

1.1.5.1 Callose wall metabolism

After tetrad formation, the callose wall breaks down catalysed by callase

(callose-degrading glucanase) from the tapetum releasing individual microspores into the locules. Both the callose formation and the timing of callose breakdown appears to be critical, as microsporogenesis is significantly affected by premature dissolution of the callose wall, demonstrated in petunia and transgenic tobacco (Frankel et al., 1969; Worrall et al., 1992; Scott et al., 2004; Wilson and Zhang, 2009).

Callose Synthase 5 (*CALS5*) plays the predominant role in the synthesis of the callose wall (Dong et al., 2005). In the (*cals5*) mutant, fertility is significantly reduced resulted from absent callose deposition during microsporogenesis. The bacula and tectum (making up the sexine) fail to develop, neither the tryphine of the sexine is deposited on the outer surface of the pollen (Dong et al., 2005). Recently, the *Cyclin-Dependent Kinase G1* (*CDKG1*) has been reported to be associated with regulating *CAS5* splicing and pollen wall formation in *Arabidopsis thaliana.* Abnormally spliced *Cas5* pre-mRNA is found in *cdkg1* mutant, leading to reduced male sterility caused by impaired callose synthesis. The mutant displays abnormal pollen wall formed with thinner and defective primexine matrix. All this evidence supports that the callose wall development dramatically affects pollen wall formation during microsporogenesis (Huang et al., 2013).

1.1.5.2 Sporopollenin metabolism

The pollen wall development involves the biosynthesis, translocation and degeneration of the chemical ingredient of each cell layer. It has been found that the biosynthesis of sporopollenin is highly associated with lipid and

phenolic metabolism. A large number of genes and their corresponding enzymes that are involved in the production of sporopollenin precursors have now been characterised, indicting that fatty acid metabolism can impact the formation of pollen exine.

The *MS2* gene is a putative fatty acid reductase catalysing the conversion of fatty acyl groups into fatty alcohol groups, which is thought to be involved in sporopollenin biosynthesis (Aarts et al., 1997). Exine formation and production of viable pollen do not occur in the *ms2* mutant due to altered biosynthesis of precursors (Aarts et al., 1997).

CYP703, which belongs to a P450 family, particularly plays a role in pollen development (Morant et al., 2007). Mutations in a certain P450 family appear to introduce the blockage of exine formation. The production of microspores in *Arabidopsis thaliana cyp703a* mutant is defective, showing a phenotype of smooth pollen surface with a complete loss of exine. *In vitro,* CYP703A2 displays a preferential activity at the C-7 position of saturated medium-chain fatty acids (C10, C12, C14, C16) in the catalytic process of mono-hydroxylation (Morant et al., 2007). In terms of *Arabidopsis* CYP704B1, which uses different substrates from CYP703A, is expressed mainly in microspores and tapetal cells. Its null mutant shows a complete absence of the exine and tryphine (Dobritsa et al., 2009). Similarly, mutations in *CYP704B2*, the orthologue of AtCYP704B1 in rice also generate immature microspores with impaired exine formation. Heterologous expression of both CYP704Bs in yeast cells are respectively

demonstrated to catalyse the production of ω-hydroxylated fatty acids with 16 and 18 carbon chains (Li et al., 2010).

Arabidopsis thaliana acyl-CoA (*ACOS5*) is proposed to encode an acyl-CoA synthetase protein producing acyl-CoA esters, which are key intermediates required for the pathway of sporopollenin biosynthesis. The enzyme shows *in vitro* preference for medium-chain fatty acids. Loss of ACOS5 activity would lead to blockage of the hydroxy-fatty acyl sporopollenin monomer secretion. An *acos5* mutant is completely male sterile without any fertile seeds generated by self-fertilization. This is caused by the abortion of pollen development, which reveals acutely defective pollen wall lacking sporopollenin or exine (de Azevedo Souza et al., 2009).

Arabidopsis thaliana tapetal specific genes *Less Adhesive Pollen 5* (*LAP5*) and *Less Adhesive Pollen 6* (*LAP6*) encode the ER localized Plant type III polyketide synthases PKSA and PKSB proteins, whose double mutants are completely male sterile due to the absence of exine deposition. Biosynthesis of phenolic constituents of sporopollenin is disturbed in the mutants by compromised gene expression, with an observation of reduced accumulation of flavonoid precursors and flavonoids in developing anthers. Both enzymes preferentially use midchain and ω-hydroxylated fatty acyl-CoAs as the substrates *in vitro*. Considering *LAP5* and *LAP6* are tightly co-expressed with ACOS5, it is suggested that ACOS5 (fatty acid hydroxylases) and LAP5/LAP6 (acyl-CoA synthetase) act sequentially to

generate the production of hydroxylated α-pyrone polyketide compounds, the potential and previously unknown sporopollenin precursors (Dobritsa et al., 2010; Kim et al., 2010).

Other genes involving the pathways of pollen wall development have also been identified as important, including *Kaonashi 2* (*KNS2*)*, Less Adhesive Pollen 3* (*LAP3*)*,* and *No Exine Formation 1* (*NEF1*).

The *Kaonashi 2* (*KNS2*) gene, previously reported as *AtSPS2F* and *AtSPS5.2*, encodes the enzyme sucrose phosphate synthase (SPS) (Langenkamper et al., 2002; Lutfiyya et al., 2007). The enzyme is proposed as required for synthesis of primexine or callose wall, both of which are important for probacula positioning. The *kns2* mutants display abnormal bacula distribution in the exine structure, but normal fertility. These fully fertile *kns2* pollen grains were covered with densely distributed bacula with smaller mesh size than the wild type.

The *Less Adhesive Pollen 3* (*LAP3*) gene encodes a putative *Arabidopsis thaliana* β–propeller enzyme essential for proper exine formation, whose insertional mutants give rise to male sterility with the loss of tectum structures. The exine turns into thinner and frail in the mutant compared with the wild type. Meanwhile, a wide range of metabolic pathways are influenced, especially severe changes in the levels of the flavonoid biosynthesis pathway (Dobritsa et al., 2009) .

The *No Exine Formation 1* (*NEF1*) gene encodes a predicted plastid integral membrane protein. Primexine of *nef1* mutant is developed crudely with no prebacular formation, the sporollenin of which aggregated and accumulated on the inner surface of the locule wall in place of the microspores. Lipid analysis indicated a significantly reduced total lipid content in the *nef* mutant when compared with wide-type. This suggested that *NEF1* is involved in lipid metabolism (Ariizumi et al., 2004).

Precursors of sporopollenin synthesised in the tapetum catalysed by a series of enzymatic reactions need to be sequentially transported onto the surface of microspores. In *Arabidopsis thaliana,* there are two ways of transportation: vesicular transport and the use of transporters (Wilson and Zhang, 2009).

Arabidopsis thaliana ABCG26/WBC27 encodes a ATP-binding cassette transporter protein located on the tapetum plasma membrane that is required for both pollen wall formation and normal pollen development (Xu et al., 2010). This enzyme is proposed to transport sporopollenin precursors from the tapetum plasma membrane into the locule. Male fertility of the *abcg26/wbc27* mutants is dramatically reduced, in which the reticulate pattern of exine is absent, with associated microspores degeneration. *ABCG26/WBC27* displays high co-expression with genes required for sporopollenin precursor synthesis, for instance, *CYP704B1, ACOS5, MS2 and CYP703A2* (Quilichini et al., 2010; Choi et al., 2011) .

In the rice tapetum, ATP-binding cassette transporter protein ABCG15 is

required for the formation of anther cuticle, orbicules and pollen exine at the young microspore stage. In the *abcg15* mutant, there is an absence of orbicules and exine formation, with associated male sterility; a glossy anther epidermal surface is also observed. It is suggested that ABCG15 acts potentially as the transporter of wax, cutin and sporopollenin precursors, whose levels show significant decrease in the *abcg15* mutant. Expression of genes engaged in lipid metabolism, including homologues of *Arabidopsis thaliana* genes, is found altered in the mutant (Qin et al., 2013).

1.1.5.3 Exine patterning

Primexine deposition is thought to be partially regulated by brassinosteroid biosynthesis (Papini et al., 1999). Even though normal exine formation occurs in the mutant of *Transient Defective Exine 1 (TDE1)*, primexine deposition and bacular formation at the early stages are defective (Ariizumi et al., 2008). The *TDE1* gene was found to be identical to the *DE-ETIOLATED2* gene involved in BR biosynthesis and treatment of the *tde1* mutant with BR was capable of rescuing the exine deposition (Ariizumi et al., 2008).

Arabidopsis thaliana Defective in Exine Formation 1 (*DEX1*) encodes a putative membrane-associated protein containing several potential calcium-binding domains. In the *dex1* mutant, the exine is absent, and obstructed primexine assembly is observed, which gives rise to delayed deposition and abnormal primexine formation. The sporollenin aggregates fails to anchor to the microspore plasma membrane, but rather is randomly

deposited to form larger aggregates onto the developing microspore and locule walls (Paxson-Sowders et al., 1997; Paxson-Sowders et al., 2001).

The *Ruptured Pollen Grain 1* (*RPG1*) gene encoding a plasma membrane protein is strongly expressed in the microspore and tapetum during male meiosis. Mutants of *RPG1* lead to male sterility due to microspores with impaired exine pattern formation. This is proposed as the consequence of defective sporopollenin deposition, which randomly deposited on the pollen surface, and that formation of the priexine is aberrant (Guan et al., 2008).

The *Faceless Pollen 1* (*FLP1*) mutant of *Arabidopsis thaliana*, which produce defective pollen with an almost smooth surface with no reticulate pattern, is identified as a conditional male sterile that can be restored at higher humidity (Ariizumi et al., 2003). As a potential membrane receptor, FLP1 may play a role in sporopollenin transfer, or polymerization to the primexine. Exaggerated tryphine deposition resulting in smooth exine is the consequence of the disordered transporting the intermediates or end products of the tryphine and sporopollenin (Ariizumi et al., 2003).

Ariizumi T and Toriyama K have developed a schematic for the genetic regulation of sporopollenin synthesis and pollen exine development to date (Figure 1.10).

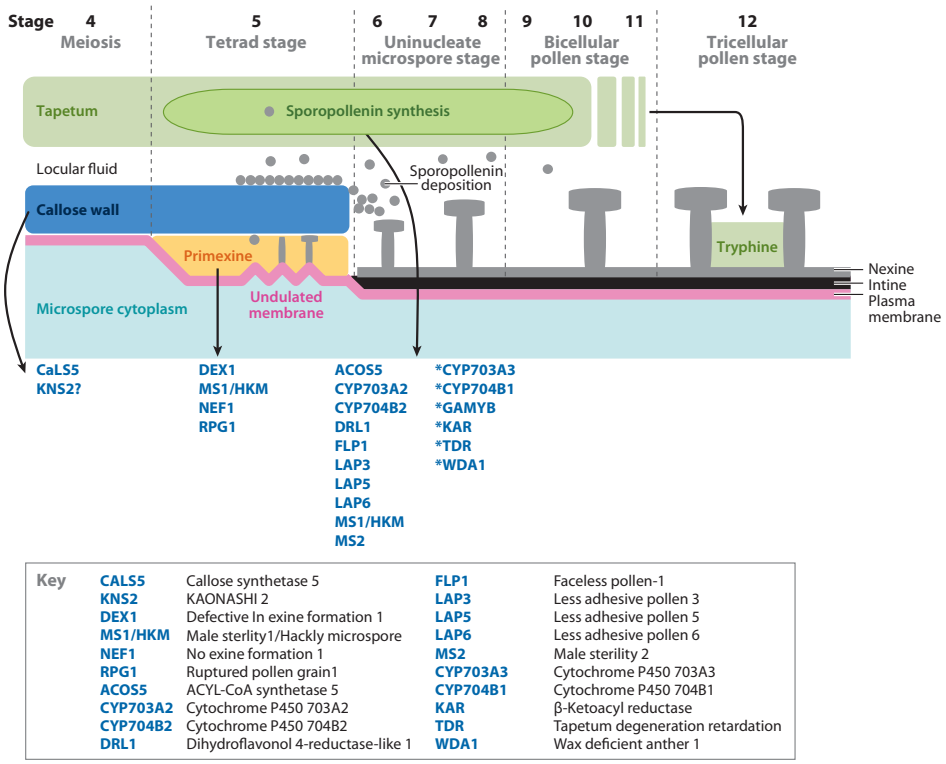


Figure 1.10 Current model of pollen wall formation in *Arabidopsis thaliana* (Ariizumi and Toriyama, 2011). Stages 4 to 12 are defined as Sanders (Sanders et al., 1999). Asterisks indicate genes identified from rice.

1.2 *ARABIDOPSIS* MALE STERILITY1 IS REQUIRED FOR TAPETAL DEVELOPMENT AND POLLEN WALL FORMATION

The *Arabidopsis MALE STERILITY1 (MS1)* gene, encoding a Plant Homedomain (PHD) transcription factor is critical for viable pollen formation (Figure 1.11 A) (Wilson et al., 2001; Alves-Ferreira et al., 2007; Ito et al., 2007; Yang et al., 2007). MS1 shows tightly regulated expression in the tapetum nuclei, from the stage of callose wall breakdown until the free microspore stage (Figure 1.11 b-f) (Wilson et al., 2001; Ito and Shinozaki, 2002; Yang et al., 2007). Yang and Ito respectively confirmed

this using fused MS1-GFP under MS1 endogenous promoter or *in situ* hybridization analysis, respectively (Ito & Shinozaki, 2002, Yang et al., 2007).

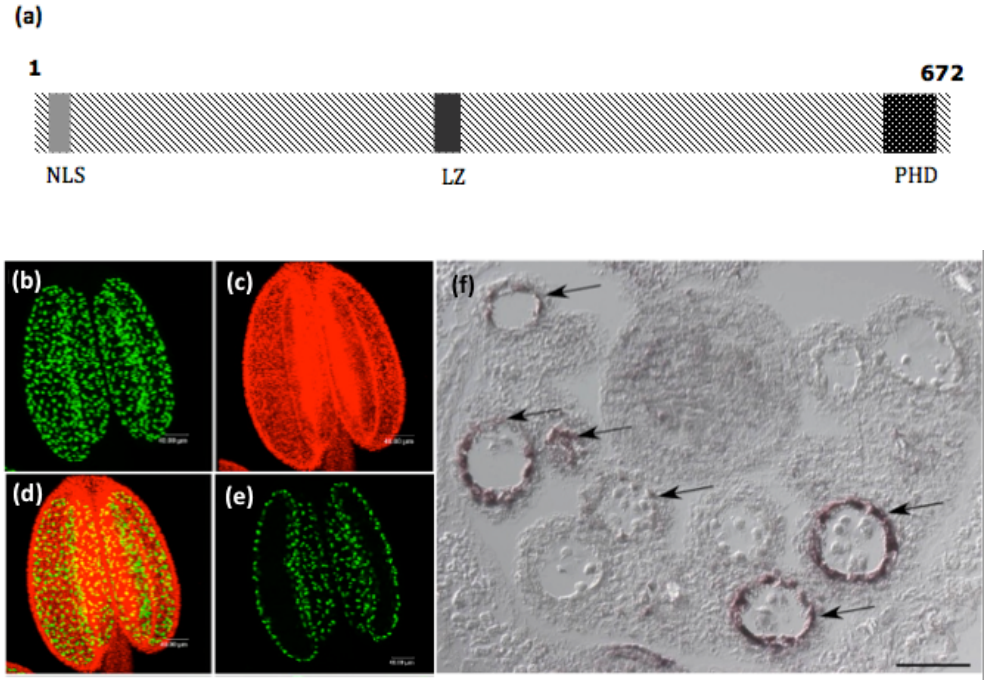


Figure 1.11 ***Arabidopsis MALE STERILITY1*** **encodes a PHD-type transcription factor that is tightly regulated in the tapetum.** (a) Domains of MS1 Protein (Ito and Shinozaki, 2002). NLS, LZ, and PHD shown by dark gray indicate nuclear localization signal, Leu zipper–like region, and PHD region, respectively. (b) (c) (d) and (e) MS1 expression in transgenic lines carrying a functional c-terminal ms1:gfp fusion protein (Yang et al., 2007). Maximal expression of the MS1: GFP fusion protein is seen during microspore release (b and d), and expression is confined to the tapetal tissue within the anthers (c and e). (f) *In situ hybridization* analysis (Ito et al., 2007). Transverse sections of the L*er* floral buds were hybridized with an antisense *MS1* probe. Arrows indicate tapetal layers expressing MS1 mRNA. Bar = 100 µm.

General plant growth of the *ms1* mutant displays no abnormity except that the inflorescences of the *ms1* mutant contain more buds, which and larger

when compared with those of wild type (Vizcay-Barrena and Wilson, 2006). Regarding pollen formation, the early stages of meiosis are not affected in the *ms1* mutant, but altered tapetal secretion and exine structure are observed since the callose wall degrades and microspores are released from the tetrads， which eventually result in the failure of functional pollen formation in the anther locule (Figure 1.12) (Wilson et al., 2001; Vizcay-Barrena and Wilson, 2006). Moreover, TUNEL (terminal deoxynucleotidyl transferase–mediated dUTP nick-end labeling) analysis revealed that the *ms1* tapetum breakdown fails to follow the normal manner of programmed cell death (PCD) as the wild type, but rather shows a necrotic-based cell death (Pacini et al., 1985; Vizcay-Barrena and Wilson, 2006).

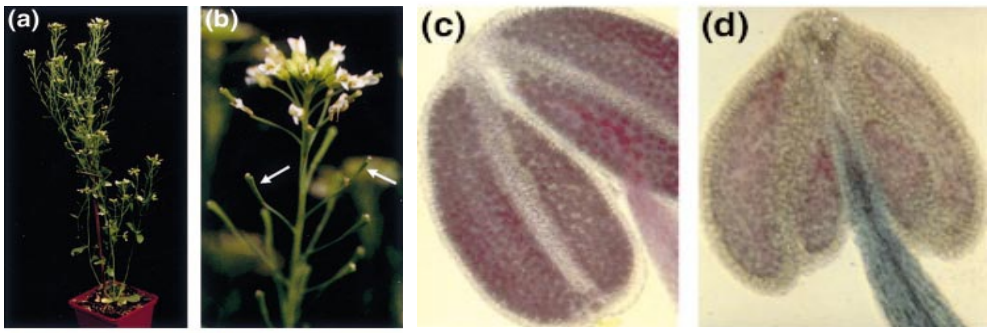


Figure 1.12. Phenotype of the *ms1* mutant (Wilson et al., 2001). (a) The *ms1* mutant plant. The plant appears as wild type except that silique development is abnormal since no viable pollen is produced and self-fertilization does not occur. (b) Detail of the *ms1* mutant inflorescence, the *ms*1 flowers do not self-fertilise and thus silique elongation does not occur (arrows). (c) L*er* wt anther, viable pollen is clearly seen stained purple (Alexander, 1969). (d) *ms1* anther, no viable pollen is seen.

Previous research strongly supported that MS1 acts as a transcription factor that plays a key role in promoting pollen development by regulating late tapetal gene expression. Large numbers of genes have been found to be down-regulated in *ms1* mutant, and identified as downstream of MS1 (Alves-Ferreira et al., 2007; Ito et al., 2007; Yang et al., 2007).

1.3 GENETIC REGULATION OF ANTHER DEVELOPMENT

A growing number of transcription factors (TFs) have been discovered in *Arabidopsis thaliana* that act as key regulators for tapetal specification and pollen wall formation (Wilson and Zhang, 2009; Zhu et al., 2011). These TFs include DYSFUNCTIONAL TAPETUM 1 (DYT1) (Zhang et al., 2006), *TAPETAL DEVELOPMENT AND FUNCTION 1* (TDF1/AtMYB35) (Zhu et al., 2008), ABORTED MICROSPORE (AMS) (Sorensen et al., 2003), and MYB DOMAIN PROTEIN 80/ MALE STERILE 188 (AtMYB80/MS188) (Higginson et al., 2003).

The *DYT1* TF encodes for a putative basic helix-loop-helix (bHLH) transcription factor that appears to be one of the earliest players in tapetal development, after initiation of the anther cell layers has occurred. It is critical for tapetal gene regulation, and required for normal expression of many tapetum specific genes (Zhang et al., 2006). In the *dyt1* mutant, the tapetum becomes highly vacuolated, PMC meiosis is initiated, but the callose wall is thin and cytokinesis rarely occurs (Zhang et al., 2006).

Expression of many tapetum genes are reduced severly in *dyt1* mutant (Feng et al., 2012).

Tapetal development and function1 (TDF1) encodes a putative R2R3 MYB transcription factor (MYB35). This gene is highly expressed in both tapetal cells and microspores, playing a critical role in controlling callose dissolution. In *tdf1* mutant, callose fails to breakdown and no expression of the *A6* gene that encodes a putative β-1,3-glucanase as part of the callase enzyme complex, is seen. Evidence indicates that *TDF1* is vital in tapetal differentiation and function (Zhu, Chen et al. 2008).

AMS belongs to the MYC subfamily of bHLH genes, showing a prolonged expression compared with many other tapetal specific genes. It remains at a high expression level until mitosis I and the bicellular microspore stage. *AMS* is required for both tapetum development and post-meiotic microspore formation. The *ams* mutant displays complete male sterile due to premature degeneration of tapetal cells and microspores before entering mitosis (Sorensen et al., 2003). Altered expression of genes involving lipid synthesis was found in *ams* mutant by microarray analysis, suggesting its importance in the regulation of normal anther development and pollen formation (Xu et al., 2010; Ma et al., 2012).

The *Arabidopsis thaliana* MYB80 (formerly AtMYB103)/ MALE STERILE 188 (MS188) is an R2R3 transcription factor that is essential for anther development by regulating tapetum development, callose dissolution and pollen exine formation. In the *ms188* mutant, the tapetum degenerates

early, without releasing oil bodies, vesicles, and plastids (Higginson et al., 2003; Li et al., 2007). This gives rise to a reduction in enzymes involved in callose degradation and, subsequently, prevents the formation of the exine (Li et al., 2007). In the *ms188* allele of *myb103*, breakdown of callose and *A6* expression are reduced with abnormal exine formation occurring (Zhang et al., 2007). Impaired cell wall modification, lipid metabolic pathways and signal transduction has also been detected as the consequence of loss-of-function of AtMYB80 (Zhu et al., 2010). Recently, homologs of *MYB80* have been cloned form wheat, rice, canola, and cotton (Phan et al., 2012).

These transcription factors are linking with each other by feed-forward loops, and appear to play a regulatory role at different stages during tapetum development and pollen formation (Figure 1.13) (Zhu et al., 2011). DYT1 is proposed to act as upstream of *TDF1, AMS, MYB80* and *MS1,* while *MS1* could be regulated by *AMS1 and MYB80* as well (Ma, Feng et al. 2012). *AtMYB99* has been proposed as the putative direct target of *MS1* (Alves-Ferreira et al., 2007) and Berr et al (Berr et al., 2010) revealed that expression of *MS1* is affected by a SET domain protein SDG2 required for global Histone H3 lysine 4 trimethylation (H3K4me3) deposition. Another SET domain protein, SDG4/ ASHR3 also shows association with AMS (Cartagena et al., 2008; Thorstensen et al., 2008).

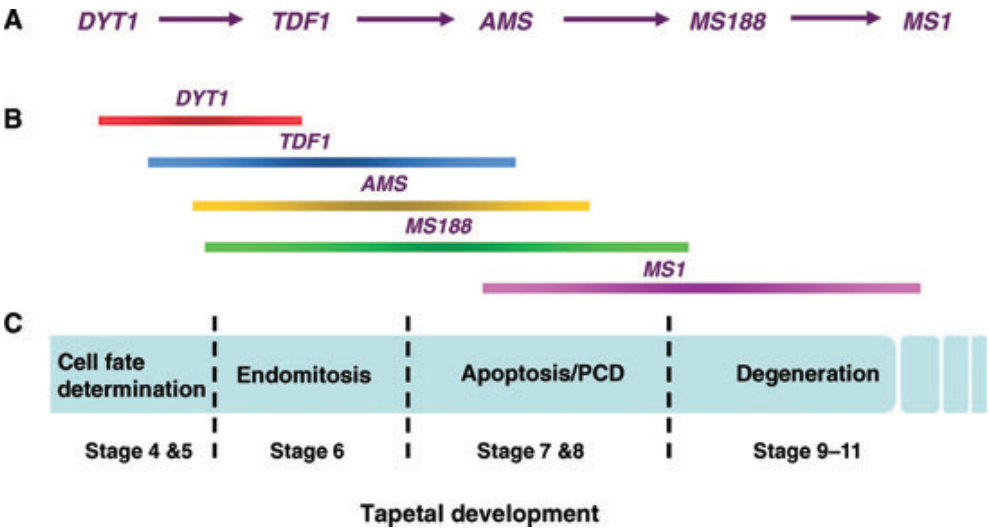


Figure 1.13. The expression of TFs in the genetic pathway and the main events during the tapetum development (Zhu et al., 2011). (A) The genetic pathway of transcription factors for tapetum development and function. (B) The expression of key transcription factors during tapetum development. The lines with different colors represent the expression pattern of different transcription factors based on *in situ* hybridization results. (C) Tapetum development showing the main events at different anther stages including cell fate determination, endomitosis, apopotosis/programmed cell death and degeneration.

Yang, Feng and Ma have concluded the regulatory frameworks so far of tapetum development and pollen formation, respectively using *MS1, DYT1* and *AMS* at the centre (Figure 1.14) (Feng et al., 2012). Nevertheless, the exact regulatory mechanism of tapetum TF regulation is still not well defined.

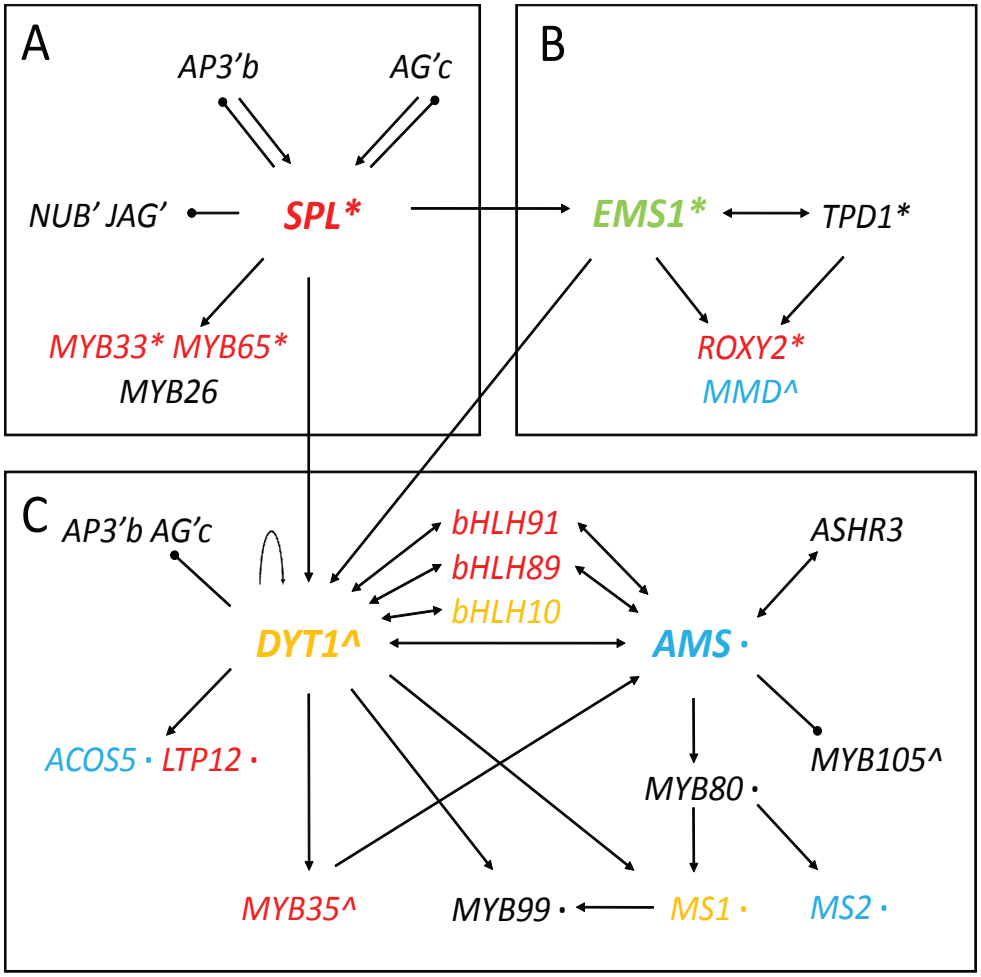


Figure 1.14 Gene regulatory network of anther development during early stages (Ma et al., 2012). Gene regulation is represented by T-bars (negatively) and arrows (positively). The direct regulation confirmed by experiment is represented in bold line. Genes encoding proteins with interaction is represented by double arrows. Gene expression patterns in different tissues are shown by colors (blue for anther specific; red for anther- preferential; green for reproductive-preferential and yellow for genes not included in ATH1 chip). Gene function in tapetum formation is marked by an apostrophe; in pollen wall formation by an asterisk; in callose dissolution by double asterisks; in stamen and petal formation by the letter b; in stamen and carpel formation by the letter c.

1.4 AIMS AND OBJECTS OF THE PROJECT

The aims of this project were to investigate putative MS1 interacting proteins and MS1 regulatory targets, to provide valuable insight into the MS1 regulatory network, and to integrate this regulatory framework into pollen development. The specific objectives of the work were as follows:

1. MS1 interacting proteins were identified using Förster resonance energy transfer (FRET) and protein pull-down assay (Chapter 3).

2. Temporal expression pattern of MS1 putative interacting protein was examined, and its biological function was investigated by manipulation of gene expression in transgenic plants (Chapter 4).

3. Direct targets of MS1 were identified by comparative transcriptome analysis and chromatin immunoprecipitation (ChIP) (Chapter 5).

4. Investigations were also conducted to confirm the nature of MS1-mediated histone modification (Chapter 6).

CHAPTER 2 MATERIALS AND METHODS

2.1 PLANT GROWTH

Seeds of *Arabidopsis thaliana* Columbia (C*ol*-0), Landsberg *erecta* (L*er*) and *ms1ttg1* mutant (L*er* background, NASC ID N1298), obtained from the Nottingham *Arabidopsis* Stock Center (NASC), were sown on Levington M3: John Innes No.3: vermiculite: perlite (6:6:1:1) compost mix supplemented with 2% (w/v) Intercept® (Scotts) and placed in the glasshouse. The plants were grown under the following conditions: 22hours of daylight (150μmol·m^{-2}·sec^{-1}) at 20±2°C and 2hours of night at 16±2°C. Plants were watered until at least 90% of the siliques had dried completely and then the plants were allowed to dry slowly on dry benches for maximum viable seed production. Seeds were harvested after the whole plants had dried out and were stored in moisture-porous paper bags in a dry atmosphere at room temperature.

2.2 SEED STERILISATION AND GROWTH OF PLANTS

Seeds of *Arabidopsis thaliana* were surface-sterilized for 10mins in 70% (v/v) ethanol with violent shaking, prior to being washed in absolute ethanol three times. Surface-sterilized seeds were spread onto Petri dishes containing 20ml of sterile half MS medium with selective antibiotics. Plates were placed in a cold room for 3 days at 4°C to synchronize germination, and then cultured under full light (140μmol·m^{-2}·sec^{-1}) at 22-24°C. After one

week, germinated seedlings were transferred into soil and grown (Section 2.1) until seeds were collected.

2.3 GENOMIC DNA EXTRACTION

2.3.1 Rapid Genomic DNA Isolation Using Sigma Kit

Rapid isolation of the genomic DNA from plant was using the Extract-N-Amp Kit (Sigma) following the manufacturer's instructions. About 10mg of samples, typically a 0.5-0.7cm section of leaf tissue, were collected from plants and then boiled for 10mins at 95°C in 50µl of Extraction Buffer. The scissors used for sample collection were wiped with 70% (v/v) ethanol between each cut to avoid cross contamination. Each sample was then diluted with equal volume of Dilution Buffer (Sigma), which needed to be used immediately, or stored at -20°C.

2.3.2 Sucrose Method for Crude DNA Extraction from Plant Materials

Sucrose method for crude DNA extraction was used for large-scale genotyping of plants as previously described (Berendzen et al., 2005). 2-3mm diameter of leaf tissues were cut from the plant and immediately transferred into a tube on ice with 100µl of the Sucrose Buffer (50mM Tris-Cl, 300mM NaCl, 300mM sucrose). After roughly crushed with a pipette tip in the tube, each sample was boiled for 10mins at 95°C prior to briefly spun at 2000–6000g for 5sec. 0.5-1µl of the extracted DNA was immediately used for PCR amplification.

2.3.3 Genomic DNA Extraction Using Qiagen Kit

Genomic DNA extraction using DNeasy Plant Mini Kit (Qiagen) was conducted according to the instruction provided by the manufacturer. 100mg of plant materials for genomic DNA extraction were collected into 1.5ml tubes and immediately frozen in liquid nitrogen, and ground into powder using a plastic pestle (Kimble Chase) pre-chilled in liquid nitrogen. Throughout the grinding process, the tubes were dipped into liquid nitrogen to ensure the samples did not thaw. Ground materials were then incubated for 10mins at 65°C in 400 µl Buffer AP1 and 4µl 100mg/ml RNase A solution. Subsequently, the lysates were added with 130µl Buffer AP2 and incubated on ice for 5mins, before passing through the QIAshredder Mini spin column by centrifuged for 2mins at 20,000g. The flow-through was transferred to a new tube with 675µl of Buffer AP3/E and mixed by pipetting. The lysate was then loaded onto a DNeasy Mini spin column in a 2ml collection tube, which was centrifuged for 1min at 6,000g afterwards. The column was loaded with 500µl Buffer AW, centrifuged for 1min at 6,000g and the flow-through discarded. This was repeated once, after which the column was centrifuged for 2mins at 20,000g. 50µl of molecular grade water (Sigma) was added to the centre of the column, which was incubated at room temperature for 5mins, and then centrifuged for 1min at 6,000g to elute the DNA into a new 1.5ml tube. Quantification of genomic DNA was carried out by NanoDrop 2000 spectrophotometer (Thermo Fisher). The DNA was stored at -20°C unless used immediately.

2.4 RNA EXTRACTION

RNA was extracted from approximately 100mg of plant tissue, typically the closed buds of 5-week-old plants unless otherwise stated, using RNeasy Plant Kit (Qiagen) according to the manufacturer's instruction. Plant material was collected and ground as described in Section 2.3.3. The disrupted tissue was vortexed in 450µl Buffer RLC containing 4.5µl β–mercaptoethanol, then incubated for 3mins at 56°C. The lysate was then transferred to a QIA shredder spin column in a 2ml collection tube, which was then centrifuged at 23,000g for 2mins. The flow through was transferred to a new tube containing 225µl of 100% ethanol and mixed by pipetting. The mixture was then loaded to an RNeasy Mini Spin column and centrifuged for 15secs at 8,000g. The samples were then treated with 1U RNase free DNase (Promega) at 37°C for 45mins. Subsequently, the column was added 700µl of Buffer RW1 and centrifuged for another 15secs at 8,000g. The flow through was discarded，and the column was transferred to a new collection tube and washed twice with 500µl of Buffer RPE, by centrifugation for 2mins at 8,000g. Flow through was discarded after each wash and the column was centrifuged for 1min more at 23,000g to ensure all liquid was removed. To elute the RNA, the column was transferred to a new 1.5ml tube and 30µl of DNase and RNase free water (QIAGEN) was added to the centre of the column. After incubation at room temperature for 1min, the column was centrifuged for 1min at 8,000g. RNA concentrations were quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher). 2µg of total RNA was run on a 1% (w/v) agarose gel at 100 V/cm to check

its integrity. RNA stocks were stored at -80°C unless used immediately.

2.5 cDNA SYNTHESIS

Following RNA purification (Section 2.4), 5µg of total RNA was used for cDNA synthesis. RNA samples was precipitated together with 200-500ng Oligo dT_{12-18} and 10mM dNTP mix, and incubated at 65°C for 5mins to destruct any secondary structures (Table 2.1). Reverse transcription (RT) was then performed using Superscript II Reverse Transcriptase (Invitrogen) at 50°C for 1 hour, and the reaction was then deactivated at 70°C for 15mins.

Table 2.1 cDNA Synthesis Reaction.

Reagent	Volume
200-500ng Oligo dT_{12-18}	1µl
10mM dNTP Mix	1µl
5µg Total RNA	11µl
Incubated at 65°C for 5mins then add:	
5xFirst Strand Buffer	4µl
0.1M DTT	1µl
40 Units RNaseOUT RNase Inhibitor (Invitrogen)	1µl
200 Units Superscript II RT (Invitrogen)	1µl

2.6 POLYMERASE CHAIN REACTION (PCR) AMPLIFICATION

2.6.1 PCR Using Red Taq® Ready Mix® PCR Reaction Mix

For all non proof-reading applications, such as colony screening, PCR amplification was conducted using Red Taq® Ready Mix® PCR Reaction Mix (Sigma-Aldrich) in a 10µl reaction volume (Table 2.2), following typical conditions: 94°C for 3mins; 40 cycles of 94°C for 30secs, 55-60°C for 30secs, 72°C for 1min/kb); then 72°C for 5mins.

Table 2.2 PCR Reaction for Red Taq® Ready Mix® PCR Reaction Mix.

Reagent	Volume
2xRed Taq® Ready Mix®	5µl
10µM Forward Primer	0.25µl
10µM Reverse Primer	0.25µl
50ng DNA Template	0.5µl
ddH_2O	Up to 10µl

2.6.2 PCR Using Phusion High-Fidelity DNA Polymerase

For high fidelity PCR amplifications that required proof-reading activity, PCR was performed using Phusion High-Fidelity DNA Polymerase (Thermo Fisher) as shown in Table 2.3, following typical conditions: 98°C for 30secs; 30 cycles of 98°C for 30secs, 55-65°C for 30secs, 72°C for 30secs/kb); then 72°C for 6mins.

Table 2.3 PCR Reaction for Phusion High-Fidelity DNA Polymerase.

Reagent	Volume
Phusion High-Fidelity DNA Polymerase	0.5µl
5xHF/GC Buffer	10µl
10mM dNTPs	1µl
10µM Forward Primer	0.25µl
10µM Reverse Primer	0.25µl
100ng/5µl DNA Template	1µl
ddH_2O	Up to 50µl

2.6.3 Real-Time Quantitative PCR

qRT-PCR (Real-Time Quantitative PCR) was performed for quantitative analysis using all regents provided by Maxima® SYBR® Green/ROX qPCR Master Kit (Fermentas) as in Table 2.4, operated on the LightCycler® 480 Real-Time PCR system (Roche Applied Science). The LightCycler® 480 software was used to operate the PCR system and analyse the data, following typical conditions: 95°C for 3mins; 40 cycles of 95°C for 30secs, 55°C for 30secs, 72°C for 30secs; 72°C for 6mins; and then the dissociation programme (1°C per cycle from 55°C to 95°C), which was designed to determine amplification specificity. All samples were run at least in duplicate. For gene expression analysis, the house-keeping ACTIN transcripts were used to normalize amplification between experimental samples. Relative expression levels were determined in comparison to ACTIN expression using the $2^{-\Delta\Delta CT}$ analysis method (Livak and Schmittgen, 2001).

Table 2.4 Real-Time Quantitative PCR Reaction.

Reagent	Volume
2xSYBR® Green Master Mix	4.5µl
10µM Forward Primer	0.1µl
10µM Reverse Primer	0.1µl
cDNA Template	0.2µl
ddH_2O	Up to 10µl

2.7 AGAROSE GEL ELECTROPHORESIS

Electrophoresis was carried out after PCR to check the size and/or amount of the products. PCR products mixed with 6xDNA loading buffer (Bioline) were run on 0.8-2.0%(w/v) agarose (Sigma-Aldrich) gels for 40mins at 100V in 0.5xTBE buffer, which was visualized by 0.2%(v/v) 10mg/ml ethidium bromide staining (Sigma-Aldrich). Sizes of the PCR products were determined using a HyperLadder I marker (Bioline) unless otherwise stated.

2.8 DNA RECOVERY AND CLEAN-UP

2.8.1 Phenol-Chloroform DNA Purification

DNA samples were thoroughly mixed with an equal volume of phenol:chloroform:IAA (25:24:1), pH8.0. After centrifuged for 2mins at 18,000g, the upper aqueous phase containing the DNA was transferred to a clean tube to be precipitated by adding 1/10 volume of Sodium Acetate (3 M, pH5.2) and 2-3 fold volume of 100% (v/v) ethanol, or equal volume of 100% isopropanol. The DNA was pelleted by incubation on ice for at least

15mins and centrifuged at 23,000g for 30mins at 4°C, then rinsed in 70% (v/v) ethanol after which the supernatant was removed with care to avoid disturbing the pellet. An additional 15mins of centrifuge at 23,000g was used to precipitate the DNA and remove all residual ethanol; the pellet was then dried for 5mins under vacuum. The purified DNA sample was subsequently dissolved in 30-50µl of ddH_2O.

2.8.2 Recovery of DNA from Agarose Gels

Recovery of DNA from agarose gels was using QIAGEN Gel Extraction Kit following the manufacturer's instructions with slight adjustments. The gel slice was weighted and 300µl of Buffer QG was added to per 100mg of excised agarose gel. After incubation at 50°C until the gel had completely dissolved, the sample was mixed with 20µl of 3M sodium acetate (pH5.2) and 100µl of isopropanol for per 100mg of gel, which was then transferred to the DNA cleanup column. Flow-through was discarded and the column was washed with 0.5ml of Buffer QG and 0.75 ml of Buffer PE, respectively, which was centrifuged at 18,000g for 1min after each wash. An additional 1 min spin was performed to ensure remaining liquid removed, followed by the elution step by adding 30-50µl of ddH_2O to the centre of the column. The DNA sample was eluted to a new 1.5ml tube by incubated at room temperature for 1min prior to centrifuged at 18,000g for 1mins.

2.8.3 Concentrated DNA Purification Using MinElute Kit (Qiagen)

DNA samples following reactions, such as PCR products, were purified using the MinElute PCR purification kit (Qiagen) following the manufacturer's instruction with slight adjustments. To obtain highly concentrated DNA, the sample was precipitated by mixed with 1/10 volumes of sodium acetate (NaAc, 3M, pH5.2) and 5-fold volumes of Buffer PBI, which was then transferred to a MinElute column on a 2ml collection tube and centrifuged for 1 min at 8,000g.The flow through was discarded and the column was washed with 750µl of Buffer PE by centrifuged for 1 min at 18,000g. After the flow-through was discarded, the column was centrifuged for an additional 1 min at 23,000g to completely the remove residual ethanol from Buffer PE. To elute the DNA sample, 10µl of water was added to the center of the column, which was transferred to a new 1.5ml tube and then centrifuged for 1 min at 18,000g.

2.9 CLONING PROCEDURES

Maps of all plasmid vectors used for cloning are in Appendices.

2.9.1 Topoisomerase Based Cloning

Topoisomerase based cloning technology allows the PCR products to be cloned into a TOPO vector (Invitrogen), such as pCR-BluntII-TOPO vector, for proof-reading sequencing (Section 2.13). Cloning of the purified PCR products (Section 2.8.2) amplified by the Phusion polymerase (Section.2.6.2)

was performed in reaction volumes of 6µl at room temperature (21°C) for 30mins (Table 2.5).

Table 2.5 TOPO Cloning Reaction.

Reagent	Volume
PCR Product	0.5-4µl
Salt Solution (1.2M NaCl, 0.06M $MgCl_2$)	1µl
TOPO Vector (15-20 ng/µl)	1µl
ddH_2O	Up to 6µl

Reactions were then transformed into chemically competent *Escherichia coli (E. coli)* DH5α cells (Section 2.11) and resultant colonies PCR screened (Section 2.9.5) using the vector specific M13 primer pair (M13F(-20) 5'TGTAAAACGACGGCCAG3' and M13R 5'CAGGAAACAGCTATGAC3'). Plasmid containing the correct sequence was extracted from the cells (Section 2.12) for further experimentation.

2.8.2 Digestion by Restriction Enzyme

DNA requiring digestion by desired restriction enzymes (New England BioLabs) was incubated in reaction volumes of 50µl at 37°C for at least 1 hour (Table 2.6), depending on the enzyme specification (New England BioLabs), and then deactivated at 65°C for 15mins. Sequential digestion was performed if multiple enzymes were required, or no enzyme buffer was compatible for each enzyme to achieve highest efficiency. Digested plasmids of the vectors used for cloning were then processed with 1 unit

calf intestinal alkaline phosphatase (CIAP) (New England BioLabs) to remove 5' end phosphates, and thereby prevent self-annealing of the plasmids and increase ligation efficiency. Reaction products were checked and purified by agarose gel electrophoresis (Section 2.8.1) prior to ligation.

Table 2.6 Typical Restriction Endonuclease Reactions.

Reagent	Volume
DNA	1µg
10x Enzyme Buffer	5µl
100xBSA	0.5µl
Restriction Enzyme 1	20 units
Restriction Enzyme 2 (Optional)	20 units
ddH_2O	Up to 50µl

2.9.3 Ligation

Ligation reactions were performed using T4 DNA ligase (Invitrogen) in a reaction volume of 10µl with overnight incubation at 16 °C for up to 18hours (Table 2.7). To achieve better ligation efficiency, the molar ratio of plasmid vector: insert was set to be 1:3. All ligation reactions were used immediately for transformation into chemically competent DH5α *E. coli* cells (Section 2.11).

Table 2.7 Ligation Reaction.

Reagent	Volume
Vector Plasmid DNA (3-30fmol)	1µl
Insert DNA (9-90fmol)	1µl
T4 DNA Ligase (1U/µl)	1µl
10xLigase Buffer	1µl
ddH_2O	6µl

2.9.4 Gateway® Technology

The Gateway® Technology provides a rapid and highly efficient method to apply DNA sequences into various vector systems (Hartley et al., 2000), utilizing the site-specific recombination properties of *Bacteriophage Lambda* (Landy, 1989). It consists of two steps, the BP recombination transferring the insert into the entry vector and the LR recombination that transfers the insert to the destination vector (Figure 2.1). The BP reaction facilitates the recombination of the PCR product (*attB* substrate) to create the entry clone. The donor vector contains two recombination sites, *attP1* and *attP2*, for recombination cloning of the *attB*-PCR product; the reaction is catalyzed by the Gateway BPclonase enzyme. The LR reaction facilitates the recombination of the entry clone (*attL* substrate) with the destination vector (*attR* substrate) to create the desired expression clone. The destination vector contains two recombination sites, *attR1* and *attR2*, for recombinational cloning of the entry clone. This reaction is catalysed by the Gateway LRclonase enzyme.

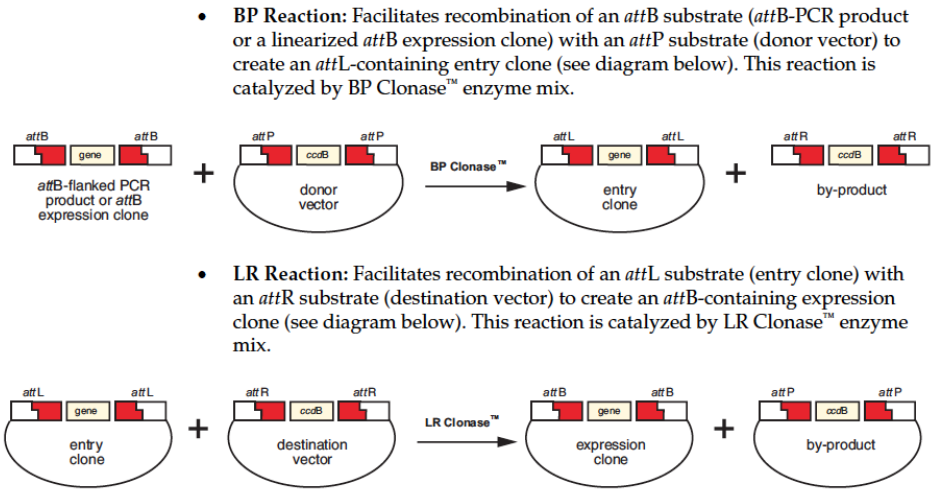


Figure 2.1. Gateway® Technology (Invitrogen Handbook). The BP and LR recombination reactions were set up following the manufacturer's instructions (Invitrogen). The BP reaction was mixed, incubated at 25°C overnight, and then deactivated by adding with 2µl of Proteinase K (Invitrogen) at 37°C for 10mins (Table 2.7). The products were subsequently transformed into chemically competent DH5α E.coli cells (Section 2.11), and clones were PCR screened (Section 2.9.5). Plasmids from clones containing the correct sequence were then used in LR reaction (Gateway LR **Clonase™** Enzyme mix, Invitrogen) with a destination vector desired. Conditions for the LR reaction (Table 2.8) followed the same ones as for the BP reaction as described above; products subsequently followed the sequential procedures for plasmid transformation (Section 2.11), colony screening (Section 2.9.5), plasmid extraction (Section 2.12) and eventually confirmation by sequencing (Section 2.13), as conducted for the BP reaction products.

Table 2.8 BP and LR Recombination Reaction.

BP Recombination Reaction	LR Recombination Reaction	Volume
attB PCR Product (40-100fmol)	Entry Clone (100-300ng)	1-10µl
pDONOR Vector	Destination Vector (150ng/µl)	2µl
5xBP Clonase Reaction Buffer	5xLR Clonase Reaction Buffer	4µl
BP Clonase II Enzyme Mix	LR Clonase II Enzyme Mix	4µl
TE Buffer pH8.0	TE Buffer pH8.0	Up to 16µl

2.9.5 Colony Screening

Colonies were picked using a 10µl pipette tip and dipped into the prepared PCR reaction. The PCR screening was performed by Red Taq® (Sigma-Aldrich) (Section 2.6.2) using primer pairs respectively specific to the insert or the vector, under the conditions: 94°C for 3mins; 40 cycles of 94°C for 30secs, 55-60°C for 30secs, 72°C for 1min/kb. Positive clones were then cultured overnight with 50ml of liquid LB medium (Section 2.10.1) at 37°C with shaking at 200rpm. These were then used for plasmid extraction (Section 2.12) and confirmed by sequencing (Section 2.13) afterwards.

2.10 SELECTIVE MEDIA

2.10.1 Luria-Bertani Medium

Luria-Bertani (LB) medium (1% (w/v) Tryptone (Sigma Aldrich), 0.5% (w/v) Yeast Extract (Sigma Aldrich), 100mM Sodium Chloride (Fisher Scientific), pH7.0 and 1.5% (w/v) Bacterial Agar for solid medium) was used for culturing bacteria (Green and Sambrook, 2012). The medium was autoclaved allowed to cool to ~50°C and antibiotic added (Table 2.9). Solid medium was poured into 9cm Petri dishes (Fisher Scientific) and allowed to set then transferred to 4°C for storage up to 14 days.

2.10.2 Half-strength Murashige and Skoog Medium

Half-strength Murashige and Skoog (MS) medium (2.2g/L MS basal salt mixture (Sigma Aldrich), 9% (w/v) agar) was used for selective and sterile seed germination (Murashige and Skoog, 1962). Plates were prepared no more than a day prior to seed sowing, the medium was autoclaved, allowed to cool to ~50°C and antibiotic was added (Table 2.8). The medium was poured into 12 cm square Petri plates (Fisher Scientific) and allowed to set then transferred to 4°C for storage.

Table 2.9 Antibiotic Stock and Working Concentrations

Antibiotic	Solvent	Stock Concentration	Working Concentration
Ampicillin (Amp)	H_2O	100mg/mL	100µg/mL
carbenicillin (Carb)	H_2O	200mg/mL	50µg/mL
Kanamycin (Kan)	H_2O	50mg/mL	50µg/mL
Hygromycin B (Hyg)	H_2O	25mg/mL	25µg/mL
Spectinomycin (Spec)	H_2O	100mg/mL	100µg/mL
Rifampicin (Rif)	MeOH	25mg/mL	25µg/mL
Tetracycline (Tet)	H_2O	10mg/mL	10µg/mL

2.11 CHEMICALLY COMPETENT CELL TRANSFORMATION

50ul of competent cells directly taken out from a -80°C freezer was thawed on ice. After gently mixed with 50ng of plasmid DNA, the cells were incubated on ice for 30mins and then transferred to a pre-heated water bath at 42°C for exactly 80secs. Immediately removed to an ice bath to stand for 10mins afterwards, the sample was then incubated with addition of 200µl of

liquid S.O.C. medium (2% (w/v) Trytone, 0.5% (w/v) Yeast Extract, 100mM NaCl, 2.5mM KCl, 10mM $MgCl_2$ and 20mM glucose) for 1hour at 37°C with shaking. Recovered transformed cells were spread onto the surface of a prepared LB agar plate with appropriate antibiotics as required (Table 2.8), which would be incubated overnight at 37°C.

2.12 PLASMID DNA EXTRACTION

2.12.1 Miniprep Using Qiagen Kit

Plasmid DNA extracted from no more than 5ml of *E. coli* cell culture was using QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's instructions. The bacterial samples were incubated overnight at 37°C with shaking was pelleted by centrifuged at 4,000g for 20mins and the supernatant discarded. The cell pellet was resuspended in 250µl of Buffer P1 and then mixed with 250µl of Buffer P2 by inverting the tube several times. 350µl of Buffer N3 was added within 5mins and several inversions of the tube applied to mix. The lysate was then centrifuged for 10mins at 18,000g and the supernatant was transferred to a QIAprep spin column in a 2ml collection tube. Then the column was centrifuged for 1min at 18,000g, the supernatant was discarded, 500µl of Buffer PB and 750µl of Buffer PE, in sequence, were added to the column to wash by centrifuged for 1min at 18,000g performed after each wash, with all flow through discarded. The column was centrifuged for an additional 1min to remove any residual wash buffer. To elute the plasmid, 50µl of sterile water was added to the centre of

the column transferred into a new 1.5ml tube, which was then centrifuged for 1min at 18,000g.

2.12.2 Midiprep Using Buffers

Large scale of purification was used to achieve higher concentration of plasmid DNA. The bacterial sample containing the plasmid of interest was cultured in 20ml of liquid LB medium in a 50ml Sterilin tube, which had grown overnight at 37°C with shaking. The cells were pelleted by centrifuged at 10,000g for 10mins. The supernatant was discarded and the pellet was resuspended in 0.5ml of Solution I (50mM Glucose, 25mM Tris-Cl, 0.01mM EDTA pH8.0, 0.1ng/µl RNase A) and transferred to a new 15ml Sterilin tube. 1ml of Solution II (0.2M NaOH, 1% (w/v) SDS) was then added to the tube, which was inverted several times until the sample was clear. To neutralize the lysate, 1.5ml of Solution III (3M KAc, 2M HAc) was subsequently added and mixed by several inversions of the tube until a white precipitate had formed in the sample. The tube was centrifuged for 10mins at 10,000g and the supernatant in volume of approximately 2ml was transferred to a new 15ml Sterilin tube, which was then purified using Phenol-chloroform method (Section 2.8.1).

Quantification of plasmid DNA was performed by NanoDrop 2000 spectrophotometer (Thermo Fisher). The plasmid DNA was stored at -20°C unless used immediately.

2.13 DNA SEQUENCING

Sequencing reactions of purified PCR products (Section 2.8) or plasmids (2.12) were performed using ABI Prism® BigDye® Terminator v1.1 Cycle Sequencing Kit (Life Technologies) (Table.2.10)，following the conditions: 96°C for 90 secs; 28 cycles of 96°C for 30secs, 55-60 °C for 20secs, 60°C for 4mins); then 28°C for 1min. Samples were run on an ABI 3130 analyzer (Applied Biosystem) and analysed using MacVector software (MacVector, Inc).

Table 2.10 Sequencing Reactions.

Reagent	Volume
DNA (100-200ng)	3-4µl
10µM Primer	0.5µl
BigDye Ready Reaction Mix	0.8µl
5xSequencing Buffer	2µl
ddH_2O	Up to 10µl

2.14 TRANSFORMATION OF *AGROBACTERIUM* BY ELECTROPORATION

50µl of competent *Agrobacterium* cells thawed on ice were gently mixed with 200ng of plasmid DNA, then transferred into a dry electroporation cuvette without any air bubbles, samples were pulsed with 1800V using a electroporator (source). 1ml of LB medum was then added and the transformed cells were incubated for 2hours at 28°C with shaking (200 rpm), these were then spread onto LB plates supplemented with rifampicin

(25μg/ml) and other appropriate antibiotics specific to the transformed plasmids. Plates were incubated at 28 °C for 2-3 days.

2.15 TRANSFORMATION OF *ARABIDOPSIS THALIANA* BY FLORAL DIPPING

Transformed *Agrobacteria* samples (Section 2.14) cultured overnight in 5ml of LB medium containing appropriate antibiotics were diluted by 20 fold and incubated at 28°C with shaking until reaching the OD_{600} of 1.0-2.0. The samples were pelleted by centrifuged for 20mins at 4,000g, and then re-suspended in 50ml of freshly made solution (5% (w/v) of sucrose and 0.05% (v/v) Silwet). Newly flowering *Arabidopsis thaliana* plants were dipped into the solution for 1min，and then kept in the plastic sleeves (source) to ensure humidity. Sleeves were gradually opened up over the next day. Seeds were harvested after the whole plants had dried out and screened on half MS medium (Section 2.10.2) containing appropriate antibiotics (Table 2.8) (Klee et al., 1987).

2.16 PROTEIN ANALYSIS

2.16.1 SDS-PAGE Gel Electrophoresis

Protein samples for SDS-PAGE gel analysis were prepared in volumes of 20μl. For 15μl of sample diluted to 100ng-20μg, 5μl 4x Sample buffer (Invitrogen) was added to reach the total volume of 20μl, which was then heated to 95°C for 10mins and briefly centrifuged. A precast 4-12%

Bis-Tris SDS-PAGE gel (Invitrogen) was prepared by removal of the well comb and adhesive strip, and the wells were flushed with prepared 1xMES Running Buffer by 20-fold dilution of 20xMES Running Buffer (Invitrogen). The gel was then placed into an assembled XCell SureLock electrophoresis tank (Invitrogen) and locked into place creating water-tight inner and outer reservoirs. 200ml 1xMES Running Buffer was poured into the inner reservoir and 500µl Antioxidant (Invitrogen) was added. 15µl of each sample was then loaded into each well of the gel. 10µl of Spectra Multicolor Broad Range Protein Ladder (Thermo Fisher) was loaded into either well at the side of the gel. The outer reservoir was then filled with 600ml 1xMES Running Buffer, and the lid was fitted onto the tank. The gel was run for 40mins at 200V until the band reached the bottom of the gel.

2.16.2 SDS-PAGE Gel Staining

Following the SDS-PAGE gel electrophoresis, gels for staining were removed from the cassette, washed in sterile water three times and then stained with 10ml per gel of SimplyBlue™ SafeStain (Invitrogen) solutions (1hour at room temperature with agitation), and then the solution removed. 100ml per gel of sterile water was used to destain the gel by incubation at room temperature with agitation, until the background turned into blank and the stained bands on the gel had reached the required clarity.

2.16.3 Western Blotting of Proteins

Following SDS electrophoresis, the protein gel, six squares per gel of 3MM

filter papers (Whatman) and one square per gel of nitrocellulose blotting membrane (Amersham) were soaked in transfer buffer (48mM Tris; 39mM Glycine; 0.04% (w/v) SDS; 20% (v/v) Methanol) for at least 5-15 minutes. The transfer stack was then prepared from the bottom in the sequence of one piece of sponge, three sheets of the pre-soaked filter paper, the pre-soaked nitrocellulose membrane, the equilibrated gel and anther three sheets of pre-soaked filter paper, and finally the other cover sponge on the top. Air bubbles were removed by rolling a pipette over the surface of each layer. The transfer step was carried out at 12V for 1 hour using BioRad Trans-Blot SD System, followed by membrane being blocked in filtered 5% (w/v) non-fat milk PBST solution (137mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 2 mM KH_2PO_4; 0.1%(v/v) Tween-20; pH 7.4) for 1 hour at room temperature under aggregation. Primary antibody was then added into the solution at appropriate dilute (1:100 – 1:5000) before incubation continued overnight under aggregation. The membrane was subsequently washed 4-6 times in PBST and then incubated for 1 hour at room temperature under aggregation, supplement with rabbit HRP-conjugate at 1:2500 dilution as secondary antibody. The membrane was washed in PBST for an additional 4-6 times and ready for signal development.

2.16.4 Western Detection

All western detection steps were carried out in a darkroom. The blot was incubated for 5mins in working solution consisting equal parts of the Stable Peroxide Solution and the Luminol/Enhancer Solution (Thermo Fisher

Pierce). Wrapped carefully with clingfilm without any excess liquid or bubbles; the protected membrane with protein side facing up was then used for explosion in the film cassette by placing a piece of Hyperfilm (Amersham) on top of the membrane. Exposure time was set as 60secs unless stated otherwise. The film after exposure was developed using developing solution until bands were visible. Following washing in water to remove any remaining solutions, the film was fixed in fixing solution until all of the film was dark in colour. Excess solutions were removed by washing in the water，and the film was hung to dry prior to photographing.

2.17 CONFOCAL MICROSCOPY

Experimental samples imaged on a Leica TCS SP2 AOBS confocal scanning microscope, equipped with a 100mW multi-line Argon laser (458nm, 476nm, 488nm, 496nm and 514nm) and a 1mW He-Ne Laser (543nm) as excitation sources. Signal was collected by the SP scanner (Table 2.11). The Leica confocal software (LCS) was used to operate the microscope and perform photographing.

Table 2.11 Excitation and Emission Settings.

Florescent protein used	Excitation	Emission Collected
Cholorophyll	488nm	650-700nm
eGFP (green)	488nm	500-520nm
YFP (yellow)	514nm	520-540nm
mRFP (red)	543nm	620-640nm

CHAPTER 3 VALIDATION OF MS1 PROTEIN-PROTEIN INTERACTIONS

3.1 INTRODUCTION

Previous research strongly supports that MS1 promotes pollen development by regulating gene expression (Wilson et al., 2001; Ito et al., 2007; Yang et al., 2007). Eukaryotic transcriptional regulation modulates gene expression in complicated ways， most transcription factors do not work alone, but rather function as a protein complex. For instance, certain transcription factors could promote or inhibit the efficient recruitment of RNA polymerase, by binding to specific DNA regulatory sequences, which always require the assistance of their interacting proteins (Margueron and Reinberg, 2011). Hence, identification of the putative interacting proteins proves to be invaluable for the characterization of transcription factors. Various approaches can be used to identify and validate protein-protein interactions (Table 3.1).

The yeast two-hybrid system is an *in vivo* method extensively used to investigate the existence of protein-protein interactions or identify new interacting candidates of a known protein (Chien et al., 1991). The protein of interest is fused to the DNA-binding domain (DB) of GAL4, the yeast transcriptional activator protein, serving as the “bait”. Meanwhile the putative interacting protein fused to the GAL4 activation domain (AD) would serve as the “prey”. Physical association of the proteins activates the promoter of the reporter gene through the GAL4 binding sites (Figure

3.1), and therefore give rise to the transcription of the reporter gene. The selectable markers *HIS3* and *lacZ* are commonly used reporter genes. *HIS3* gene allows the transcription activation to be monitored by enabling growth of cells on plates lacking histidine and containing 3-AT. Induction of the *lacZ* gene results in a blue colour when assayed with X-gal (5-bromo-4-chloro-3-indolyl-â-D-galactopyranoside).

Table 3.1 Protein-Protein Interaction Methods.

Approaches	Description
Yeast-2-Hydrid (Y2H)	Two putative interacting proteins are respectively used as the "bait" and the prey, and then co-transformed into yeast. Neither of the fusion protein alone is sufficient to activate the transcription of reporter genes. Only after association between the two proteins occur does transcription of the reporter gene become activated (Joung et al., 2000).
Protein Pull-down	Proteins are co-expressed in *Escherichia coli* (*E. coli*), and then purified using an antibody specific to one of the proteins, or to a tag it is fused with. Subsequent western analysis using an antibody specific to its putative interacting protein confirms the presence of the interaction if they two has been co-purified as a protein complex (Sambrook and Russell, 2006).
Bimolecular Fluorescence Complementation (BiFC)	This involves two proteins respectively fused with the C-terminus and the N-terminus of the YFP protein and then co-expression in plant cells, either onion epidermis/ protoplasts or *Nicotiana benthamiana* leaves. YFP fluorescence can be detected if the proteins physically interact with each other, and therefore form a integrated YFP protein (Hu et al., 2002).
Fluorescence Resonance Energy Transfer (FRET)	Different fluorochromes are attached to either protein, and co-expressed in living cells, such as the *N. benthamiana* leaves. The fluorescent the "acceptor" molecule is excited by transfer of energy from the "donor" (Ecker et al., 2004).
Co-Immunoprecipitation (Co-IP)	Using the antibody specific to a known protein to pull-down the particular protein complex, out of a cell lysate pool of various proteins.

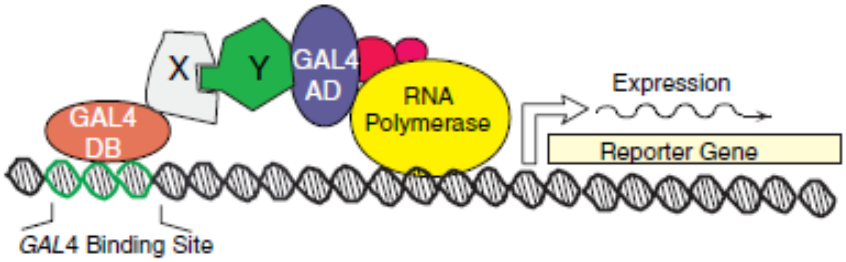


Figure 3.1 Principle of the yeast two-hybrid system (ProQuestTM Two-Hybrid System with Gateway® Technology, Invitrogen). Yeast cell expressing both the GAL4 DB-X fusion protein and the GAL4 AD-Y fusion protein. When X and Y physically interact with one another, the GAL4 AD-Y fusion protein is localized to the promoter, and transcription is activated.

To identify the MS1 putative interacting proteins, Yang and Wilson (unpublished data) generated and screened an *Arabidopsis thaliana* stamen–specific yeast-2-hybrid library comprising all stages of anther and filament tissue, using the MS1 protein. The MS1 protein was deleted for the 3' PHD-motif due to auto-activation by this sequence (Yang and Wilson, unpublished). Using MS1 as the "bait" protein, a number of candidates have been identified as putative MS1-interacting proteins.

Clone 54 (At1g58210/NET2A), termed as Y2H54, gave the strongest interaction with MS1 and comprised 50% of the identified clones; and an alternative protein clone 19 (AT2G46260/ LRB1), termed as POB2, also exhibited a strong interaction (Figure 3.2). The Y2H54 protein belongs to the NET actin-binding superfamily (Deeks et al., 2012), the POB2 protein with the BTB/POZ domain previously identified as getting involved in

protein ubiquitylation through the interaction with the CULLIN 3 proteins (Gingerich et al., 2005). Both proteins display expression changes through the pollen grain germination and pollen tube growth progress (Wang et al., 2008), suggesting their engagement in microsporogenesis.

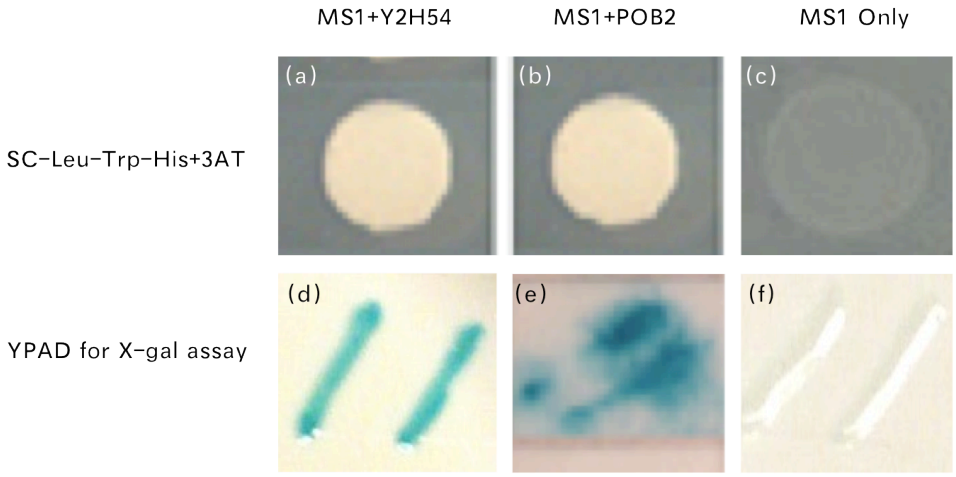


Figure 3.2 MS1 interaction proteins demonstrated by yeast 2 hybrid (Yang and Wilson, unpublished). The MS1 protein without the PHD-finger motif was used as the "bait", and a stamen specific library containing the anther and filament tissues of all stages was screened. The interaction between MS1 and other proteins lead to the activation of the transcription of the reporter genes, *HIS3* and *LacZ*. Yeast strains were patched from isolated colonies onto an SC-Leu-Trp master plate and incubated for 18hours at 30°C. 8 candidates were identified as putative MS1-interacting proteins (data not shown), among which clone 54 (Y2H54) and clone 19 (POB2) showed strong interactions (a, b, d, e). (a) and (d), isolated colonies were plated onto SC-Leu-Trp-His+3AT (10mM) plate. Yeast cells were grown on the plate lacking Histidine, indicating the effective activation of the *HIS3* reporter gene. (b) and(e), cells of Y2H54 and POB2 from the SC-Leu-Trp-His+3AT (10mM) plate were plated onto the YPAD plate containing a nylon membrane for X-gal assay, and the blue color suggested the activation of the reporter *gene LacZ.* (c) and (f), yeast negative control cells transformed only with MS1 as the "bait" were neither capable of growing on the SC-Leu-Trp-His+3AT (10 mM) plate (c), or give a blue colour for the X-gal assay (f).

Through we attempted to avoid the auto-activation problem by eliminating the PHD-finger domain of the MS1 protein, the presence of false positive is still one of the major issues when using the yeast-2-hybrid technology. Occasionally, proteins of interest can non-specifically associate with multiple types of proteins, giving rise to the identification of the false positives. Therefore, validation of the protein-protein interactions using additional alternative approaches is critical for confirming the "true" interaction between proteins. Two commonly used methods, FRET and the protein pull-down assay, are performed in this chapter.

The Fluorescence Resonance Energy Transfer (FRET) as the sensor of protein-protein interactions in living cells provides direct proof of physically association of the proteins (Ecker et al., 2004). In the FRET assay, energy of the "donor" such as the fluorochrome in the electron excited state can be transferred to the "acceptor" fluorochrome, once they are within close proximity. This therefore causes the increase in the emission intensity of the "acceptor", yet at the cost of reduction in the fluorescence intensity of the "donor" as well as its lifetime in the excited state. In case of studying intermolecular interaction between two separate proteins, protein partners are respectively fused to suitable "donor" and "acceptor" fluorescent dyes. This usually involves two different genetically encoded fluorophores such as the CFP and the YFP, and then co-expression in living cells. FRET can occur when the two chromophores come in close

spatial proximity, providing direct evidences for protein-protein association (Figure 3.3) (Bhat et al., 2006).

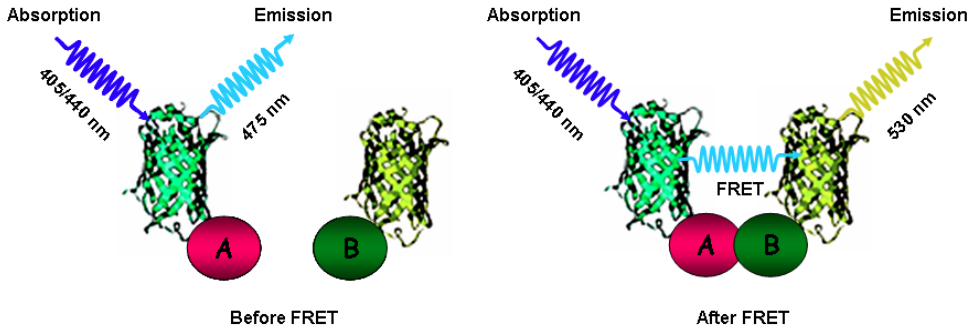


Figure 3.3 Detection of protein-protein interactions via FRET (Bhat et al., 2006). FRET between cyan fluorescent protein (CFP) as a donor fused to protein A and yellow fluorescent protein (YFP) fused as an acceptor to protein B. Under favorable spatial and angular conditions, interaction between A and B causes a decrease in the intensity of donor (CFP) fluorescence concomitant with an increase in acceptor (YFP) fluorescence. CFP and YFP are depicted as cyan and yellow ribbon models fused to putative interacting proteins A and B, respectively.

The protein pull-down assay *in vitro* is a method to determine the interactions between known interacting partners. In the pull-down assay, the "bait" protein is fused with a particular tag and then immobilized with affinity ligands specific for the fusion tag, which therefore forms the "second affinity support" for the interacting partner protein (Figure 3.4) (Green and Sambrook, 2012). This is then incubated with the source of the "prey" protein, such as cell lysate. The protein complex pulled-down by the antibody specific to the "bait" can be subsequently detected by western blot using the antibody specific to the "prey" protein or the tag fused to it. In this chapter, MS1 and its putative interacting partners were fused to commonly

used glutathione-S transferase (GST) tag and 6× histidine (6×His) fusion tag, respectively, since no effective antibodies specific to the MS1 protein or its putative partners have been developed.

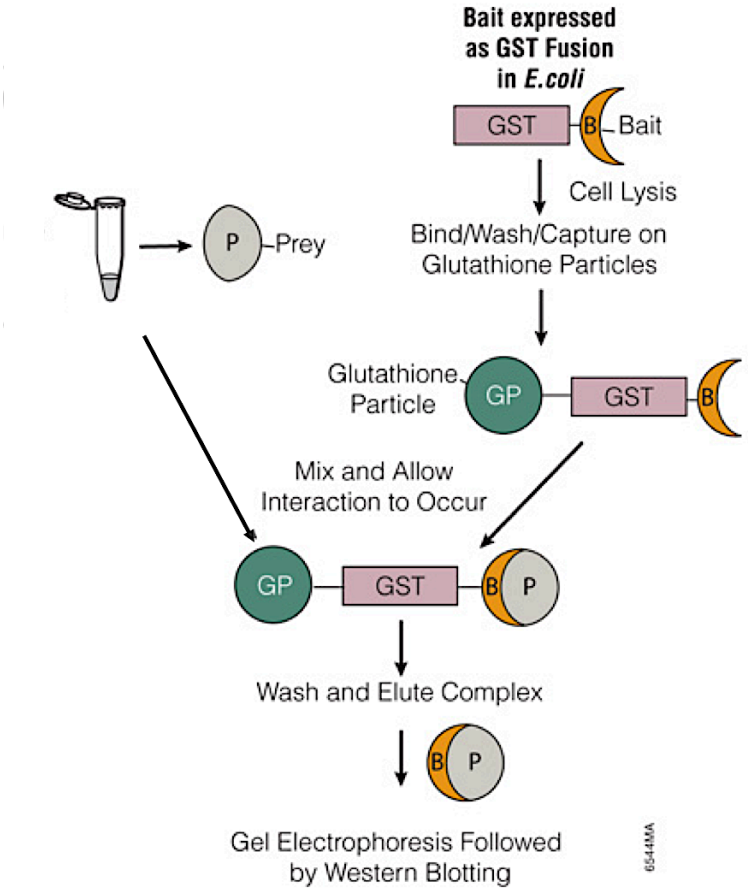


Figure 3.4 Schematic of pull-down assay using bacterial expression of bait protein and the prey protein (Promega, 2009).

3.2 MATERIALS AND METHODS

3.2.1 Generation of Constructs

3.2.1.1 Constructs for transient expression *in planta*

Entry clones of the *Y2H54* and *POB2* genes containing the full-length cDNA were kindly provided by Dr. Caiyun Yang (University of Nottingham), using Gateway entry vector pENTR/D-TOPO(Invitrogen). Each of them was then recombined into the Gateway compatible destination vector pUBC-DEST-RFP (Appendix III), by the LR reaction (Section 2.9.4), which contained mRFP at the C terminus of the gene inserted (Grefen et al., 2010).

The LR reaction products were subsequently transformed into *E. coli* DH5α chemically competent cells (Section 2.11), and screened on LB medium containing 100μg/mL spectinomycin (Section 2.10.1). Plasmids were extracted from colonies containing the correct recombinant constructs (Section 2.12), identified by colony screening (Section 2.9.5) using selective primers (Table 3.2), which were then transformed into the *Agrobacterium* competent cells GV3101 with pSoup by electroporation (Section 2.14).

Table 3.2 Primers Used for Colony Screening of Constructs for FRET.

Primer Name	Primer Origin	Primer Sequence (5'-3')
RFP_R2	pUBC-DEST-RFP	GAGCCGTACTGGAACTGAGG
MS1_5168L	*MS1*	TGAGTGTGGAGCAACGGAAG
54_58210_2443F	*Y2H54*	GCTCAATCAGGTGGCAATAA
Pob2_1402F	*POB2*	CTCGCCCGCAATGTGTAGTG

The *MS1* gene was previously overexpressed under the control of the cauliflower mosaic virus 35S promoter (CaMV35S) in a gateway binary vector PGWB5 (Yang et al., 2007). This verified construct, previously transformed into *Agrobacterium* C58 strain, contained the eGFP at the C terminus of MS1 and was used for the FRET assay.

3.2.1.1 Constructs for protein expression in *E. Coli*

To generate the constructs for protein pull-down assays, full length cDNAs of the *MS1, Y2H54* and *POB2* without the stop codon were cloned from the bud cDNA of *Arabidopsis thaliana* L*er* (Section 2.3-2.6), using primers with restriction sites added at the 5' ends (Table 3.3). PCR products of all three genes as the inserts were digested with appropriate restriction enzymes (Table 3.3) (Section 2.8.2), and then ligated into each single vector as described in Table 3.4. Each of vectors (Appendix IV and V), pET28a, pET30a, pET41a (Novagen) and pGEX-4T-1, pGEX-6P-1 (GE Healthcare Life Sciences), was previously linearized using two enzyme pairs, respectively, *Bam*H1 plus *Eco*R1 and *Sal*1 plus *Eco*R1. Sequential digestion was performed (Section 2.8.2) in which the desired DNA was first incubated with *Eco*R1 for 1hour, and then either *BamH*1 for 1hour or *Sal*1 for 10hours.

Table 3.3 Primers Used for Protein Expression Constructs.

Gene	Enzyme	Primer Name	Primer Sequence (5'-3')
Y2H54	BamH1	54_1_F	CGC*GGATCC*GCGATGTTGCAG AGAGCAGCGAGCAA
	EcoR1	54_End _R	CCG*GAATTC*CGGTTATTCAGG GAGCTTCCCAGGTGGCCT
POB2	BamH1	POB2_1 _F	CGC*GGATCC*GCGATGAGAGG TTCCAATAACACCGATCTA
	EcoR1	POB2_End _R	CCG*GAATTC*CGGTCAGTGCAG GTCTGAGGAACGTTTA
MS1	EcoR1	MS1_1 _F	CG*GAATTC*CGGATGGCGAATC TGATTCGAACA
	Sal1	MS1_End _R	CGC*GTCGAC*GTCGGCCATAGC GGCCGCGGAATTAGGGTAAA AAAGAGAGAGGAATAA

Table 3.4 Vectors Used for Recombinant Protein Expression in *E. coli.*

Vector	Fusion Tags	Selection Antibiotic	Cloning Host Strain	Expression Host Strain
PET 28a	N-His; C-His	Kan	DH5α	BL21 (DE3) STAR
PET 30a	N-His; C-His	Kan	DH5α	BL21 (DE3) STAR
PET 41a	N-GST; C-His	Kan	DH5α	BL21 (DE3) STAR
pGEX-4T-1	N-GST	Amp	DH5α	BL21 (DE3) STAR
pGEX-6P-1	N-GST	Amp	DH5α	BL21 (DE3) STAR

The ligation products were subsequently transformed into *E. coli* DH5α chemically competent cells (Section 2.11), and screened on LB medium containing desired selective antibiotics (Section 2.10.1). Plasmids were extracted from colonies containing the correct recombinant constructs

(Section 2.12), identified by colony screening (Section 2.9.5) using selective primers (Table 3.5), which were then transformed into the BL21 (DE3) STAR ™ One Shot Chemically Competent *E. coli* (Invitrogen) (Section 2.11).

Table 3.5 Primers Used for Colony Screening of Constructs for Protein Expression.

Primer Name	Primer Origin	Primer Sequence (5'-3')
T7_Term	pET vectors	GCTAGTTATTGCTCAGCGG
pGEXstop	pGEX vectors	GGCAGATCGTCAGTCAGTCA
MS1_5168L	*MS1*	TGAGTGTGGAGCAACGGAAG
54_58210_2443F	*Y2H54*	GCTCAATCAGGTGGCAATAA
Pob2_1402F	*POB2*	CTCGCCCGCAATGTGTAGTG

3.2.2 Transient Expression in *Nicotiana benthamiana* Epidermal Cells

Nicotiana benthamiana seeds were grown at 22°C+/-2°C in pots containing Levington M3: John Innes No.3: vermiculite: perlite (6:6:1:1) compost immersed in 0.2% (w/v) Intercept® (Scotts). Plants were grown for 5-6 weeks at which point they were used for transient assays. Constructs containing genes of interest and the p19 suppressor from tomato bushy stunt virus (TBSV) were respectively transformed into *Agrobacterium* competent GV3101 with pSoup (Section 2.14) (Hellens et al., 2000; Voinnet et al., 2003). Single colonies of recently transformed cells were cultured in 10ml

of liquid LB medium at 28°C with selection antibiotics added, including rifampicin (for GV3101 strains), tetracycline (for pSoup plasmids) plus either of kanamycin (for p19 plasmids) or spectinomycin (for pUBC-DEST-RFP plasmids). Following overnight growth, cells were pelleted by centrifguation at 3,000g for 10mins and then resuspended in solutions containing 10mM $MgCl_2$ and 0.19% (w/v) MES. Cultures were diluted to OD_{600} of 0.4 and acetosyringone added to the final concentration of 20nM. Following overnight co-incubation at room temperature, cells containing genes of interest and those transformed with p19 were co-infiltrated into the abaxial leaf surface by pressure inoculation using a blunt 5ml syringe. Plants were then incubated at 22°C for 3-5 days before experimental analysis.

3.2.3 Co-Localisation and Förster Resonance Energy Transfer

Several 0.5-0.7cm sections of leaf tissues were cut from transformed *N. benthamiana* plants and rinsed in water prior to imaging. Coexpression of the two constructs and subcellular localisation of the two proteins was confirmed by scanning using a confocal microscope. The FRET specimen was generated by co-transformation of MS1-eGFP and Y2H-RFPs，plus two reference specimens of tobacco leaf tissues expressing either MS1-eGFP or Y2H-RFP serving as GFP reference and RFP reference, respectively. The

samples were initially observed under UV light to identify transformed cells, and then imaged under Leica SP2 confocal microscopy (Section 2.17). The FRET sensitized emission method was adapted from the manufacturer's instructions (Leica). In the donor channel setting, the acoustooptic tuneable fibre (AOTF) for the 488 nm laser was set to 15% and the photomultiplier tube (PMT) voltage for GFP was adjusted to slightly below detector saturation. In the acceptor channel, AOTF for the 543 nm laser and the PMT voltage for RFP were also adjusted to just below detector saturation. The laser AOTF and detector PMT were fine-tuned to avoid detector saturation for every sample specimen. The acceptor and donor channel settings were saved as defaults and all other specimens were examined under exactly the same conditions. Once both the AOTF and PMT setting were optimised for all specimens (FRET and two references), FRET images of each sample were recorded according to the manufacturer's instruction. Briefly, when laser light excited the donor selectively, the FRET signal resulted from the energy transfer was detected in the acceptor florescence channel. The donor and acceptor references were used to generate correction factors for excitation and emission cross-talk corrections. The donor only, acceptor only, FRET and background were subsequently chosen in the FRET application wizard (FRET Sensitized Emission method, Leica) to calculate FRET efficiency.

3.2.4 Optimization of Protein Expression in *E. coli*

Recombinant constructs were individually transformed into BL21 (DE3) STAR ™ One Shot Chemically Competent *E. coli* (Invitrogen) for recombinant protein expression. Induction of protein expression was based on the manufacturer's instructions. One or two transformants were cultured in liquid LB medium containing the desired antibiotics (Table 3.4) with shaking at 37°C until the OD_{600} reached 0.6 to 1, which were then diluted by 20-fold using fresh liquid LB medium with desired antibiotics. Upon reaching mid-log phase (OD_{600}=~0.4, 2 to 3hours), IPTG was added to half of the cultures (final concentrations of 0.1mM, 0.2mM, or 1mM were tested), whist the other half was set aside as the non-induced control. Both samples were kept incubating with shaking (37°C, 28°C, and room temperature (21°C) were tested). Samples were harvested individually at different time points (0hour, 2hours, 4hours, 6hours, and overnight) for the following analysis.

3.2.5 Protein Extraction

The cell cultures were centrifuged at 4,000g at 4°C for 20mins to discard the supernatant. 3ml of ice-cold Lysis Buffer (150mM NaCl; 5% Glycerol; 1mM EDTA, pH8.0; 50mM Tris-Cl, pH8.0) containing 80 µl of 10mg/ml lysozyme and 4µl of 100mM PMSF was added into every 1g of the pellets to resuspend them. The samples were incubated on ice for at least half an hour, and then sonicated on ice for pulses of 15 sec at 6 micron amplitude (40 times, with 30 sec between pulses) in a Soniprep 150 sonicator (Shah et

al.), until it turned into the colour of greyish cream. The cell lysates were treated at 4°C for 30mins by adding DNase and RNaseI (Invitrogen) to a final concentration of 5 μg/ml, and subsequently centrifuged at 4°C for 30mins at 4,000g to separate the soluble and insoluble fractions. The pelleted inclusive bodies were then washed twice in Wash Buffer (150mM NaCl; 5% Glycerol; 0.5% Triton X-100; 50mM Tris-Cl, pH8.0,) and centrifuged at 4°C for 10mins at 4,000g after each wash. Small portions of the soluble and insoluble fractions were respectively mixed with SDS Loading buffer for SDS PAGE electrophoreses analysis (2.16.1). In case of large scale of experiments, the most efficient conditions of IPTG concentration, inducing temperature and time were retained for protein expression.

3.2.6 Resolubilisation and Refolding of Proteins

The insoluble fraction was then dissolved in 1ml of Solubilisation Buffer (8M Urea; 50mM Tris-Cl, pH8.0; 5% Glycerol) and incubated at 4°C for 1 hour. The solution was then centrifuged at 4°C for 20mins at 13,000g. Protein concentration was estimated using a NanoDrop 2000 spectrophotometer (Thermo Fisher) at OD_{280}. Solubilisation Buffer was then added to the sample to achieve a final concentration of 10mg/ml. This solution was transferred into freshly prepared dialysis membrane (Sigma D9777), which was cut to size depending on the sample loaded and both ends clipped. The bag was then submerged in 100-fold volume of ice-cold

Refolding Buffer (50mM Tris-Cl, pH7.8; 1mM $CaCl_2$; 0.8M L-arginine) containing glycerol in a gradient of concentrations (5%, 10%, 20%, and 30% were tested) to achieve highest recovery rate. After an overnight incubation at 4°C with stirring, the solution inside the bag was then transferred into a new tube for spinning. 10µl of the supernatant was mixed with SDS loading buffer for subsequent SDS PAGE electrophoresis (2.16.1) and staining analysis (Section 2.16.2). In case of large scale of protein extraction, the most suitable glycerol concentration in Refolding Buffer, that obtained highest rate of recovery, was retained for refolding proteins.

3.2.7 Protein Pull-down Assay

The refolded GST tagged protein was transferred into a 2ml low binding tube (Eppendorf) containing 50µl of Pierce Glutathione Magnetic Beads (Thermo Fisher) resuspended in 1ml of PBS buffer (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4，2mM KH_2PO_4, pH8.0) with 10µl of 100mM PMSF added. After incubation on a rolling platform for 1hour at room temperature to gently mix, the tube was placed on the magnet until the liquid turned clear. The supernatant was discarded and the beads washed four times in 300µl of PBS buffer. The solution was placed on the magnet after each wash to remove the supernatant. The refolded protein extract tagged with 6xHis was subsequently mixed with the beads in the Binding/ Wash buffer

(125mM Tris-Cl, pH8.0; 150mM NaCl; 5% (v/v) Glycerol) with 3µl of 100mM PMSF added, and incubated for 1 hour at room temperature (21°C) on a rolling platform. After the incubation, the beads were washed four times in Binding/Wash Buffer and supernatant removed after each wash by placed on the magnet rock until the solution turned clear. To elute the proteins, 50µl of Elution Buffer (50mM Reduced Glutathione in Binding/Wash Buffer) was added to the beads prior to incubation at room temperature for 5 minutes. This was repeated 3 times and targeted proteins pooled in a new tube for subsequent SDS PAGE electrophorese (2.16.1) and gel staining analysis (Section 2.16.2).

Sequential western blotting assay was performed to confirm the existence of protein interaction (Section 2.16.3). Rabbit polyclonal 6xHis tag® antibody (HPR) (Abcam, ab1187) was used as the primary antibody and diluted 1:4000. Pierce® Anti-rabbit HRP-conjugate (Thermo Fisher) at 1:2500 dilution served as the secondary antibody. The film was exposed for 60secs prior to photographing.

3.3 RESULTS

3.3.1 Validation of Interaction by FRET

3.3.1.1 MS1 is transiently expressed in *Nicotiana benthamiana* with high efficiency

The previously verified PGWB5::35S:MS1-eGFP construct (Yang et al., 2007), transformed into the *Agrobacterium* C58 strains, was used to test the efficiency of transient expression in tobacco epidermal cells. In more than 90% of the cells checked, clear and strong nuclear localised expression of the eGFP fused MS1 protein was observed by confocal microscopy (Figure 3.5), which indicates the high efficiency of transient expression in tobacco cells.

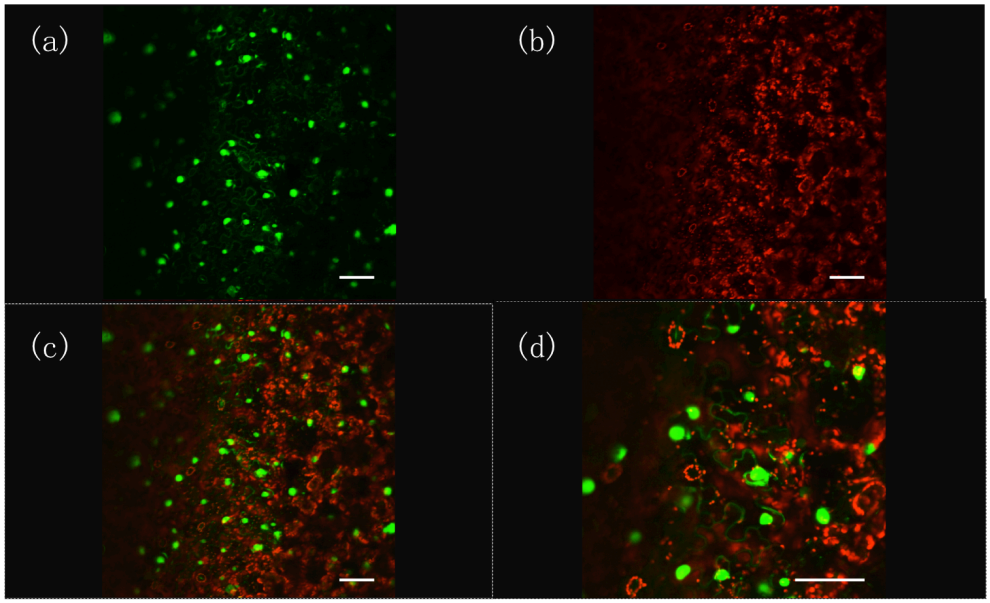


Figure 3.5 Transiently expressed MS1-eGFP in the nuclei of tobacco epidermal cells. Leaves were harvested 4 days after Agrobacterium-infiltration transformation. Images were taken using sequential scan confocal microscopy. (a) nuclear-localised MS1-eGFP (green, excitation 488 nm, emission collected 500-520 nm); (b) image of the chlorophyll in the chloroplasts (red, excitation 488nm, emission collected 650-700nm); (c) and (d) merged images of MS1-eGFP and chlorophyll images. (a) (b) and (c), 20x magnification; (d), 40x magnification . Bar= 500 μm.

3.3.1.2 Co-localisation of MS1 and Y2H proteins

Transient transformation by Agrobacterium-infiltration was used to determine whether the putative MS1-interacting proteins showed co-localisation with the MS1-fusion protein. Nuclear, co-localisation of POB2-RFP and MS1-eGFP was observed in the tobacco epidermal cells (Figure 3.6). However, only membrane-associated expression was observed in the Y2H54 transformed cells (Figure 3.7). This suggests that it is less likely that the Y2H54 co-localises with MS1. This implies that the identified

yeast 2 hybrid interaction between MS1 and Y2H54 could be a false positive, although it is possible that the tobacco cells lack certain factors, that are specific to the tapetum, which are capable of localising the Y2H54 transcript to the nucleus. Nevertheless, Y2H54 was excluded from subsequent analyses.

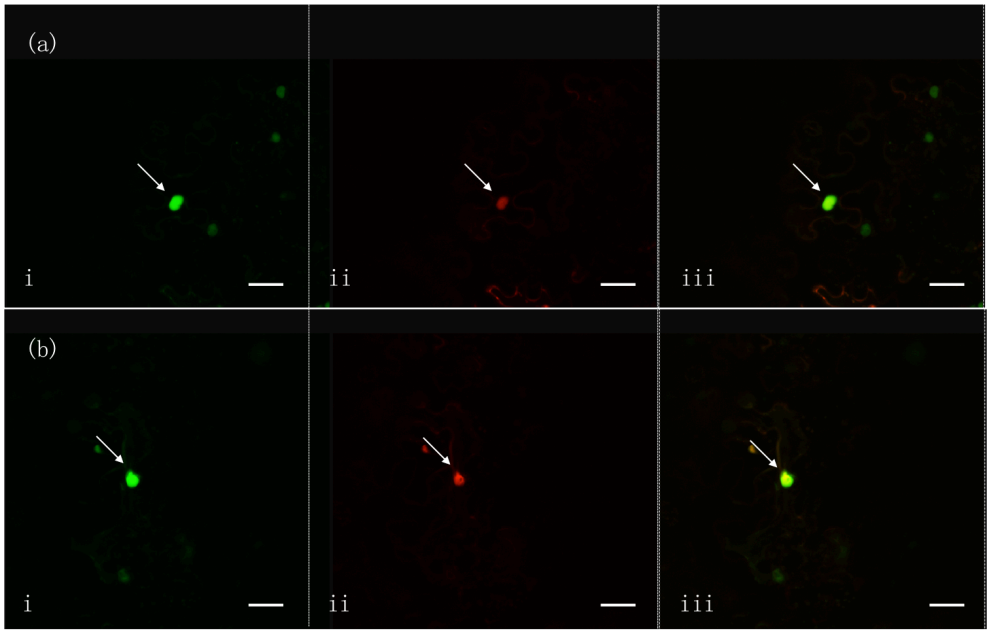


Figure 3.6 Co-localisation of MS1 and the Y2H protein POB2 when co-expressed in the tobacco leaves. Images were taken using sequential scan confocal microscopy. (a) and (b) Two individual tobacco epidermal cells. i, florescence from MS1-eGFP (excitation 488 nm, emission collected 500-520 nm); ii, florescence from POB2-mRFP (excitation 543 nm, emission collected 620-640 nm); iii, overlay of GFP and RFP. Arrows point at tobacco nucleus. Bar=100 μm.

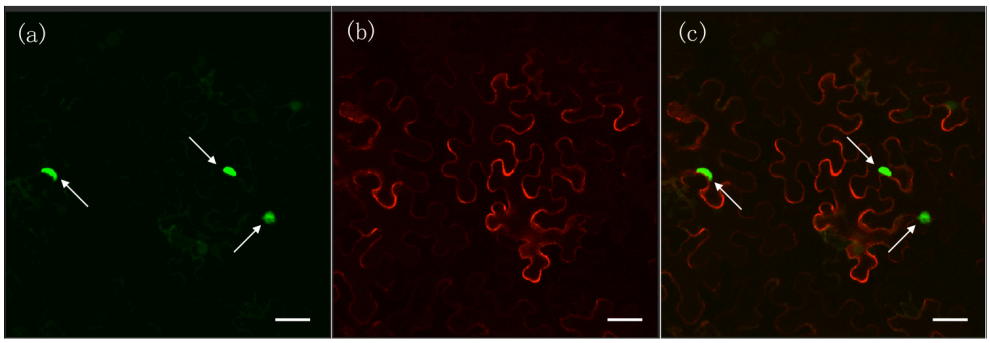


Figure 3.7 Subcellular localisation of MS1 and Y2H54. Images were taken using sequential scan confocal microscopy. (a) Nuclear-localised MS1-eGFP (green, excitation 488 nm, emission collected 500-520 nm). (b) Membrane-localised Y2H54-mRFP (red, excitation 543 nm, emission collected 620-640 nm). (c) Overlay of GFP and RPF images. Arrows point at tobacco nuclei. Bar = 100 μm.

3.3.1.3 FRET analysis of the interaction between MS1 and POB2 proteins

Previously a Yeast 2 hybrid assay indicated the existence of physical interaction between POB2 and MS1 (C. Yang and Z. A. Wilson, unpublished). POB2 is a transcription factor belonging to the BTB/POZ domain family, which has been demonstrated to be co-localised with MS1 in the nucleus by transient expression in the epidermal cells of tobacco leaf tissues (Figure 3.6). To determine the association of the two proteins *in planta*, FRET measurements of MS1-eGFP and POB2-mRFP were performed on the transformed tobacco leaf epidermal cells. To allow adequate expression of the proteins and therefore to achieve better protein

interactions, plants were allowed to grow for 4 days after transformation before being checked by confocal microscopy. FRET efficiency between MS1-eGFP and POB2-mRFP, determined by the FRET application wizard (FRET Sensitized Emission method, Leica), was measured in 12 independent transformations (Figure 3.8), and the average value of multiple regions (ROI) calculated from each measurement. Y2H54 had been previously shown to not co-localise with MS1 (Figure 3.7); cells co-transformed with MS1 and Y2H54 were used to measure the cross-talk between the proteins, which showed an average FRET efficiency of 0.374%.

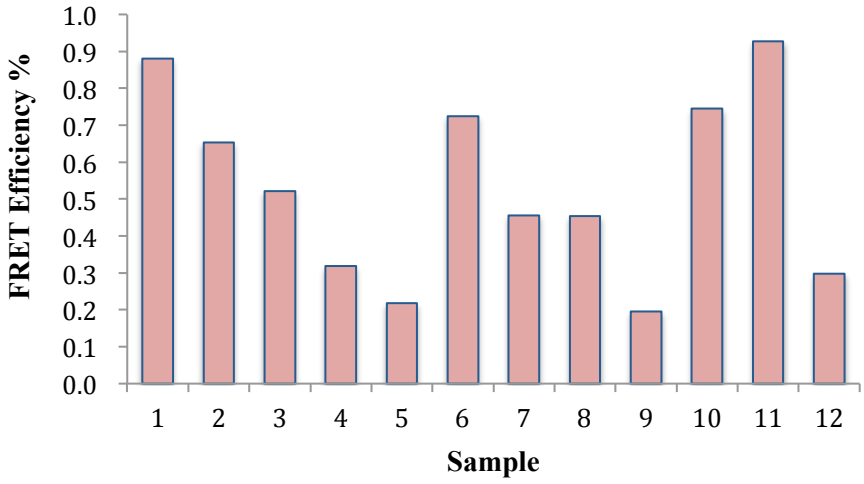


Figure 3.8 All FRET experiments for MS1-eGFP and POB2-mRFP. FRET efficiency between MS1-eGFP and POB2-mRFP was measured in 12 independent transformed cells.

In the typical FRET experiments, MS1-eGFP and POB2-mRFP were co-expressed in the same tobacco cell. Briefly, under the excitation spectra of the donor GFP，if the protein partners of interest could interact, GFP emission would be dampend whereas the acceptor RFP sensitized. However, no expected emission of the acceptor RFP was detected upon the excitation of the donor GFP (Figure 3.9). Samples co-transformed with MS1-eGFP and POB2-mRFP did not show significant differences from those used as the negative references, in which the two proteins were expressed separately. The FRET efficiency values were after adjustment for excitation and emission cross-talks, as low as 0.533% in the samples expressing both proteins (Figure 3.8). This suggested that energy transfer (FRET) between MS1-eGFP and POB2-mRFP probably did not occur, and therefore it is unlikely that physical interaction was occurring between the MS1 and POB2 proteins in the tobacco epidermal cells.

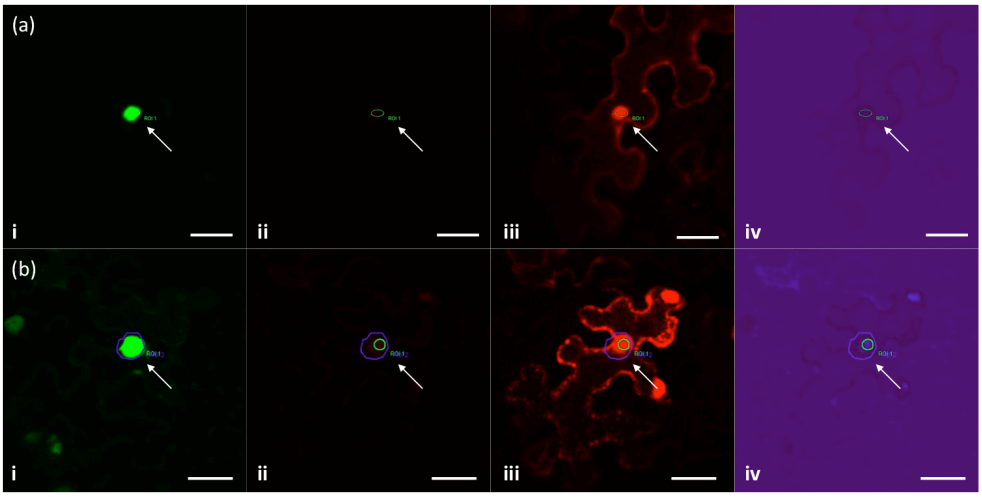


Figure 3.9 FRET report for two individual experiments. (a) and (b), two individual tobacco epidermal cells. i, Donor (Donor exc.), donor (GFP) emission under donor (GFP) excitation; ii, Acceptor (Donor exc.), acceptor (RFP) emission under donor (GFP) excitation; iii, Acceptor (Acceptor exc.), acceptor (RFP) emission under acceptor (RFP) excitation; iv, FRET Efficiency calculated by the FRET application wizard according to manufacturer's instructions (Leica).

3.3.2 Determination of Protein Interactions by Protein Pull-Down Assays

3.3.2.1 Protein expression and purification

Levels of protein expression in *E. coli* are determined by multiple factors, including the temperature, the inducer concentration, the host codon preference, etc. (Green and Sambrook, 2012). To achieve the highest yield of protein expression in *E. coli* BL21 (DE3) STAR strains, a collection of conventional vectors were used to clone genes of interest (Table 3.4). Considering the Y2H54 protein was membrane-localised as indicated in the FRET analysis (Figure 3.7) and therefore less likely to interact with the

MS1 protein, it was excluded from subsequent protein expressions. The conditions for protein expression were optimized through pilot experiments (data not shown); optimal expression was obtained using 0.2mM IPTG and overnight incubation at room temperature (21°C) with shaking at 200rpm.

Pilot experiments to optimize MS1 and POB2 expression in *E. coli* using a range of vectors were conducted in parallel (Table 3.6). Following a series of experiments using 20ml of cell cultures (data not shown), constructs stably expressing high level of proteins were chosen for large scale of protein expression. To meet the experimental requirements, the protein partners needed to be tagged differently, either the GST tag or the 6xHis tag; the pET28a construct was used to express the 6xHis tagged MS1 protein and the pGEX-4T-1 construct for the GST tagged POB2 protein.

Table 3.6 Construct Generation Summary[a].

	POB2					**MS1**				
	PET 28a	PET 30a	PET 41a	pGEX -4T-1	pGEX -6P-1	PET 28a	PET 30a	PET 41a	PGEX 4T-1	PGEX 6P-1
DH5a Transformation	+	+	+	+	+	+	+	+	+	+
Colony Screening	+	+	+	+	+	+	+	+	+	+
Plasmid Extraction	+	+	+	+	+	+	+	+	+	+
BL21 (DE3) STAR Transformation	+	-	+	+	-	+	+	+	+	-
Colony Screening	+			+		+	+		+	
Protein Expression	+			+		+	+		-	

a. '-', experiments that failed; Gap, experiments that were not conducted

Abundant protein bands of ~ 75kDa and ~80kDa were detected in the insoluble fractions, corresponding to the 6xHis-MS1 protein and the GST-POB2 proteins, respectively (Figure 3.10a). Protein extraction from the inclusive bodies was performed through full cell lysis by lysozyme and sonication, with the entire soluble fractions discarded by multiple wash steps (Section 3.2.6). The inclusive bodies were dissolved in denaturing buffers containing a high concentration of urea, which needed to be gradually diluted to refold the denatured proteins. To obtain best results from recovery of the native proteins, glycerol in gradient concentrations (5%, 10%, 20%, 30%) was added into each portion of Refolding Buffer to stablise the proteins during pilot experiments. SDS PAGE analysis suggested that the buffer with 5% glycerol added achieved the highest yield of protein refolding, which was then applied to large scale of experiments, in which 500ml of cells were used for protein extraction.

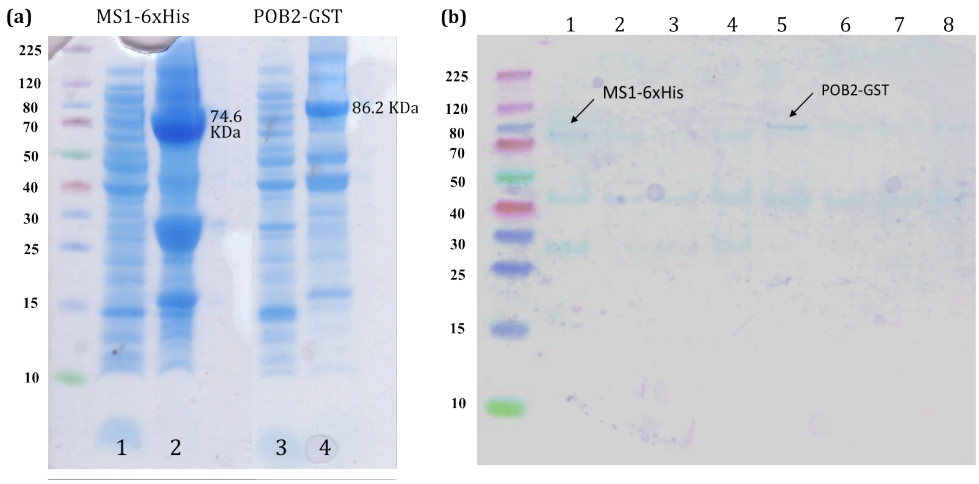


Figure 3.10 Pilot protein expression experiments. (a) Stained SDS PAGE analysis of expression from the pET28a vector containing the whole MS1 cDNA with an N-terminal 6xHis tag, and the pGEX-4T-1 construct containing the whole POB2 cDNA with an N-terminal GST tag. Enriched bands of ~75kDa and 85kDa respectively indicating that MS1 and POB2 are expressed from the vectors. Lane 1, Soluble fraction from the pET28a-MS1; Lane 2, Insoluble fraction from the pET28a-MS1; Lane 3, Soluble fraction from the pGEX-4T-1-POB2 ; Lane 4, Insoluble fraction from the pGEX-4T-1-POB2. **(b)** Stained SDS PAGE analysis of refolded MS1-6xHis tagged protein (lanes 1-4) and POB2-GST tagged protein (lanes 5-8). Proteins were purified from the inclusion bodies first (Section 3.2.5) and then resolubalised in buffers (section 3.2.6). Specific bands of ~75kDa and 85kDa respectively indicated refolded the MS1 and POB2 proteins. Lane 1-4, refolded extracts from the pET28a-MS1 protein; Lane 5-6, refolded extracts from the pGEX-4T-1-POB2 protein. Lane 1 and 5, 5% glycerol in refolding buffer; Lane 2 and 6, 10% glycerol in refolding buffer; Lane 3 and 7, 20% glycerol in refolding buffer; Lane 4 and 8, 30% glycerol in refolding buffer.

3.3.2.2 Protein pull-down assay

Refolded GST tagged POB2 protein was immobilized with affinity ligands specific to the GST tag, forming a “second affinity support” for POB2. As long as the interaction between MS1 and POB2 occured, 6xHis tagged MS1

proteins would also associate with the beads through the “second affinity”. In that case, the MS1 protein could be detected in the eluted protein complex by western blot when anti-6xHis tag antibody was used as the primary antibody. Two references were used as negative controls in parallel experiments, including elution from the GST beads with which the POB2 protein was immobilized, yet pulled-down with blank Binding Buffer; as well as the blank beads without any protein immobilized but pull-downed with His tagged MS1 proteins.

In the western blot analysis, a specific band representing a protein of ~74.6 KDa was identified as the MS1 protein (Figure 3.11). In the control samples, several other bands were detected, possibly representing non-specific bindings. This demonstrated the existence of the interaction between MS1 and POB2 proteins *in vitro*, and provided supporting evidence for the results generated by the yeast-2-hybrid analysis.

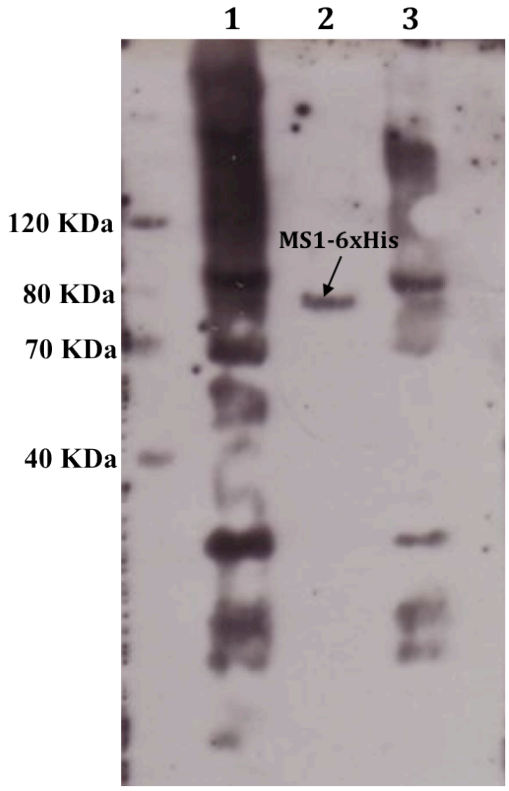


Figure 3.11 Analysis of protein pull-down. Western transferred blot, probed with 1:4000 rabbit polyclonal 6xHis antibody. In the pull-down sample, the specific band detected is ~74.6 KDa in size, identified as 6xHis tagged MS1 (black arrow). In the control samples, several other bands are detected possibly representing non specific bindings. 1, blank beads without POB2 immobilized but incubated with His tagged MS1 proteins in the pull-down assay; 2, protein pull-down sample, beads with POB2 immobilized and incubated with His tagged MS1 protein; 3, beads with POB2 immobilized, and subsequently incubated with blank binding buffer in the pull-down assay.

3.4 DISCUSSION

To further validate whether interactions are occurring between the MS1 putative interacting proteins Y2H54 and POB2, identified by yeast-2-hybrid library screening (Yang and Wilson, unpublished), FRET and protein pull-down analysis were performed.

In the FRET assay, MS1 putative interacting proteins were fused with a C-terminal RFP tag, and then transformed into the tobacco leaf epidermal cells, together with the MS1 fused with a GFP tag at the C terminus. Subcellular expression suggest that Y2H54 is not co-expressed with MS1 in the nucleus, but rather acts as a membrane protein. This indicates that the interaction between Y2H54 and MS1 that was identified by yeast 2 hybrid analysis is probably a false positive. Alternatively, the localization of Y2H54 may not equate to that of the native anther, this could potentially be due to factors that help to localise the Y2H54 protein to the nucleus being expressed specifically in the tapetal cells during a specific development stage; nevertheless this seems unlikely

The POB2 protein, which exhibited clear co-localisation with MS1 protein in the nucleus, was then used for FRET analysis. The MS1-eGFP fusion protein served as a “donor” while POB2-mRFP as the “acceptor”. If these two proteins physically interact with each other, the “acceptor” RFP would be excited by the emission of the “donor” GFP under the GFP excitation spectra, leading to a increase in the RFP intensity. Nevertheless, no significant energy transfer from the GFP chromophores to the RFP has been

detected, suggesting that the MS1 and POB2 proteins in living cells do not come into sufficient spatial proximity.

In addition to verification by FRET analysis, pull-downs were conducted to confirm potential interactions for MS1 and POB2. The MS1 and POB2 proteins were respectively tagged with 6xHis and GST at the N terminus for the pull-down analysis. A clear band of a size similar to the MS1 protein was detected in the elution using glutathione buffers specific to GST tagged POB2. This result is consistent with that identified by yeast 2 hybrid analysis, which provide an evidence for the existence of the interaction between MS1 and POB2 proteins *in vitro*.

The discrepancy observed between the *in vivo* and the *in vitro* experiments may be caused by multiple reasons. Technically, the FRET sensitized emission method requires equimolar concentrations of the fluororohores, which is difficult to guarantee in different cell samples (Sekar and Periasamy, 2003; Bhat et al., 2006). The concentrations of the fluororophores rely on the expression levels of the proteins in the cell, which are diverse and variable in individual cells. Besides, partial of the cross-talks that the acceptor is directly excited at the donor excitation spectra, cannot be subtracted by adjusting instrumental settings. In some experiments, chlorophyll pigments of the tobacco epidermis cells may absorb part of the fluorescence (Gadella Jr et al., 1999), especially those

from the RFP emissions, which cannot be fully co-corrected by using the negative references. All these factors would have an influence on the measurement of the FRET, and may result in false FRET values. Consequently, the possibility of the existence of the interaction between MS1 and POB2 *in vivo* cannot be fully excluded based on the absence of FRET.

On the other hand, the interaction between MS1 and POB2 *in vitro* is demonstrated by protein pull-down analysis, which indicates the likelihood that POB2 contains particular structures capable of binding MS1. If so, these domains might have homologies in other MS1 interacting proteins. The POB2 protein is encoded by *LRB2* which belongs to the Bric-a-Brac/Tramtrack/Broad Complex (BTB) family. Members of this family direct the selective ubiquitination of proteins following their assembly into Cullin3-based ubiquitin ligases (Grefen et al.). Considering MS1 is tightly regulated in the tapetum during a transient period, from callose breakdown until free microspore stage (Vizcay-Barrena and Wilson, 2006; Ito et al., 2007; Yang et al., 2007), it is presumed that the degeneration of the MS1 protein may involve ubiquitylation and targeted proteolysis of the protein. However, previous characterisation of *LRB2* indicated that LRB2 acts redundantly with LRB1 as nuclear-localised BTB proteins, whose double mutants are relatively normal phenotypically (Christians et al., 2012). The mutants are compromised in multiple photomorphogenic processes, including seed germination, cotyledon

opening and expansion, chlorophyll accumulation, shade avoidance, and flowering time, but no investigation specific to anther development was performed, which may suggest that fertility and pollen development occur normally. Preliminary analysis of newly developed POB2-GUS lines suggest that POB2 expression may be delayed compared with MS1, however further analysis of these lines is required (Simpson and Wilson, unpublished). This suggests that there may be a discrepancy in the temporal expression pattern of MS1 and POB2, raising further questions as to whether POB2 interacts with MS1 *in Arabidopsis thaliana*. This still requires further experiments to validate.

The alternative interacting protein Y2H54 belongs to a novel actin-binding family member providing specialized sites for actin-membrane association (Deeks et al., 2012). Stable *Arabidopsis thaliana* transformants of the full length NET2A-GFP fusion under the control of endogenous NET2A promoter show that it is specifically expressed in the pollen grains and at the plasma membrane of the shank zone in the growing pollen tubes (Deeks et al., 2012). Y2H54 exhibits expression changes throughout pollen grain germination and pollen tube growth as indicated by transcriptome analysis of *Arabidopsis thaliana* pollen grains (Wang et al., 2008). Nevertheless although Y2H54 failed to prove itself as a nuclear-localised protein when co-expressed with MS1 in the tobacco epidermal cells, making it less prospective to act as a MS1 direct interacting protein (Figure 3.7), the

association of Y2H54 with pollen development pathways is still noticeable. This may suggest a role for Y2H54 in regulating plant fertility. Thus, Y2H54 was still retained for subsequent characterising to have deeper insights in its function in pollen development regulation.

CHAPTER 4 CHARACTERISATION OF MS1 PUTATIVE INTERACTING PROTEIN Y2H54

4.1 INTRODUCTION

MS1 putative interacting protein Y2H54 (At1g58210), originally identified as EMBRYO DEFECTIVE 1674 (EMB1674), is reported as required for normal embryo development ending in seed dormancy (Tzafrir et al., 2004). Nevertheless, recent research has revealed its role as a novel actin-binding family member providing specialized sites for actin-membrane association (Deeks et al., 2012). The founder member of this family, the NET1A protein，initially discovered from an *Arabidopsis thaliana* cDNA-GFP fusion expression library，exhibits co-localization with microfilaments when transiently transformed into the leaves of *Nicotiana benthamiana*. Residues 1-94 of NET1A are identified as the minimal region responsible for actin binding, termed as the NET actin-binding (Saiga et al.) domain. Further research then discloses a total of 13 homologous sequences containing the NAB domain within the *Arabidopsis thaliana* proteome, composing four sub-families of the NET superfamily (Figure 4.1) (Deeks et al., 2012).

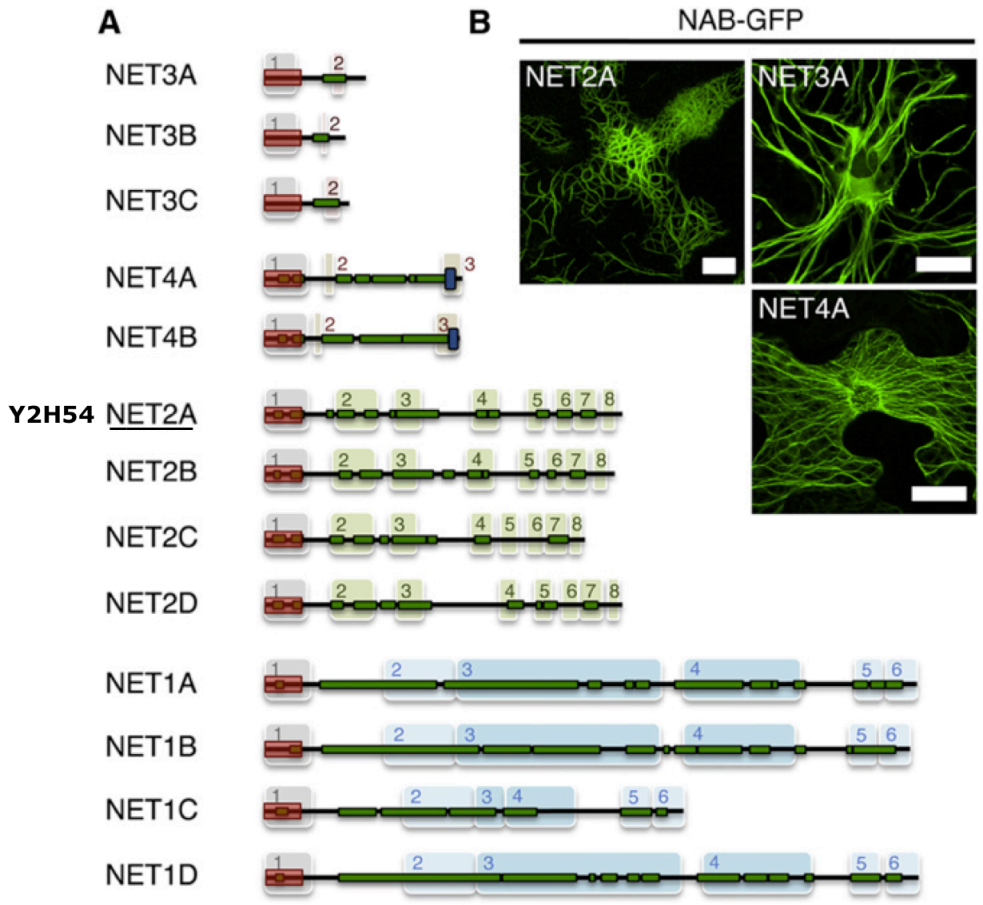


Figure 4.1 The NAB domain defines a NET superfamily (Deeks et al., 2012). (A) The 13 members of the *Arabidopsis thaliana* NET family, showing relative positions of the NAB domain and predicted coiled-coil regions (dark green). Zones of homology shared exclusively between members of each subfamily are denoted by numbered and colored tiles. (B) Representative images showing actin association of NAB domain GFP fusions derived from the NET2, NET3, and NET4 subfamilies expressed in *N. benthamiana* pavement epidermal cells. Scale bars represent 20 μm.

The MS1 putative interacting protein Y2H54 belongs to the NET2 subclade and shows preferential expression in male gametophyte. Correspondingly, stable *Arabidopsis thaliana* transformants of the full length Y2H54-GFP fusion under the control of endogenous Y2H54 promoter showed the GFP expression specifically in the pollen, which subsequently appeared at the plasma membrane of the shank zone in the growing pollen tubes (Figure 4.2) (Deeks et al., 2012).

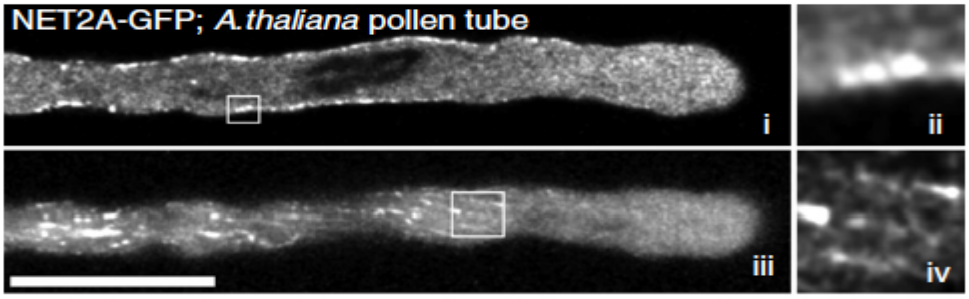


Figure 4.2 Y2H54 panels show plasma membrane-associated using laser scanning confocal microscopy (Deeks et al., 2012). Y2H54-GFP expressed using the endogenous Y2H54 promoter in *A. thaliana* pollen tubes isolated from stable transformants. White boxes indicate areas of higher magnification in neighboring panels that show membrane and filament-associated NET. Scale bars represent 10 μm.

Actin arrays are believed to be associated with plant cell growth, and the activity of actin-binding proteins is proving to be essential for proper cell morphogenesis (Hussey et al., 2006). Particularly, KIP1, the petunia NET2 homologous protein, directly interacts with a predominantly pollen-expressed receptor-like kinase (RLK), that is required for microspores to proceed to the bicellular stage from the unicellular stage (Tzafrir et al., 2004). Correspondingly, Y2H54/ NET2A as an actin-binding protein exhibits expression changes through the pollen grain germination and pollen tube growth progress in the transcriptome analysis of *Arabidopsis thaliana* pollen grains (Wang et al., 2008). In summary, though it seems unlikely that there is an *in planta* interaction between Y2H54 and MS1 (Chapter 3), the association of Y2H54 with pollen development pathways is still noticeable.

Over recent decades, manipulating gene expression in living plants has become a common way to determine the function of particular genes of

interest. There are ways to achieve mutagenesis of plant genome gene expression, and the availability of transgenic plants generated by *Agrobacterium tumefaciens*-mediated technique allows us to directly monitor the resulting change of phenotypes (Klee et al., 1987; Valvekens et al., 1988). Expression under the control of a strong promoter such as cauliflower mosaic virus (CaMV) 35S promoter has been widely used for increasing ectopic gene transcription to cause the over-expression in the infected plants (Odell et al., 1985; Fang et al., 1989). Insertional mutagenesis using foreign T-DNA that is inserted into the gene of interest allows mutated plants to be easily identified based on the mark of the T-DNA insertion (Krysan et al., 1999).

For almost a decade, RNA interference (RNAi) technology has been used to discover and validate gene function (Waterhouse et al., 2001). RNAi is a post-transcriptional gene-silencing phenomenon triggered by double-stranded RNA (dsRNA) produced by experimental induction or endogenous RNAs transcripts from transgene. The molecular mechanism of RNAi is initiated when the dsRNA recognized by member of the RNAase III family, the Dicer, is converted into the 21-25nt small interfering RNA (siRNA) duplexes. These siRNA containing two-nucleotide 3' overhangs and 5'-phosphate termini is then unwound into two strands, one of which preferentially incorporated into the multicomponent nuclease, the RNA-induced silencing complex (RISC). The RISC accordingly uses the siRNA as a guide to select its homologous mRNA sequence, and cleaves the siRNA-mRNA hybrids near the centre of siRNA, causing the degradation of

specific mRNA. RNAi as an efficient knockdown technology has been used as a powerful tool to analyze gene function in various organisms. In plants, this is usually achieved by enforcing the expression of the silencing trigger, usually as a hairpin RNA (hpRNA) derived from inverted repeat sequences spaced with an unrelated intron sequence, and driven by a strong promoter such as CaMV35S promoter (Smith et al., 2000; Waterhouse and Helliwell, 2003). Until recently, RNAi has been developed as an efficient method to achieve the down-regulation of particular gene in many examples, especially in the model plant *Arabidopsis thaliana* (Chuang and Meyerowitz, 2000; Stoutjesdijk et al., 2002).

4.2 MATERIALS AND METHODS

4.2.1 Plant Materials

Seeds of the *Arabidopsis thaliana* homozygous lines, stably transformed with Y2H54-GFP under its endogenous promoter, were kindly provided by Prof. Patrick J. Hussey (University of Durham) (Deeks et al., 2012).

Seeds of *Arabidopsis thaliana* T-DNA insertional mutagenesis SALK 020898 (C*ol*-0 background, NASC ID N565759) and SALK 065759 (C*ol*-0 background, NASC ID N20898) were obtained from Nottingham *Arabidopsis* Stock Centre.

Overexpression lines (L*er* background) of *Y2H54* under the control of CaMV35S promoter using Gateway destination vector pGWB14 (Nakagawa et al., 2007) (Appendix VII) were generated and maintained in our lab by Dr. Caiyun-Yang (University of Nottingham). The T1 generation was previously crossed with the hybrids of the L*er* wild type and the *ms1ttg1* mutant, giving rise to the progenies that overexpress Y2H54 yet in the *ms1ttg1* mutant background.

Seeds of the transgenic *Arabidopsis thaliana* were sterilized (Section 2.2) and screened on half MS medium (Section 2.10.2) with appropriate antibiotics added (Table 2.8), Hygromycin B for Y2H54-GFP lines and Kanamcyin for the other plasmids. Plates were placed in a cold room for 3 days at 4°C to synchronize germination, and then cultured under full light

(140μmol·m-^{2}·sec^{-1}) at 22-24°C. After 10 days, seedlings that did not turn white were transformed into soil and grown in a glasshouse (Section 2.1).

4.2.2 Identification of T-DNA Insertional Mutants

Indentification of T-DNA insertional mutants was conducted according to the instructions from the SIGnAL T-DNA-Express facility (Salk Institute Genomic Analysis Laboratory http://signal.salk.edu/cgi-bin/tdnaexpress) (Figure 4.3). By using the three primers (LBb1.3+LP+RP) for SALK lines, users for WT (Wild Type - no insertion) should get a product of about 900-1100 bps (from LP to RP), for HM (Homozygous lines - insertions in both chromosomes) will get a band of 410+N bps (from RP to insertion site 300+N bases, plus 110 bases from LBb1.3 to the left border of the vector), and for HZ (Heterozygous lines - one of the pair chromosomes with insertion) will get both bands.

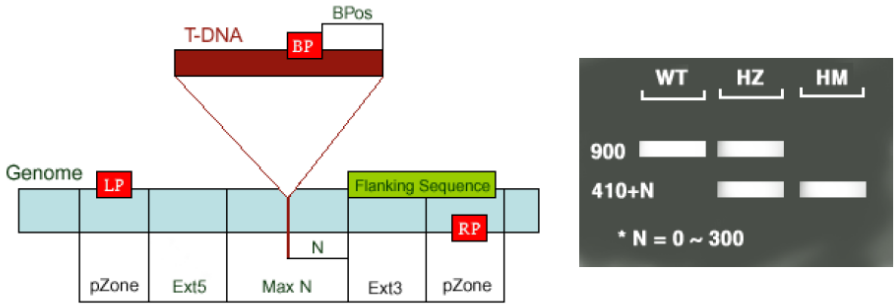


Figure 4.3 Principle of SALK T-DNA insertion identification (http://signal.salk.edu/tdnaprimers.2.html). N, Difference of the actual insertion site and the flanking sequence position, usually 0 - 300 bases; MaxN, Maximum difference of the actual insertion site and the sequence, default 300 bps;pZone, Regions used to pick up primers, default 100 bps;Ext5, Ext3, Regions between the MaxN to pZone, reserved not for picking up primers; LP, RP - Left, Right genomic primer; BP - T-DNA border primer LB - the left T-DNA border primer; BPos - The distance from BP to the insertion site.

Arabidopsis thaliana T-DNA insertional Mutagenesis of Y2H54, SALK_020898 and SALK_065759, were identified by genotypic analysis as previously described (Section 2.3.3), using T-DNA primer Lbb 1.3R plus gene specific primer SALK020898F and SALK065759R (Table 4.1).

Table 4.1 Primers for Identification of T-DNA Insertional Mutants.

Primer	Sequence 5'-3'
Lbb 1.3R	ATTTTGCCGATTTCGGAAC
SALK_020898R	GCAAAAGTGTCTCCATCTTCATC
SALK_065759F	GATTTTGGACGATGTTGTTGG

4.2.3 Expression Analysis

To determine the expression pattern of genes of interest in wild type, expression analysis was conducted using the various tissues, namely the old bud, the young bud, the immature bud, the open flower, the leaf, the stem, the root and the young silique, from the L*er* wild type plants. Buds were numbered from the outmost unopened bud, petals of which were unseen, and then grouped based on the bud morphology (Figure 4.4). Buds numbered 1 to 4 were grouped as old buds (pollen mitosis II through to dehiscence), 5 to 8 as young buds (formation of the sporogeneous tissues to pollen mitosis I), and the rest as immature buds (prior to sporogeneous tissue formation). The siliques tested were picked from the same position on each branch of each plant. Materials were harvested from over 10 plants of 5 weeks old after sowing. The expression analysis was performed by the Real-Time quantitative PCR (Section 2.6.3) using gene specific primers

(Table 4.2), unless stated otherwise.

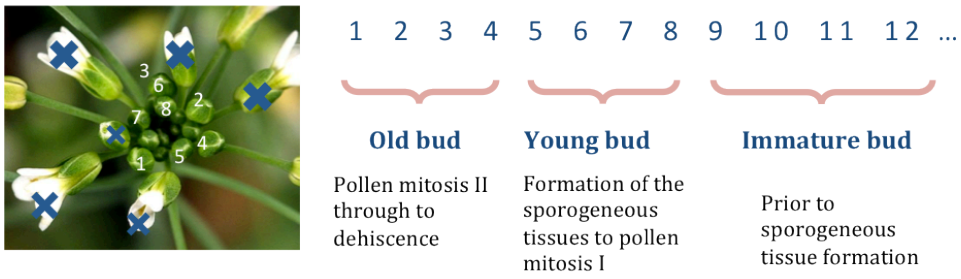


Figure 4.4 *Arabidopsis thaliana* buds staged based on bud morphology (Vizcay-Barrena, 2005). Buds were numbered from the outmost unopened bud without visible petals. Buds numbered 1 to 4 were grouped as old buds (pollen mitosis II through to dehiscence), 5 to 8 as young buds (formation of the sporogeneous tissues to pollen mitosis I), and the rest as immature buds (prior to sporogeneous tissue formation). Buds 7 onwards are not supposed to be seen on the graph.

To investigate the changes in expression levels of genes due to transformation, individual whole bud of the first inflorescence from each plant of 5 weeks old after sowing was collected. The expression analysis were performed using semi-quantitative PCR with Taq® Ready Mix® PCR Reaction Mix (Sigma-Aldrich) (2.6.1) or the real-time quantitative PCR (Section 2.6.3), unless stated otherwise. The semi-quantitative PCR was following typical conditions: 94°C for 3mins; 28 cycles of 94°C for 30secs, 55°C for 30secs, 72°C for 40secs); then 72°C for 5mins. cDNAs from equal amount of the RNA samples were used as templates, and gene specific primers designed accordingly (Table 4.2).

Table 4.2 Primers for Expression Analysis.

Primer	Gene	Sequence 5'-3'
54_1607F	*Y2H54*	CGGAAGACGAGGATGAAGAG
54_1918R	(At1g58210)	CTTGTTCATGTTCCAACTGGTT
54L_1847F	*Y2H54_Like*	AAAGAAACCGGGCTGAAAGT
54L_2187R	(At1g09720)	CCCAGCATTACCTTGTTGCT
Actin 7_512F	*Actin*	GCCATTCAGGCCGTTCTTTCT
Actin 7_876R		CGGAATCTCT CAGCTCCGATG

4.2.4 RNA Interference

To generate the RNA interference lines of Y2H54, 380bp of DNA fragments was amplified from Y2H54 cDNA extracted from *Arabidopsis thaliana* buds (Section 2.6.2). Primers 1600_RNAiF and 1980_RNAiR were added with *att* adapters (Table 4.3) to be cloned into Gateway compatible vector pK7GWIWG2 (Karimi et al., 2002), as previously described (section 2.9). The recombinant construct was screened by colony PCR using gene specific primer 1980R, respectively working with the primer 35S_For derived from the CaMV35S promoter on the vector and ntpII_new_R primer from the NTPII gene encoding for kanamycin resistance.

Table 4.3 Primers Used for RNA Interference.

Primers	Sequence 5'-3'
YH54_1600_RNAiF	GGGGACAAGTTTGTACAAAAAAGCAGGCT CAGCTGAAGATCTGGTGACGGAAGA
YH54_1980_RNAiR	GGGGACCACTTTGTACAAGAAAGCTGGGT CGGTGGCAACCGAAAAGTTAGAAG
1980R	GGTGGCAACCGAAAAGTTAGAAG
35S_For	CACAATCCCACTATCCTTCGCAAGAC
ntpII_new_R	ATACTTTCTCGGCAGGAGCA

The recombinant construct transformed into *Agrobacterium* competent cells C58 (Section 2.14) was subsequently used for floral dipping transformation into the *Arabidopsis thaliana* L*er* wild type (Section 2.15)**.** Seeds of the transformed plants were harvested when entirely dried out (Section 2.1).

Seeds of the transformed *Arabidopsis thaliana* were sterilized (Section 2.2) and screened on half MS medium with Kanamycin (Section 2.10.2). Plates were placed in a cold room for 3 days at 4°C to synchronize germination, and then cultured under full light (140μmol·m-2·sec-1) at 22-24°C. After 10 days, seedlings that did not turn white were transformed into soil (Section 2.1). Expression analysis of the RNAi lines were conducted using three different Y2H54 primer pairs designed at different positions on the cDNA of Y2H54 (Table 4.4).

Table 4.4 Primers Used for Expression Analysis of the Y2H54 RNAi lines.

Primer	Sequence
P1_1502F	GCGATGCTAGGACGTTGATGG
P1_1065R	CTTCTCCGCAAGCTTCACAACA
P2_1607F	CGGAAGACGAGGATGAAGAG
P2_1918R	CTTGTTCATGTTCCAACTGGTT
P3_2443F	GCTCAATCAGGTGGCAATAA
P3_2773R	GGGATGATGCTGATTGCTTT

4.2.5 Bioinformatic Analysis

The *Arabidopsis* DNA and protein sequences were obtained from database TAIR (www.Arabidopsis.org). NCBI (National Center for Biotechnology Information; www.ncbi.nlm.nih.gov) BLAST (Basic Local Alignment Search Tool ; Altschul et al., 1990) analysis was carried out for identifying homologous genes to Y2H54. MacVector (MacVector, Inc, PMB 150; USA) was used for sequence editing, primer designing, PCR products prediction, and multiple sequence alignment. Primers were designed using online web tool Primer3 (http://104rabi.wi.mit.edu/primer3/). The expression pattern of the genes were predicted using AtGenExpress Visualization Tool (https://www.genevestigator.com/gv/) and *Arabidopsis* e-FP Browser http://bar.utoronto.ca/efp_arabidopsis/cgi-bin/efpWeb.cgi).

4.2.6 RACE PCR

5' RACE-PCR was performed to obtain the 5' end sequence of the *Y2H54* gene (Figure 4.5), using the GeneRacer^{TM} kit (Invitrogen) based on the

manufacture's instruction. Total RNA extracted from buds of the L*er* wide type (Section 2.4) was first treated with calf intestinal phosphatase (CIP, 10U/µl) at 50°C for 1 hour to remove the 5' phosphates and eliminate truncated mRNA and non-mRNA for the subsequent ligation with the GeneRacer RNA oligo. This dephosphorylated RNA was then incubated with tobacco acid pyrophosphatase (TAP, 0.5U/µl) at 37°C for 1 hour to remove the 5' cap structure from intact, the full length mRNA. The GeneRacer RNA Oligo within the kit used to provide a known priming site for GeneRacer amplification was ligated to the 5' end of the treated mRNA using T4 RNA ligase (Section 2.9.3). The ligation was subsequently reverse-transcribed using the GeneRacer oligo dT primer (Invitrogen) (Section 2.5) to prepare the first-strand cDNA for RACE.

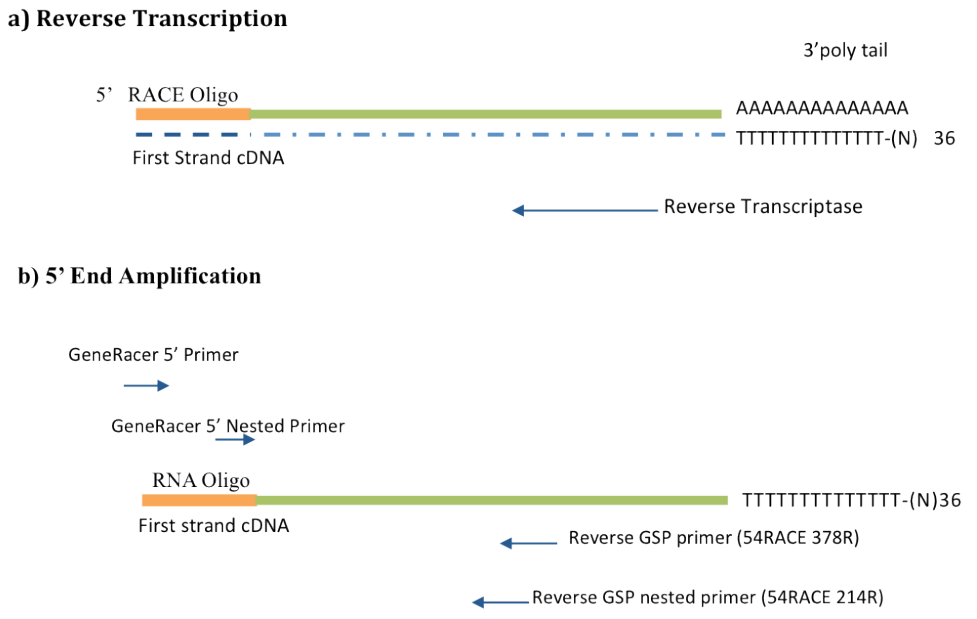


Figure 4.5 5' RACE-PCR reaction principle. (a) Reverse transcription using an Oligo dT primer which included a PCR primer site. (b) 5'end amplification. GeneRacer 5' primer and Reverse Gene Specific primers were used for 5' amplification.

Gene specific primers specific to the GeneRacer RNA oligo and the *Y2H54* predicted cDNA sequence of L*er* wide type were designed (Figure 4.6). Amplification was initially conducted by 'Touchdown PCR' with designed primers (Table 4.4) using Phusion High-Fidelity DNA Polymerase (Thermo) (Section 2.6.2), following the conditions: 98°C for 30secs; 5 cycles of 98°C for 20secs, 72°C for 2mins; 5 cycles of 98°C for 15secs, 72°C for 2mins; 30 cycles of 98°C for 15secs, 60-65°C for 25secs, 72°C for 2mins; then 72°C for 10mins. Gradient temperatures of 60.3°C, 62°C, 63.3°C, and 65°C were set up to maximize specific amplification products accumulation.

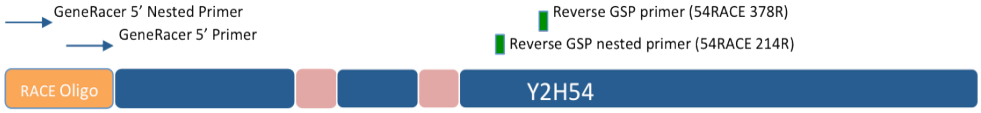


Figure 4.6 Structure of the *Y2H54* gene for 5'RACE PCR reaction. Position of the RNA oligo (orange box) and primers used for the amplification of the 5 ends are shown (Table 4.4; Figure 4.6). GSP primers designed based on previously published sequence on the *Arabidopsis* Information Resource (TAIR) (www.arabidopsis.org), to be used with the 5' GeneRacer primers provided in the kit. Blue bars indicate exons and the pink represent the introns.

Table 4.4 Primer Sequences Used for 5' RACE PCR.

Primer	Sequence 5'-3'
54RACE_214R	CAGCTAAAGCGCGGTATGAACGAAAAGC
54RACE_378R	CGGTTTTCGTGGCCTTCCGTCGT
GeneRacer 5' Primer	CGACTGGAGCAAGAGGACACTGA
GeneRacer 5' Nested Primer	GGACACTGACATGGACTGAAGGAGTA

In order to increase the specificity and sensitivity of 5' RACE-PCR products, 1µl of the "Touch-down" PCR product was used as the template to perform a further nested PCR using the gene specific primers 54RACE-5'-214F and Gene Racer Nested 5' Primer provided in the kit (Table 4.4, Figure 4.6). The PCR was conducted using Phusion High-Fidelity DNA polymerase (Thermo) (Section 2.6.2), following the conditions: 98°C for 2mins; 30 cycles of 98°C for 30secs, 65°C for 30secs, 72°C for 2mins; and then 72°C 10mins. The reactions following the same conditions were performed within four separate tubes in parallel to each other. The PCR products from two of the tubes were prepared for sequencing using the Gene Racer Nested 5' Primer (Section 2.13). The rest was then cloned into the pCR-BluntII-TOPO vector (Section 2.9), and plasmids isolated from selected colonies for sequencing (Section 2.12.13).

4.3 RESULTS

4.3.1 Determination of the Transcript Sequences of the *Y2H54* Gene

Previously, the publically available full-length sequence of the *Y2H54* (At1g58210) gene contains three exons separated by two introns (Figure 4.7). However, it is suggested by our lab that the transcription of *Y2H54* initiates at a different site, forming two exons separated by one intron (Figure 4.7) (Yang and Wilson, unpublished). To determine the transcription initiation site of the *Y2H54* gene, 5' end RACE PCR as well as RT-PCR using selected primers around the site was sequentially performed.

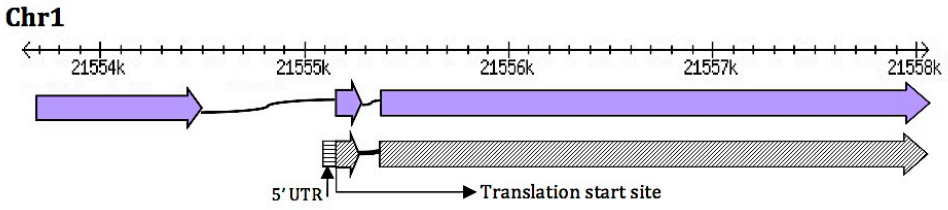


Figure 4.7 Putative genomic structure of *Y2H54*. Structures of the publically available sequence and the one predicted by our lab were shown on the *Arabidopsis thaliana* Chromosome 1. The purple bars indicate the structure of *Y2H54* based on publically available information, and the grey one represents the result generated in our lab (Dr Cai-yun Yang and Zoe Wilson). Arrow bars indicate exons.

The full-length 5' end sequence of Y2H54 was identified by RACE-PCR, using gene specific primers designed as previously described (Figure 4.8). The first-step "Touchdown" PCR to maximize specific amplification products accumulation was performed using gradient annealing temperatures between 60-65°C (Section 4.2.6), in which products of

approximately 500bp were generated (Figure 4.8a). The PCR products with the highest enrichment, annealing at 62°C, was used as the template of the subsequent nested PCR.

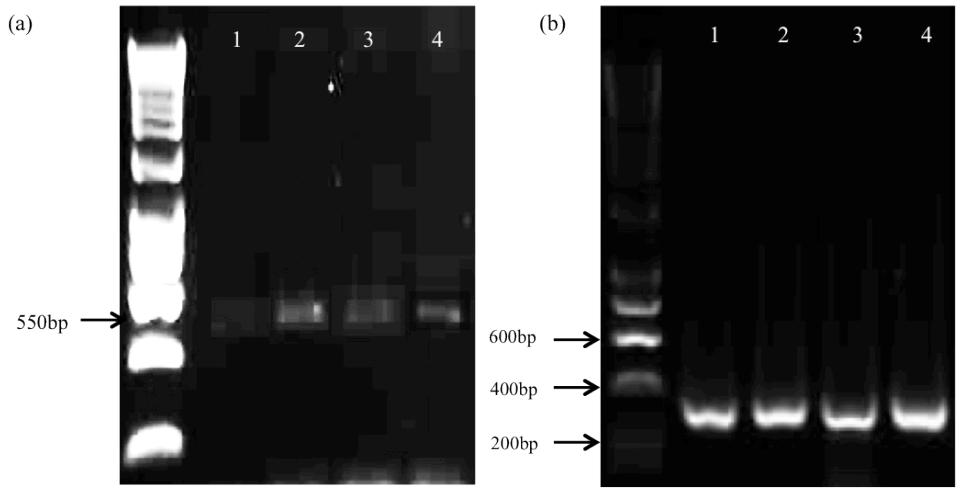


Figure 4.8 RACE-PCR reaction. (a) Touchdown PCR reaction performed using 54RACE-5'-378R and GeneRacer 5' primers (Figure 3.5). Lane 1, 2, 3 and 4 respectively shows the amplification under the gradient temperature of 60.3°C, 62°C, 63.3°C, and 65°C. (b) 5' RACE nested PCR was conducted using 54RACE-5'-214R and GeneRacer 5' primers (Figure 3.5). Each lane shows the amplification of individual samples as replicates at a fixed temperature of 65°C.

Specific bands of about 320bp were amplified in the nested PCR (Figure 4.8b), which were then cloned into pENTR/D-TOPO vectors for sequencing. 17 samples derived from individual colonies were sequenced. Various transcription initiations of the *Y2H54* gene were found by sequence alignment (Figure 4.9), only two exons with one intron were identified, corresponding to previous prediction by our lab. Despite the slight variation in transcript size, the translation start codon was proposed to be the same to

generate a functional amino-acid sequence, which is 1465bp upstream to the one of the publically available sequence on the chromosome.

Chromosome Regions and Sequences of the Transcription Starts	Frequency
21,553,994 |GTCTGATCTTGAAACTCTCAAGTTCTTACGTTGATGACTTGGCTATGTTTGAACCTTCCTCTGCT...AAAGATGTT...	1
21,555,001 |CTTGAAACTCTCAAGTTCTTACGTTGATGACTTGGCTATGTTTGAACCTTCCTCTGCT...AAAGATGTT...	1
21,555,002 |TTGAAACTCTCAAGTTCTTACGTTGATGACTTGGCTATGTTTGAACCTTCCTCTGCT...AAAGATGTT...	2
21,555,004 |GAAACTCTCAAGTTCTTACGTTGATGACTTGGCTATGTTTGAACCTTCCTCTGCT...AAAGATGTT...	2
21,555,013 |AAGTTCTTACGTTGATGACTTGGCTATGTTTGAACCTTCCTCTGCT...AAAGATGTT...	2
21,555,021 |ACGTTGATGACTTGGCTATGTTTGAACCTTCCTCTGCT...AAAGATGTT...	3
21,555,023 |GTTGATGACTTGGCTATGTTTGAACCTTCCTCTGCT...AAAGATGTT...	2
21,555,026 |GATGACTTGGCTATGTTTGAACCTTCCTCTGCT...AAAGATGTT...	3
21,555,054 |CTGCT...AAAGATGTT...	1
Total	17

Figure 4.9 Sequence alignment of the 5' end of *Y2H54*. Transcription starts at various locations in the total 17 samples. The translational start codon keeps the same to generate a functional amino-acid sequence, which is the "ATG" in orange. Numbers at the top indicate the sequence location on the chromosome.

To confirm this, a series of RT-PCR experiments were performed using the reverse primer 54_R at the end of the transcripts, in combination with a set of forward primers (1051F, 1212F, 1407F, 1418F, and 1465F) (Table 4.5), which all commenced at the "ATG" prior to or at the top of the confirmed transcripts (Figure 4.10).

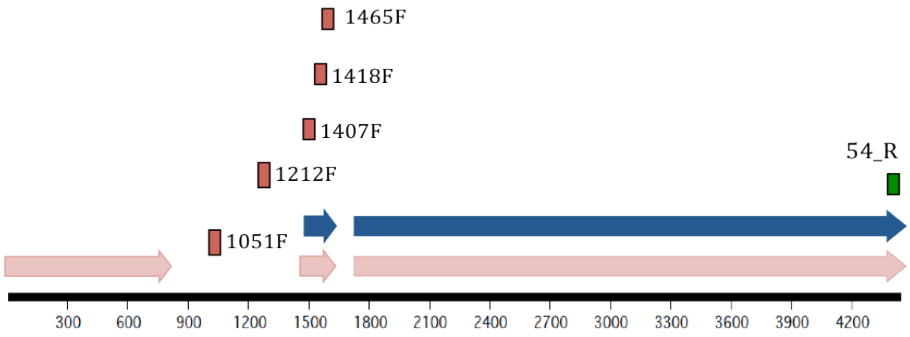


Figure 4.10 Locations of the selected RT-PCR primers on the genome. The locations of the primers were shown. The pink arrow bars indicate exons of the publically available sequence, and the blue ones presents the exons of the sequence confirmed by our lab. Forward "ATG" primers (1051F, 1212F, 1407F, 1418F, 1465F) were shown as red bars and reverse primer exhibited as the green one. Scale bar:bp.

Table 4.5 Primers for Confirmation of The Translational Start Codon.

Primer	Sequence (5'-3')
1051F	ATGTTCTTCATTGACTGCATTCTTCTG
1212F	ATGCGGATAAGAAATAAATGTCATATATT
1407F	ATGACTTGGCTATGTTTGAACCTTCC
1418F	ATGTTTGAACCTTCCTCTGCTTCTTC
1465F	ATGTTGCAGAGAGCAGCGAGCAA
54_R	TTATTCAGGGAGCTTCCCAGGTGGCCT

In the RT-PCR analysis, specific bands of approximately 3kb were generated in three out of the five reactions, when using forward primers 1407F, 1418F and 1465F (Figure 4.11), consistent with the RACE sequencing results (Figure 4.9). However, only the translation starting at the ATG of primer 1465F is capable of generating a predicted functional polypeptide of 951aa, separated by a single intron of 127bp (Appendix VIII). BLAST analysis identified a gene highly homologous to *Y2H54,* termed as

Y2H54_Like (At1g09720), which showed 73% similarity at the cDNA level and 77% at the amino-acid sequence (Appendix VIII). The homology was previously identified as a NET2B, also a member of the NET2 subclade of the NET superfamily, in which Y2H54 is refered to as NET2A (Deeks et al., 2012).

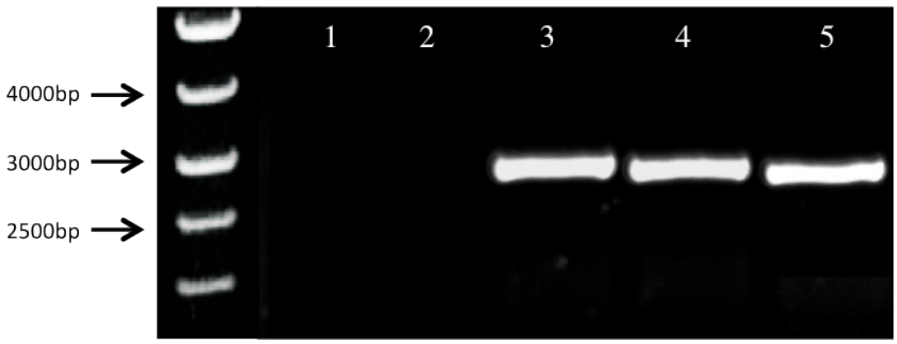


Figure 4.11 RT-PCR of the *Y2H54* gene using ATG primers (Table 3.4). Lane 1-5 indicate the PCR products using the forward primers 1051F, 1212F, 1407F, 1418F and 1465F, respectively. Reaction using forward primers 1407F, 1418F, 1465F generated specific bands of 3kb, whereas the rest did not.

4.3.2 Expression Pattern Analysis

4.3.2.1 Bioinformatic analysis

The expression patterns of the *Y2H54* gene and its homologue, the *Y2H54_Like* gene, were predicted by the AtGenExpress and e-FP Browser (Figure 4.12; Figure 4.13). According to these databases, the *Y2H54* is specifically regulated during flower development in *Arabidopsis thaliana*,

showing significant enrichment in the buds, flowers, siliques and the seeds (Figure 4.12); whereas *Y2H54_Like* exhibited almost even expression across the *Arabidopsis thaliana* life cycle (Figure 4.13). Noticeably, both of the two genes were up-regulated in the stamen and pollen tissues, suggesting putative roles of the genes in microsporogenesis.

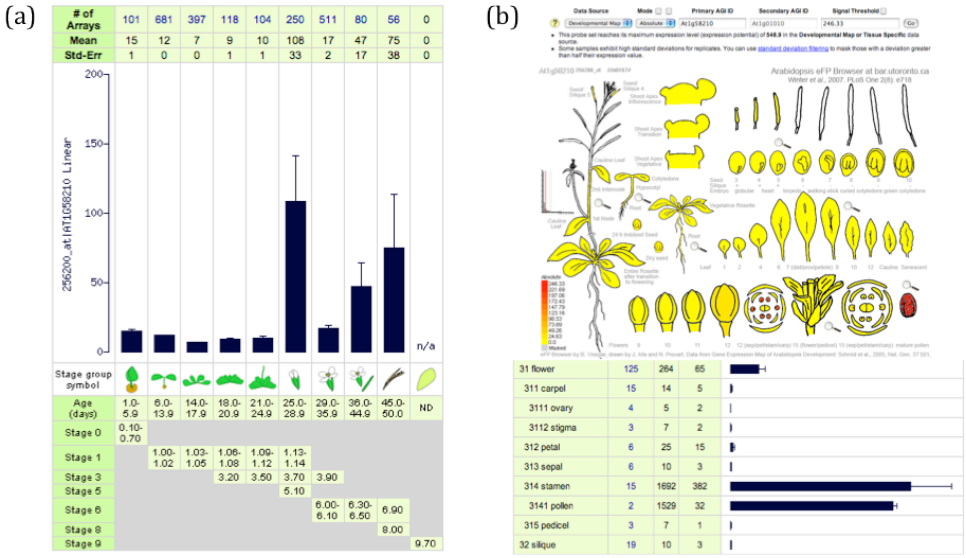


Figure 4.12 Bioinformatic analysis of the Y2H54 expression pattern. (a) Expression pattern of Y2H54 predicted by AtGenExpress Visualization Tool (https://www.genevestigator.com/gv/). (b) Expression pattern of Y2H54-At1g58210 predicted by *Arabidopsis* e-FP Browser (http://bar.utoronto.ca/efp_arabidopsis/cgi-bin /efpWeb.cgi).

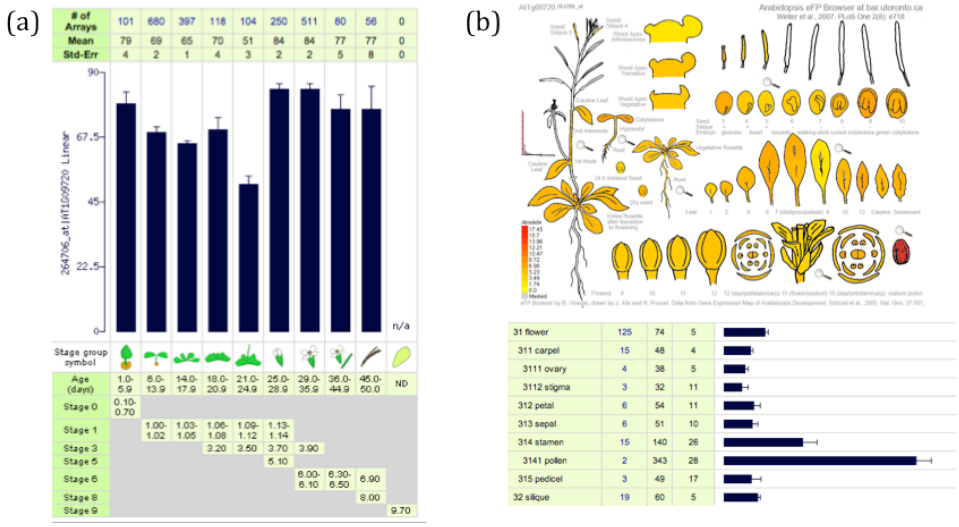


Figure 4.13 Bioinformatic analysis of the Y2H54_LIKE expression pattern. (a) Expression pattern of Y2H54_LIKE predicted by AtGenExpress Visualization Tool (https://www.genevestigator.com/gv/). (b) Expression pattern of Y2H54_LIKE predicted by *Arabidopsis* e-FP Browser (http://bar.utoronto.ca/efp_arabidopsis/cgi-bin /efpWeb.cgi).

4.3.2.2 Quantitative PCR analysis

The expression patterns of the *Y2H54* and *Y2H54_Like* genes were further tested by qRT-PCR, using tissues collected from L*er* wild type as previously described (Section 4.2.3). To avoid false pairing due to the high similarity of the two sequences, primers specific to each gene were first tested in semi-quantitative RT-PCR reactions (data not shown), whose products were then sequenced in the pENTR/D-TOPO vectors.

The highest expression level of *Y2H54* was detected in the old bud group, and became lower in the young bud, the open flower and the leaf. In the immature bud, the stem, the root, and the young silique, expression of the *Y2H54* was barely detected (Figure 4.14a). The expression of the

Y2H54_Like was enriched in the bud tissues as they matured and peaked in the open flowers, which was decreased in the young silique, the leaf, the stem, and the root (Figure 4.14b). Based on the quantitative RT-PCR analysis, the *Y2H54* and *Y2H54_Like* exhibited significantly different expression patterns, despite the high similarity between their sequences. Considering the *Y2H54_Like* failed to interact with the MS1 protein in previous yeast-2-hybrid assays, it was excluded from subsequent analysis.

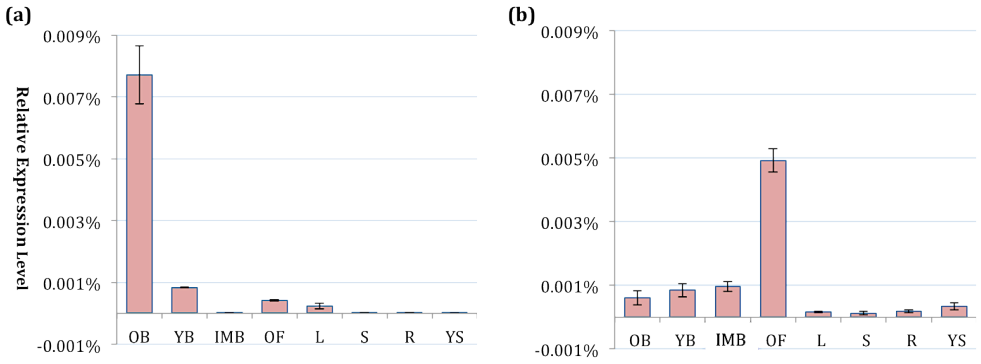


Figure 4.14 Quantitative expression analysis of Y2H54 and Y2H54_Like in *Arabidopsis thaliana* tissues. PCR reactions were carried out using *Y2H54* gene specific primers (Table 4.2). (a) Expression pattern of Y2H54 detected by quantitative expression analysis (b) Expression pattern of Y2H54_Like detected by quantitative expression analysis. The *Actin7* gene was used to normalise the expression level. OB, old bud; YB, young bud; IMB, immature bud; OF, open flower; L, leaf; S, stem; R, root; YS, young silique.

4.3.2.3 GFP fusion analysis in *Arabidopsis thaliana*

In order to understand the expression pattern of Y2H54 more precisely, *Arabidopsis thaliana* (C*ol-0* background) stably transformed with the Y2H54-GFP fusions under the endogenous promoter of *Y2H54* was analysed using confocal microscopy. The Y2H54-GFP fusion protein was detected on the tapetum membrane during anther stage 7-8, yet was absent

from the tapetum nuclei (Figure 4.15 a,b). Faint GFP signal could also be detected in the microspores at anther stage 12-13 (Figure 4.15 c,d), when the locules were about to open to release mature pollen grains. This result is consistent with the quantitative expression analysis (Section 4.2.2), in which the expression level of *Y2H54* is higher in the old buds than the young ones, if considering the existence of more microspores than the tapetum cells.

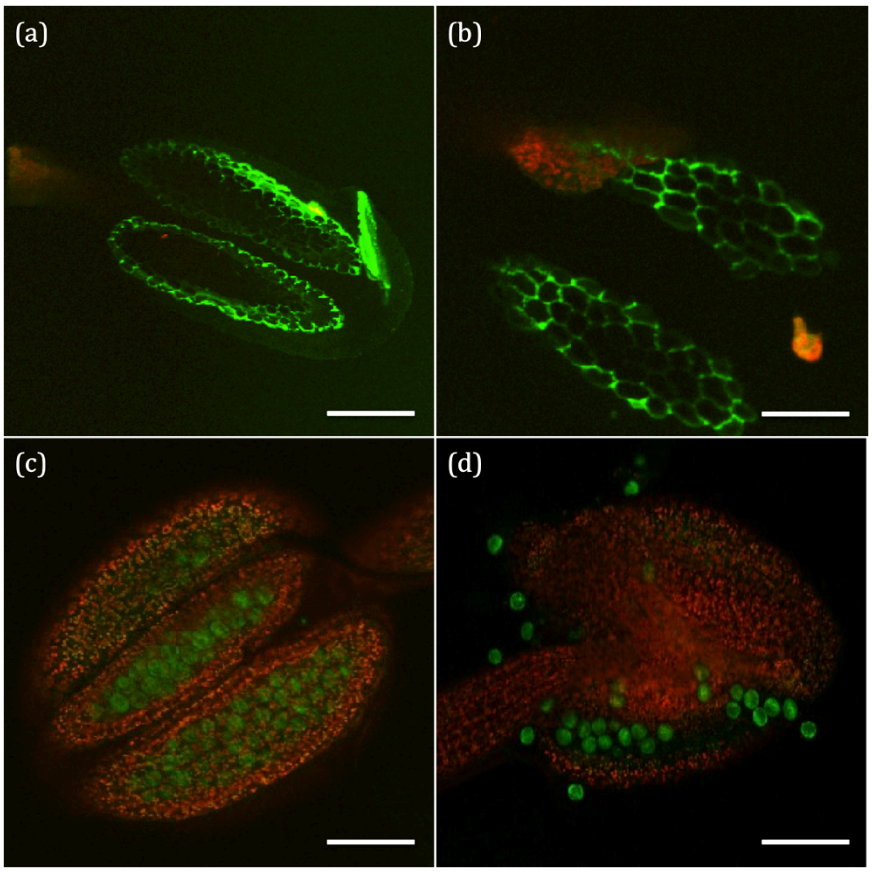


Figure 4.15. Expression of Y2H54-GFP in the anthers of *Arabidopsis thaliana*. Y2H54-GFP fusions were stably expressed in *Arabidopsis thaliana,* under the control of the endogenous promoter of *Y2H54*. Staging was based on the bud morphology (Vizcay-Barrena, 2005). (a) and (b) The tapetum membrane at anther stage 7-8; (c) Anther locules filled with microspores at anther stage 12; (d) Microspores released from the locules at anther stage 13.

4.3.3 Characterisation of the Y2H54 T-DNA Insertional Mutants

4.3.3.1 Genotypic analysis

Arabidopsis thaliana T-DNA insertional mutagenesis of Y2H54, SALK_020898 and SALK_065759, were identified by genotypic analysis as previously described (Section 4.2.2) using T-DNA primer Lbb 1.3R with gene specific primer SALK020898F and SALK065759R (Table 4.1). The T-DNA insertions of both lines were located in the promoter region, upstream of the transcription start, as confirmed by RACE-PCR, rather than the first intron as the publically available information indicated (Figure 4.16).

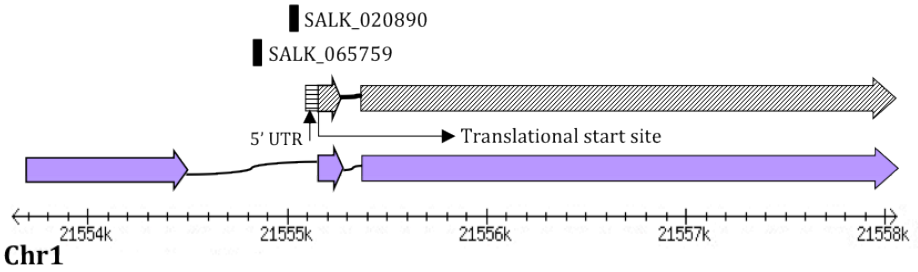


Figure 4.16 Predicted genomic structure of Y2H54 and T-DNA insertion location. The location of the T-DNA insertion and the genomic structure of the *Y2H54* gene are shown. The purple arrow bars indicate the publically available cDNA sequence, and the grey one represents the sequence confirmed by RACE-PCR. Scale bar shows the position on the *Arabidopsis thaliana* Chromosome 1.

For SALK_020898 lines, 3 of the 12 plants tested were identified as homozygous plants and another 5 as heterozygous (Figure 4.17 a), with the rest wild type. For SALK_065759 lines, a total of 12 out of 12 plants were

identified as homozygous for the insertion (Figure 4.17 b).

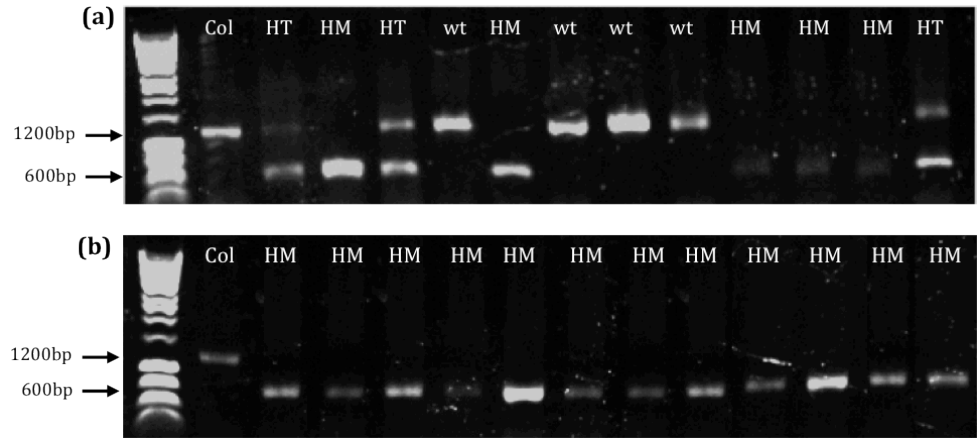


Figure 4.17 Genotyping of the SALK lines carrying the T-DNA insertions. (a) Genotyping of the SALK_020898 lines. (b) Genotyping of the SALK_065759 lines. Wild type controls (wt) show a single band of approximately 1200bp. Homozygous (HM) plants show a single band of approximately 600bp and heterozygous (HT) samples contain both bands.

4.3.3.2 Expression analysis

Homologous lines of both T-DNA insertional mutants were used for semi-quantitative expression analysis with gene specific primers. The *ACTIN 7* primers were used to check the integrity of the cDNAs, and wild type C*ol* cDNA as the positive control. Expression levels of *Y2H54* were not altered in the mutants compared with wide type (Figure 4.18), suggesting that the gene had not been silenced by the T-DNA insertions.

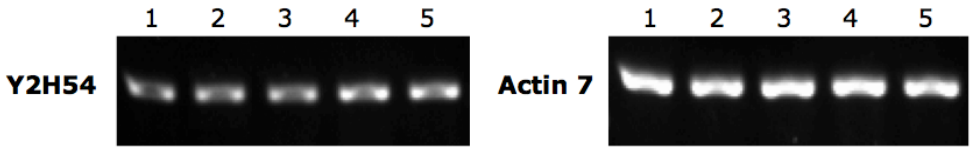


Figure 4.18 Representive semi-quantitative expression analysis of T-DNA insertional lines. *Actin7* was used as the control to check the integrity of the cDNAs. Lane 1, wt C*ol.* Lane 2 and 3, SALK_065759. Lane 4 and 5, SALK_020898.

4.3.3.3 Phenotypic analysis of T-DNA insertion line

Phenotypic analysis of T-DNA mutants was carried out by checking the anther development and pollen formation. As expected from the results of the expression analysis, no significant phenotypes were observed in the plants carrying T-DNA insertions when compared with wild type C*ol* (Figure 4.19 a-f). Mutants were fully fertile with viable pollen grains produced, as indicated by Alexander staining analysis (Figure 4.19 g,h).

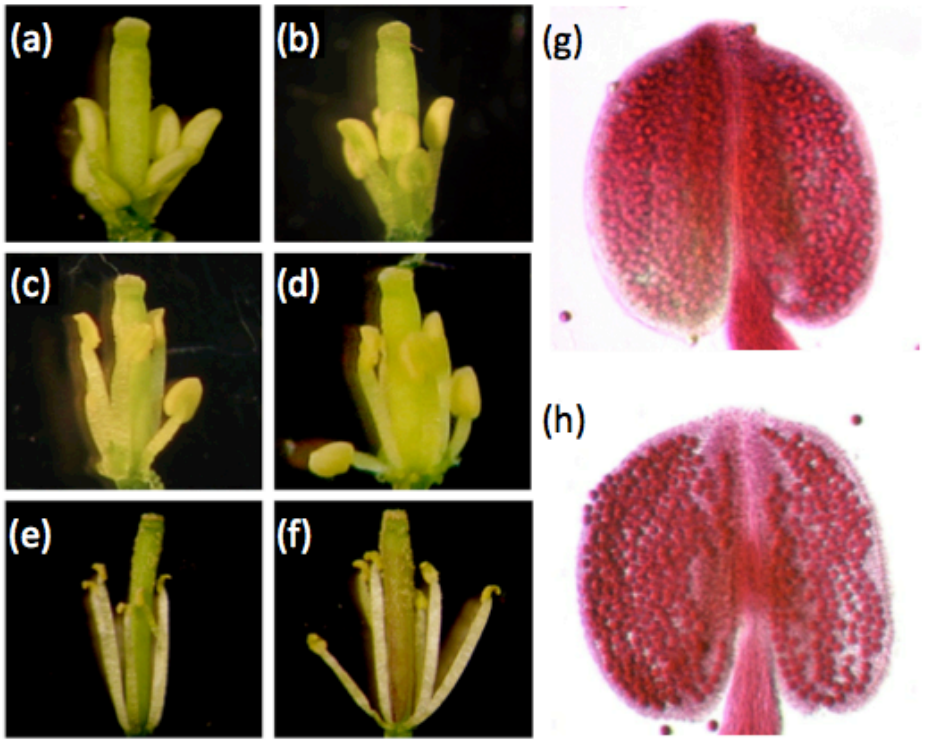


Figure 4.19 Phenotypic analysis of the T-DNA insertional lines. Anther development was checked by microscopic analysis, using young buds (a-b), old buds (c-d) and open flowers (Guan et al.). (a)(c)(e) and (g), wild type C*ol*; (b)(d)(f) and (h), T-DNA insertional mutant. No significant phenotypes were observed for the mutant when compared with wt. Pollen development was checked by Alexander staining assay (Aexander, 1969; Frankel et al., 1969), showing that mutants (h) still produced fully viable pollen grains as seen in the wt (g).

4.3.4 RNA Interference

4.3.4.1 Generation of RNAi construct

To obtain a null mutant for Y2H54, RNA interference lines were generated using a 380bp selected cDNA region that is not conserved among the NET superfamily proteins (Figure 4.20a). After amplified from the cDNA by PCR, it was inserted into the GATEWAY destination vector PK7GWIWG2

(Figure 4.20b). The recombinant construct was screened by colony PCR using gene specific primer 1980R and primer on the original vector, either 35S_For from CaMV35S promoter or ntpII_new from NPTII, the gene encoding for kanamycin resistance. It was confirmed that the DNA fragment used for RNAi had been correctly inserted to the intended sites on the vector (Figure 4.20c), forming a double-stranded hpRNA in the transformed plant.

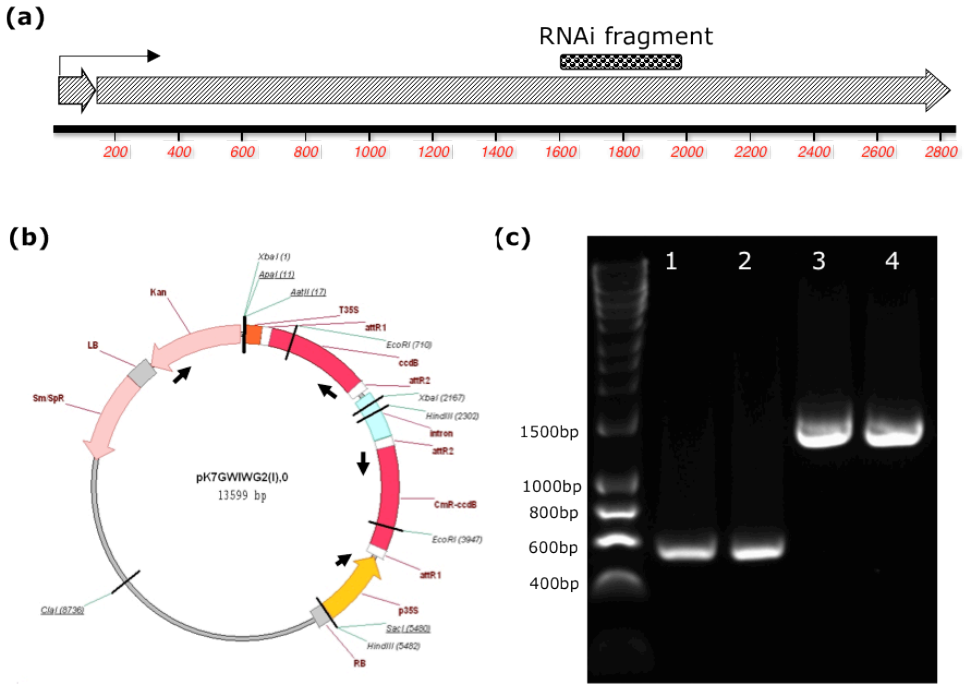


Figure 4.20 Generation of RNAi construct. (a) Diagram showing the region for RNAi on the cDNA of Y2H54. Arrow boxes indicate exons, black arrow represents the ATG start codon. Scale bar:bp. (b) Vector PK7GWIWG2 used for RNAi. Primers for colony screening are shown as black arrows. (c) PCR screening of the recombinant RNAi construct. Lane 1 and 2, independent colonies screened using a primer from the CaMV35S promoter and a gene specific primer. Lane 3 and 4, independent colonies screened using gene specific primer and primer from kanamycin resistance encoding gene NPTII.

4.3.4.2 Genotypic analysis of RNAi transformants

The construct confirmed by DNA sequencing was transformed into an *Agrobacterium* C58 competent cells used for plant infection. Seeds of the infected plants were harvested as T1 generation and PCR screened for genotypic analysis using wild type L*er* genomic DNA as a negative control (Figure 4.21). 4 out of 30 checked plants were identified as transgenic, these seeds were grown to the T2 generation to be further examined by expression and phenotypic analysis.

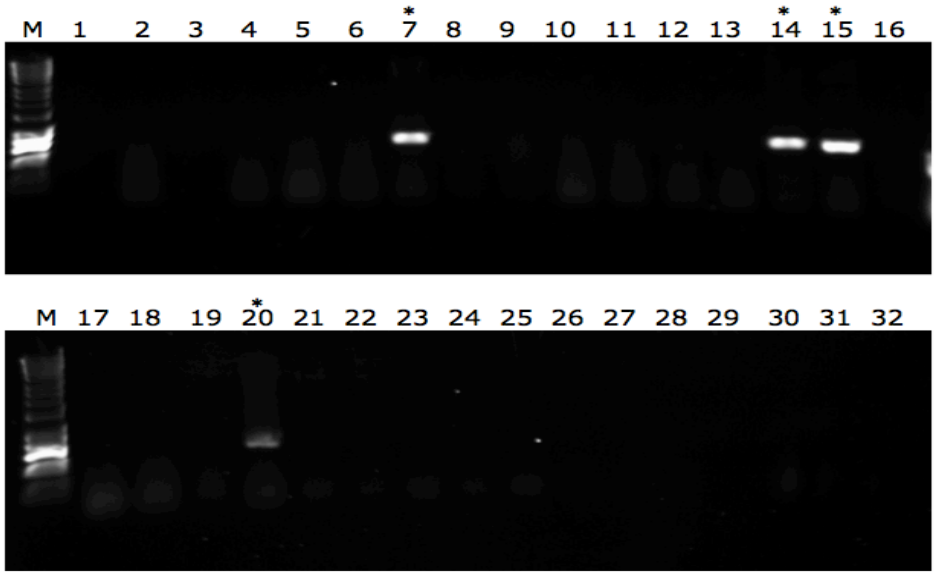


Figure 4.21 Genotypic analysis of RNAi T1 generation. Plants identified as transgenic are indicated. M, Hyperladder I (Bioline). Lane 1 and 17, PCR using wt L*er* genomic DNA as templates. Lane 2-16 and 18-32, T1 generation plants of the Y2H54 RNAi line. Primers are using gene specific primer 1980R plus primer derived from CaMV35S promoter as previously indicated (Table 4.3).

4.3.4.3 Expression analysis of Y2H54 RNAi lines

Expression level of Y2H54 in 2-3 individual plants of each transgenic line was checked by qRT-PCR, using three different Y2H54 primer pairs intentionally designed at different positions on the cDNA of Y2H54 (Table 4.4). PCR results were consistent within the three pairs of primers used, however, the efficiency of RNA interference varied among different transgenic lines, resulting in various levels of gene silencing. Approximately 50%, 70% and 90% of the Y2H54 expression was decreased in the line T1-2, T1-3 and T1-4 respectively, while no more than 10% was reduced in the line T1-1 (Figure 4.22).

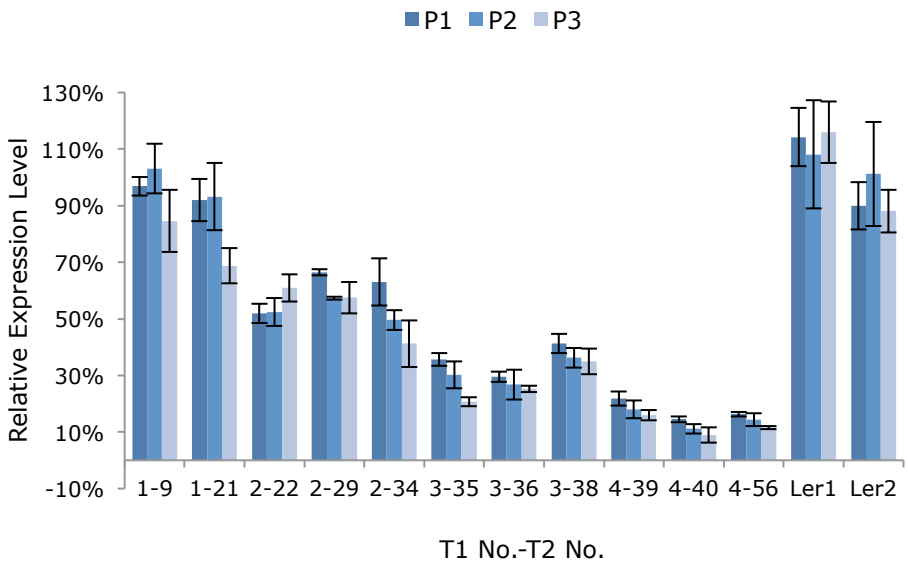


Figure 4.22 Quantitative PCR analysis of Y2H54 RNAi lines. Each gene was examined using three pairs of primers (P1, P2 and P3). *Actin7* gene was used as the control.

4.3.4.5 Phenotypic analysis of RNAi lines

Progenies of line 3 and 4 with 70% and 90 %, respectively, reduced expression of Y2H54 were used for phenotypic analysis (Figure 4.23). The pollen development was checked using Alexander staining (Aexander, 1969), and showed that no abnormal pollen was produced in the Y2H54-RNAi lines (Figure 4.23a). In agreement with this, the siliques of the transgenic lines did not show a significant difference in length compared to wild type L*er* (Figure 4.23b). No other phenotypes were observed in the Y2H54 RNAi lines, compared with the wild type. This may be caused by the residual function of Y2H54 in the transformants, that served as a 'knock-down' rather than the null mutant, or alternately, due to the presence of proteins that are functionally redundant with Y2H54.

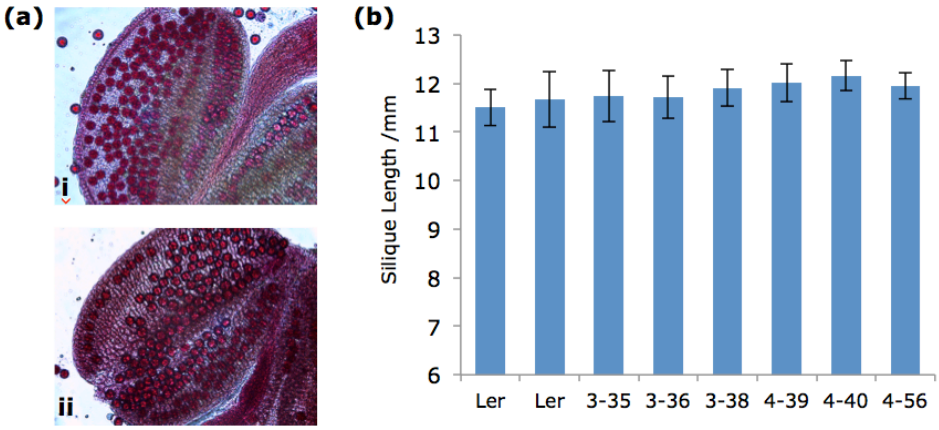


Figure 4.23 Phenotypic analysis of RNAi lines. (a) Pollen viability checked by Alexander staining (Aexander, 1969). The RNAi lines (i) are able to produce viable pollen as seen in the wt L*er* (ii). (b) Sillique length of plants. Siliques from 10 individual plants of each line are harvested when at least 10 siliques occur on each branch.

4.3.5 Analysis of Y2H54 Over-expression Lines

Y2H54 gene was overexpressed under the control of the CaMV35S promoter in both wt L*er* and the *ms1ttg* mutant. Transgenic lines were screened and identified by genomic DNA PCR. The over-expressed *Y2H54* in both backgrounds were confirmed by RT-PCR (Figure 4.24).

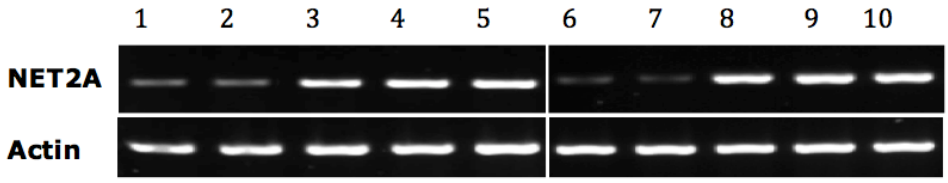


Figure 4.24 Semi-quantitative expression analysis of *35S::Y2H54* lines. Left, expression analysis of overexpression in wt L*er* background; overexpression of *Y2H54* was detectable in the transgenic lines (Lane 3, 4, and 5) when compared with wt L*er* (Lane 1 and 2). Right, expression analysis of overexpression in *ms1ttg1* mutant background; overexpression of *Y2H54* was detectable in the transgenic lines (Lane 6, 7, and 8) when compared with *ms1ttg1* mutant (Lane 5 and 6).

General plant growth of the overexpression lines was enhanced compared with wt L*er*, with larger rosettes, higher stems and more branches observed (Figure 4.25). Interestingly, these lines also exhibited an assembling phenotype with *ms1ttg* mutant in inflorescence sizing, generating more and larger open flowers and buds in each inflorescence, compared with wide type L*er*. This phenotype was strengthened when Y2H54 was overexpressed in the *ms1ttg* background, showing the largest inflorescences compared with all the other lines (Figure 4.26).

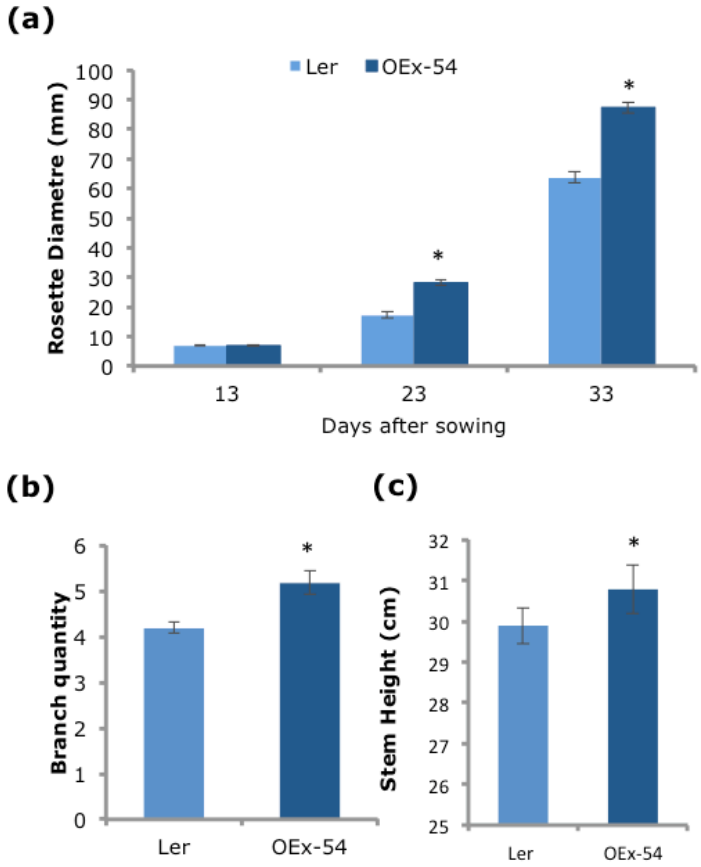


Figure 4.25 Quantitative phenotypic analysis of general plant growth. Plants used for phenotypic analysis were screened as described in Section 2.1, and grown under the same condition to minimize the environmental influences. Statistical analysis of growth using t-test was carried out on a total of 15 individual plants from each line. (a) Rosette diameter. (b) Total branch quantity per plant. (c) Final stem height.

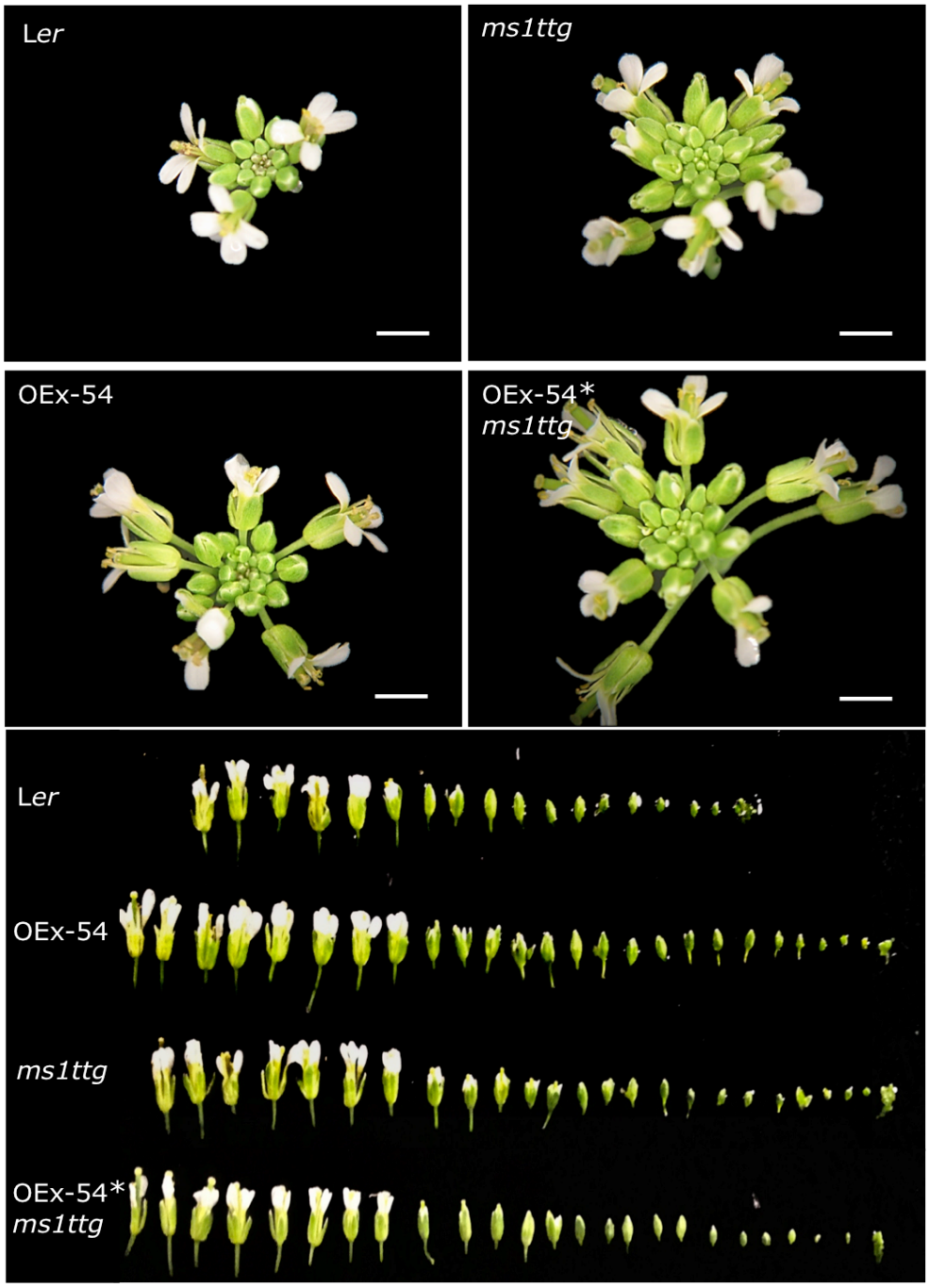


Figure 4.26 Phenotypic analysis of floral development. Overexpression of Y2H54 results in plants with larger inflorescences containing more and larger open flowers, compared with non-transgenic plants. The overexpression levels of Y2H54 in both L*er* and *ms1ttg* background were confirmed by semi-quantitative PCR (Figure 4.22). The first inflorescence from 28-day-old plants was harvested.

4.4 DISCUSSION

Y2H54, namely NET2A, belongs to the NET actin-binding superfamily that serves as the actin-membrane nexus, whose NAB motifs have been proved capable of binding F-actin (Deeks et al., 2012). This superfamily consists of four subclades, NET1 to NET4 groups, which have been found recruited to diverse membrane targets in tissues. NET1, NET2 and NET3 group members respectively localize to the plasma membrane and plasmodesmata, vacuole membrane, endoplasmic filaments and nuclear membrane, according to previous research (Deeks et al., 2012; Hawkins et al., 2014). Y2H54, corresponding to NET2A, is regarded as a pollen-specific protein, that has been found to form submicron foci particularly at the growing *Arabidopsis thaliana* pollen tubes membrane, but was absent from the pollen tube tip (Deeks et al., 2012). However, it is interesting for us to note that Y2H54-GFP associated with the somatic tapetum membrane， which may indicate that it has involvement in multiple biological roles.

In petunia, interaction between the NET2 protein KIP1 and membrane-integrated receptor-like kinase PRK1 (Skirpan et al., 2001), may serve as an archetype for NET2A interactions in *Arabidopsis thaliana*. As the first identified pollen-expressed leu-rich receptor-like kinase, PRK1 in petunia is required for unicellular stage microspores approaching bicellular stage. Reduced levels of PRK1 expression gave rise to aborted pollen grains arrested at the unicellular stage of microspore development. It showed strong interaction with KIP1 proteins at its own kinase domain in the yeast

two-hybrid screens. Temporal expression pattern analysis revealed that KIP1 transcripts were detected in the anthers from microspore mitosis and peaked at the stage of mature pollen grains (Skirpan et al., 2001), which is similar with that of NET2A that was observed in *Arabidopsis thaliana* by qRT-PCR (Figure 4.14). Total expression levels of NET2A were higher in the mature pollen grains. If this analogy to petunia holds true, characterization of NET2A may lead to the identification of not only further novel cytoskeletal proteins, but also proteins that regulate signal transduction pathways in the pollen grains, or even the tapetum.

KIP1 and PRK1 expressed in the microspores has been confirmed by *in situ* hybridlization (Skirpan et al., 2001), but the NET2A-GFP fusion proteins were observed on the tapetum membrane around the unicellular microspore stage (Figure 4.15 a-b). This might indicate that NET2A acts in a parallel mechanism in the tapetum to the pollen grains. An equivalent to PRK1 in the tapetum may get involved in the cell patterning mechanism by coordinating cell proliferation and differentiation, which is critical for proper anther development. To date there are 223 *Arabidopsis thaliana* genes encoding leucine-rich repeat receptor-like kinases (LRR-RLK), some of which have been shown to be critical for anther and pollen development, including EMS1/EXS, SERK1/2 and BAM1/2.

One of the LRR-RLKs EXCESS MICROSPOROCYTES 1 (EMS1) plays a crucial role in specifying tapetum cell layers. Mutants of *ems1* appear to form extra meiocytes at the expense of tapetal cells, whose initials fail to

proliferate (Canales et al., 2002; Zhao et al., 2002). EMS1/EXS has been proposed to interact with putative co-receptors SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1/2 (SERK1 and SERK2), whose double mutant displays resembling phenotype of defected anther as that of *ems1* (Shah et al., 2002; Albrecht et al., 2005; Colcombet et al., 2005). BARELY ANY MERISTEM1/2 (BAM1/2) act redundantly to define anther somatic cell layers at early stages, including the endothecium, middle layer, and tapetum. Double mutants of BAM1 and BAM2 are defective in these layers, and PMCs degenerate subsequently (Hord et al., 2006). Moreover, *Arabidopsis thaliana* RECEPTOR-LIKE PROTEIN KINASE2 (RPK2) is required for normal cell layer differentiation. In the *rpk2* mutants, the middle layer was absent, with PMCs undergoing meiosis but degenerating afterwards, which eventually cause defective anther dehiscence and pollen maturation (Mizuno et al., 2007). In summary, mutation of these LRR-RLK genes commonly causes male sterility, but the signaling pathways behind these defects have not been well defined.

On the other hand, coordinated actin-binding proteins and other aspects of cytoskeletal behavior have been hypothesized to be essential for maintaining proper cell morphology by affecting cell growth (Hussey et al., 2006). Particularly, filamentous actin (F-actin), which the NAB motifs of NET proteins bind to, is responsible for coordinating the active movement of vesicles and organelles through the cytoplasm to the cell surface when cell growth takes place (Valster et al., 1997).

Evidence shows that manipulation of the actin-binding proteins may cause the abnormality of cell growth by altering F-actin organization (Hussey et al., 2006). Dong tested the consequences of overexpression and inhibition of ADF1 (Grefen et al.), an actin binding protein in *Arabidopsis thaliana*. Overexpression of ADF1 causes irregular cellular and tissue morphogenesis and reduces the growth of cells and organs, inhibiting the pollen tube growth (Chen et al., 2002). Interestingly, *Arabidopsis thaliana* plants overexpressing profilin isoform PFN1 do not show growth inhibition, with some cell types showing excessive expansion instead (Ramachandran et al., 2000); this phenomenon is also observed in overexpression of Y2H54. Another example is overexpression of RIC4, causing the formation of F-actin in the pollen tube apex stimulated (Gu et al., 2005).

Abnormal plant morphology have been observed in many organs by overexpression of NET2A. These transgenic lines grow faster with enlarged size of rosette leave, visibly sturdier stems and more branches compared with the wild type. Despite the low likelihood that NET2A and MS1 interact with each other, it is still interesting for us to see enlarged inflorescences as a common phenotype in both the overexpression lines of NET2A and the *ms1ttg* mutants, the phenotypes of both which are intensified further. It appears that knocking out of MS1 lead to simulated cell growth, which may prove to be a hint for the linkage between controlled actin activities and MS1 regulating pathways.

In summary, we predict that NET2A may play a role in regulating anther development via the receptor-like kinase mediated signal transduction pathways, or actin cytoskeleton activities, or by a combination of both.

Unfortunately, we failed to observe significant phenotypes in the RNAi lines, which actually represent the 'knock-down' rather than a null mutant of Y2H54. Transgenic lines with maximal silencing efficiency still maintained approximately ten percent expression level, which can be sufficient to preserve the biological function in plants. Since not all small RNAs are equally effective when at silencing a gene, it might be worth changing the structural region used for RNAi.

There are four *Arabidopsis thaliana* isoforms in total belonging to the NET2 protein subclade, all of which appears to be pollen specific based on publicly available DNA microarray data visualized with gene investigator (Hawkins et al., 2014). Y2H54_Like (NET2B) shares the highest homology with Y2H54 (NET2A), but displays a very different expression pattern, according to our qRT-PCR results. The possibility of functional redundancy existence cannot be eliminated until further experiments are conducted, such as characterisation of the double mutants of the two genes. To investigate the biological function of Y2H54 in regulating anther development, further characterization of the gene should be carried out by using knockouts generated in alternative ways to achieve complete silence, such as artificial mRNA. Double mutants, or multiple mutants of Y2H54 and other members of the NET2 subclade may help to avoid possible functional redundancy.

Simultaneously, isolation of the putative interacting proteins of Y2H54 could potentially identify novel PRK proteins, or associations with known actin proteins, which would provide insights into how Y2H54 functions in the regulatory mechanisms of anther development.

CHAPTER 5 IDENTIFICATION OF MS1 DIRECT REGULATORY NETWORKS

5.1 INTRODUCTION

Regulation of the tapetal specification and pollen wall formation involves complex genetic pathways embracing large numbers of transcription factors, particularly DYT1, AMS, MS1 and MYB99 (Wilson et al., 2001; Sorensen et al., 2003; Zhang et al., 2006; Alves-Ferreira et al., 2007; Ito et al., 2007; Yang et al., 2007; Xu et al., 2010; Feng et al., 2012; Ma et al., 2012) .

The *DYT1* gene encodes a putative basic helix-loop-helix (bHLH) transcription factor, critical for early tapetum development and tapetal gene regulation (Zhang et al., 2006). Reduced expression level of *DYT1* gives rise to a highly vacuolated tapetum (Zhang et al., 2006), and large numbers of the tapetum transcription factors are severly reduced in *dyt1* mutant (Feng et al., 2012). Microarray analysis shows that DYT1 positively regulate genes involved in lipid metabolism and transport, cell wall modification, as well as secondary metabolism, particularly, AMS, MS1, ACOS5 and MS2 (Feng et al., 2012).

AMS, belongs to the MYC subfamily of bHLH genes, required for both tapetum development and postmeiotic microspore formation (Sorensen et al., 2003). The *ams* mutant is completely male sterile due to tapetal cell and microspore degeneration. In the mutant, expression of genes involving transport and metabolism of lipid, carbohydrate and secondary metabolites

are down-regulated (Xu et al., 2010; Ma et al., 2012); 13 genes have been identified as AMS direct targets by ChIP-qPCR (Xu et al., 2010; Ma et al., 2012) and more recently further target involved in pollen wall development have been identified (Xu et al., 2014)

As mentioned in Chapter 1, MS1 is responsible for normal sporopollenin biosynthesis and pollen exine formation (Wilson et al., 2001; Ito et al., 2007; Yang et al., 2007). Defected exine formation and organisation are found in the *ms1* mutant (Ito et al., 2007; Yang et al., 2007). Genes associated with biosynthesis of long-chain fatty acids, putative components of the sporopollenin, are proposed as MS1 downstream genes by Ito (Ito et al., 2007). Correspondingly, among the genes exhibiting expression changes in the *ms1* mutant, as indicated by microarray analysis, a high proportion are involved in wall biosynthesis and lipid metabolism (Yang et al., 2007). The major role that MS1 plays in the control of late pollen development is confirmed as regulating lipid metabolism, cell wall biosynthesis pathways, and lipid transfer proteins that participate in the biosynthesis of sporopollenin and pollen coat materials (Yang et al., 2007).

The NAM protein MYB99 was identified as an immediate target of MS1, which shows highest level of repression in *ms1* mutant (Alves-Ferreira et al., 2007; Ito et al., 2007; Yang et al., 2007). This has been confirmed by using a dexamethasone (DEX)-inducible MS1 construct that MYB99 is induced by MS1 expression without the requirement for *de novo* protein synthesis (Ito et al., 2007; Yang et al., 2007). Furthermore, MYB99 has been found to

be regulated by ASHH2 via H3K36 trimethylation, a protein that is thought as a collaborating partner of the MS1 protein in regulating down-stream targets (Grini et al., 2009).

These transcription factors assemble complex regulatory pathways via feed-forward loops, or through direct protein interaction. Down-regulated in both the *dyt1* and *ams* mutants, MS1 is proposed as one of the commonly shared down-stream targets of DYT1 and AMS (Sorensen et al., 2003; Zhang et al., 2006). Additionally, DYT1 exhibits the ability to bind to the promoter region of MS1, which suggests that MS1 is a putative direct target of DYT1 (Feng et al., 2012). A few regulation forward loops have been proposed, as DYT1-MS1-MYB99, DYT1-MYB35/TDF1-AMS, and AMS-MYB103/MYB80-MS1 (Feng et al., 2012). Moreover, both as bHLH transcription factors, the DYT1 and AMS proteins potentially interact with each other, and are capable of interacting with other bHLH proteins to form heterodimers (Xu et al., 2010; Feng et al., 2012; Ma et al., 2012). Alternative interacting proteins of DYT1 include AtbHLH089 (At1g06170), AtbHLH010 (At2g31220), and AtbHLH091(At2g31210) (Feng et al., 2012); AMS has also independently been shown to interact with AtbHLH 089, AtbHLH 091 and also ATA 20 (Xu et al., 2010).

To get deeper insight of the correlation between members in the regulatory networks, a new bioinformatic method Transclust (Transitivity Clustering) was used to analyse our experimental datasets. Transitivity Clustering is a method for partitioning biological data objects into clusters, within each of

which objects share more similarity than those from other clusters (Wittkop et al., 2011). General steps for clustering include calculation of the correlation coefficient of the log2-transformed gene expression values, estimation of their threshold values, based on which subsequent clustering can be applied, and finally visualization of the clustered results graphically (Figure 5.1). This provides a powerful tool for grouping of gene expression data sets. Our lab established a 'FlowerNet' gene expression framework by combining lists of Affymetrix microarray chips from different experiments that have been identified into groups with similar gene expression patterns (Dr Simon Pearce, University of Nottingham). Subsequent Gene Ontology (GO) overrepresentation analysis of each of the clusters, a commonly used approach for functional studies, has helped aid in discovery of both the biological processes and molecular functions of genes of interest, facilitating the hypotheses of how their functions relate to the experimental conditions. Mapping the 'FlowerNet' clusters with particular microarray dataset, such as *MS1,* would help to identify direct and indirect regulatory targets of MS1.

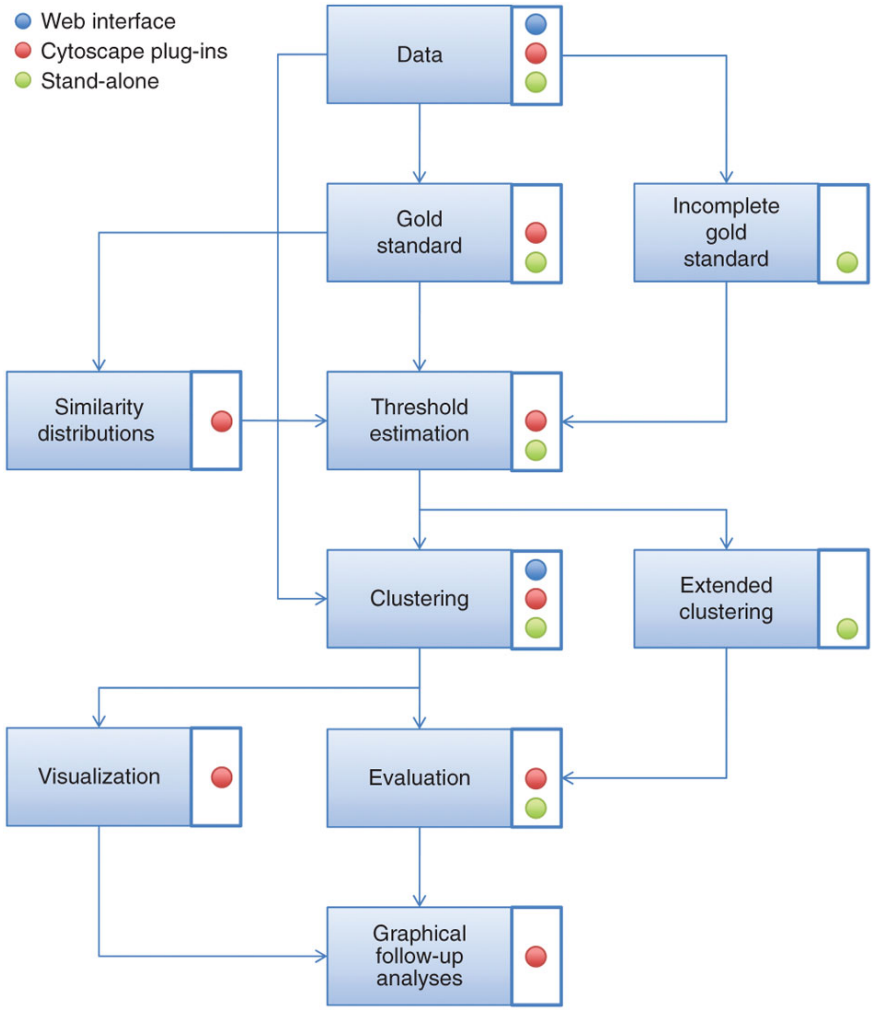


Figure 5.1 Overview of the transitivity clustering functionalities and user interfaces. This figure outlines the organization of this protocol as well as the structure of typical clustering workflows. Steps marked with a blue dot can be performed with the web interface, steps with a red dot with the Cytoscape plug-ins and steps with green dots with the stand-alone software.

The Chromatin immunoprecipitation (ChIP) technology provides a way to find target genes of the transcription factors, by studying interactions between transcription factors and DNA *in vivo* (Kaufmann et al., 2010). General ChIP procedures consist of several basic steps (Figure 5.2). Plant tissue is fixed using formaldehyde in order to crosslink protein–DNA interactions. The chromatin is then purified from cell debris and sheared

into small fragments. Subsequently, a specific antibody is applied to pull-down the protein-DNA complex, following by DNA recovered using reverse cross-linking, which is then used for Realtime quantitative PCR (qChIP-PCR) analysis. The DNA obtained in ChIP experiments can be identified using direct sequencing (ChIP-SEQ) (Bowler et al., 2004; Haring et al., 2007) or hybridization to whole genome arrays (ChIP-CHIP) (Johnson et al., 2007). This methodology has provided numerous applications in studying DNA methylation, chromatin structure/histone modifications, and the cooperative binding of transcription factors (Greb et al., 2007; Zhong et al., 2007; Massie and Mills, 2009).

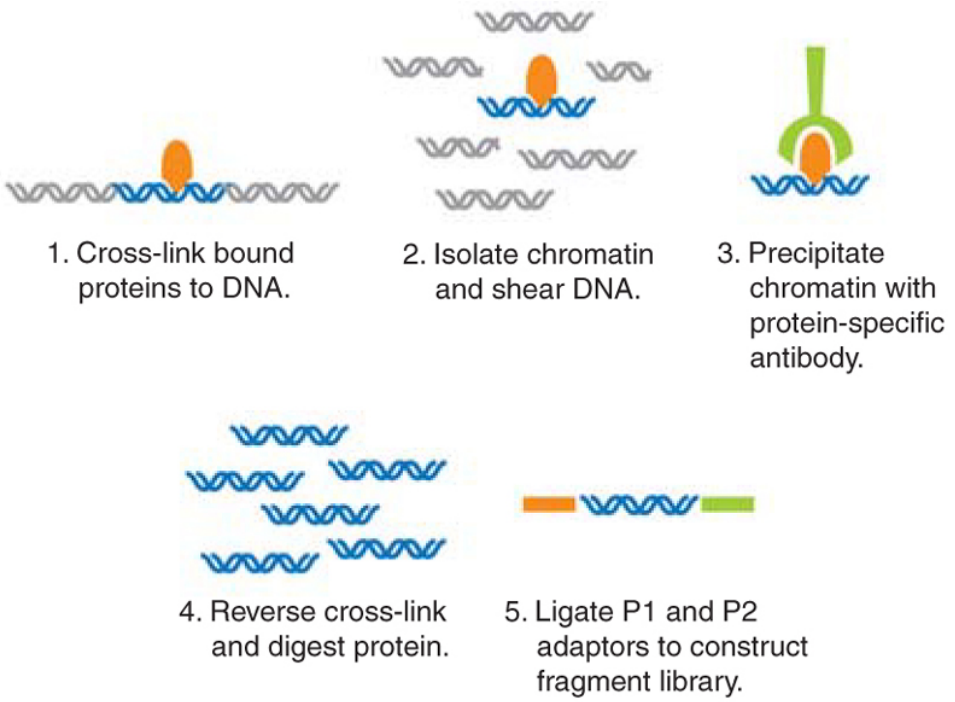


Figure 5.2 DNA enrichment by chip procedures and library construction (Shah, 2009). Cross-link the protein–DNA with formaldehyde. Isolate the chromatin and shear the DNA by sonication. DNA–protein complex are immunoprecipitated using specific antibodies. Reverse the cross-linking of DNA–protein complexes to release the DNA and digest the proteins. Finally, DNA precipitation and PCR amplification are performed.

In this chapter, as a putative direct target of MS1 (Ito et al., 2007; Yang et al., 2007) (Alves-Ferreira et al., 2007), MYB99 was used to verify the ChIP assay of MS1, by checking the enrichment of the MYB99 gene. This approach would provide a way to test insights of MS1-DNA binding, and also provide a positive control for further ChIP experiments with MS1.

5.2 MATERIALS AND METHODS

5.2.1 Comparative Transcriptome Analysis

An *Arabidopsis thaliana* correlation network (FlowerNet) of global transcriptional interactions associated with anther development was generated in our lab by Dr Simon Pearce, using publicly available gene expression data and data maintained in our own lab. This networks comprised of 66 Affymetrix Microarray chips in total from experiments including whole buds, isolated stamens or pollen samples from different sources (Appendix IX). The 66 chips were re-normalised using the CustomCDF and RMA normalisation, to make the expression values comparable across the experiments. Genes with no expression in any of the 66 samples (log2 value less than 6, determined by the histogram) were removed, and the pairwise correlations between the remaining 17,807 genes were calculated. Pairs of genes with a correlation above 0.88 were connected by an edge in the resulting network. There were 10,797 genes in this network, with 605,686 edges between them.

The network was then clustered using TransClust analysis to find subsets with almost all the possible connections, with the largest (Cluster #1) being 171 genes. To visualize and analyze the gene profiles of those regulated by MS1 and other transcription factors, the network was then mapped with available microarray data from the transcription factors, typically MS1

microarray data (available from NASC: http://affy.arabidopsis.info/narrays/experimentpage.pl? experimentid=23) (Yang et al., 2007), using Cytoscape software (http://www.cytoscape.org). Gene Ontology (GO) Overrepresentation analysis (generated in topGO in R) was carried out on each of the clusters identified as putative regulatory targets, revealing both the biological processes and molecular functions.

5.2.2 Plant Material Fixation

The *Arabidopsis ms1ttg* mutant that was restored by a functional MS1-GFP fusion under the control of MS1 endogenous promoter is maintained in our lab (Yang et al., 2007). Seeds were screened on half MS medium plus kanamycin (Section 2.10.2) and cultured for 10 days under full light ($140\mu mol \cdot m^{-2} \cdot sec^{-1}$) at 22-24°C (Section 2.2). Healthy seedlings that still carried the *ttg* phenotype, lacking trichomes on the leaves, were transferred into soil (Section 2.1). Approximately 4 weeks after sowing, fertile plants that carried at least five full length siliques were used for RT-PCR analysis to confirm the existence of the MS1-GFP fusions, using cDNA from the whole buds (Section 2.4 and 2.5) and gene specific primers for MS1 and GFP (Table 5.1). Whole buds of these rescued *ms1ttg1* mutants were harvested into a 50ml tube on ice.

Table 5.1 Primers used for RT-PCR.

Target DNA	Primer	Sequence 5'-3'
GFP	EGFP F	ATGGTGAGCAAGGGCGAGGA
	EGFP R	CTTGTACAGCTCGTCCATGCC
MS1	MS1 F	CCATTGCCAATATGTTGGTTG
	MS1 R	CAGCCTCAACTCCATTCCTT
MS1-GFP	MS1_5168L F	TGAGTGTGGAGCAACGGAAG
	EGFP R	CTTGTACAGCTCGTCCATGCC
Actin	Actin 7_512 F	GCCATTCAGGCCGTTCTTTCT
	Actin 7_876 R	CGGAATCTCTCAGCTCCGATG

Chromatin immunoprecipitation assay was adopted from Haring et al.(Haring et al., 2007). All steps were carried out at 4°C and all solutions were kept at 4°C unless stated otherwise. Whole buds of the plants were harvested and cross-linked in 30ml of Extraction Buffer 1 (0.4 M sucrose, 10 mM Tris-HCl, pH 8.0, 5 mM β- mercaptoethanol, 0.1 mM PMSF, protease inhibitors (Sigma)) containing 1% (v/v) formaldehyde for 10 min under vacuum. Cross-linking was quenched in 0.125 M glycine for 5 min under vacuum. The materials were then washed four times with ddH_2O, and stored at -80°C until used.

5.2.3 Chromatin Isolation and DNA Fragmentation

1-5 grams of plant materials were ground into a fine powder in liquid nitrogen, resuspended in a 50ml tube containing 30ml of pre-cold Extraction Buffer 1, and incubated for at 30mins at 4°C with gentle agitation. Solutions were filtered through four layers of Miracloth (Calbiochem) before centrifuged at 2,880×g for 20mins at 4°C. Supernatant carefully removed

and the pellet resuspended in an Eppendorf tube carrying 1ml of ice-cold Extraction Buffer2 (0.25M sucrose, 10mM Tris-HCl, pH 8.0, 10mM $MgCl_2$, 1% (v/v) Triton X-100, 5mM β- mercaptoethanol, 0.1M PMSF, protease inhibitors (Sigma)). Solutions were centrifuged at 12,000 x g for 10mins at 4°C, supernatant discarded. The pellet was then resuspended in 300µl of ice-cold Extraction Buffer 3 (1.7M sucrose, 10mM Tris-HCl, pH 8.0, 0.15% (v/v) Triton X-100, 2mM $MgCl_2$, 5mM β- mercaptoethanol, 0.1M PMSF, protease inhibitors (Sigma)). The nuclei suspension was then carefully overlayed on top of a cushion of 500µl of ice-cold Extraction Buffer 3 contained in an microcentrifuge tube and centrifuged at 16,000×g for 1 hour at 4°C.

Supernatant was discarded, the chromatin pellet re-suspended in 300µl of ice-cold Nuclei Lysis Buffer (50mM Tris-HCl, pH 8.0, 10mM EDTA, 1% (w/v) SDS, 0.1 M PMSF, protease inhibitors (Sigma)). A 10µl aliquot was removed to a new tube and kept on ice, representing 'unsheared' chromatin, and stored at -80°C. The chromatin solution was subsequently sonicated on ice for pulses of 15secs at 3micron amplitude (4 times, with 1min rest between pulses) in a Soniprep 150 sonicator (Sanyo), to shear DNA to manageable sized fragments of 200-800 bp or 200-400 bp for ChIP-seq.

5.2.4 Chromatin Immunoprecipitation

The sonicated chromatin suspension was centrifuged at 16,000g for 5mins to pellet debris. Supernatant containing sonicated chromatin was transferred to a new tube on ice, a 10µl aliquot was taken from the solution as the

'sonicated chromatin'. Remaining chromatin was diluted with 1660µl of ChIP Dilution Buffer (1.1% (v/v) Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl, pH8.0, 167mM NaCl, 0.1M PMSF, protease inhibitors (Sigma)) and incubated with 40µl of Dynabeads® protein A agarose beads (Invitrogen) for 1hour with gentle agitation (rotation at 12 rpm).

The tubes were then placed on a magnetic rack to pellet the beads, a 55µl aliquot taken from the chromatin solution as 'input' sample and stored at -80°C. Remaining solutions were divided into three proportions and transferred into new tubes, containing 550µl of chromatin each, one of which would serve as the 'no-antibody (NoAb)' control. Subsequently, 1-10µl of commercial rabbit polyclonal anti-GFP antibody (AbCam, ab290) was added to each of the IP tubes, equal volume of non-relevant polyclonal anti-HA (Hemagglutinin) antibody (Sigma-Aldrich, H6908) added to the NoAb tube. Inmomnopreciptation was performed by incubation overnight with gentle agitation (rotation at 12 rpm). 20µl of Dynabeads® protein A agarose beads (Invitrogen) was added to the immunoprecipitates and incubated for at least 2.5 hours with gentle agitation (rotation at 12rmp).

Beads were pelleted on a magnetic rack and supernatant discarded, which was then washed with each of the following buffers for 10mins at 4°C with gentle agitation: 1 x Low-Salt Buffer (150 mM NaCl, 0.1% (w/v) SDS, 1% (v/v) Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0), 1 x High-Salt Buffer (500 mM NaCl, 0.1% (w/v) SDS, 1% (v/v) Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0), LiCl wash buffer (0.25 M LiCl, 1% (v/v)

NP-40, 1% (w/v) sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0), and then 2x with TE buffer (10mM Tris-Cl, pH 8.0, 1mM EDTA). After each wash, beads were pelleted using a magnetic rack and supernatant discarded. Immune complexes were eluted by incubation in 250µl of Elution Buffer (1% [w/v] SDS, 0.1 M $NaHCO_3$) at 65°C for 15mins with gentle agitation. Beads were pelleted on a magnetic rack and eluates transferred to a new tube. This was repeated and the eluates pooled together.

5.2.5 Reverse Cross-linking and DNA Isolation

To reverse cross-link the samples, 20µl of 5M NaCl was added to the eluates of the two IP tubes and the NoAb tube, and the 'Input' sample and 100µl of TE buffer, 6.5µl of 5 M NaCl and 8µl of 20% (w/v) SDS added. All tubes together with previous aliquots of the 'unsheared' chromatin and 'sonicated' chromatin were incubated at 65°C overnight. The two aliquots were then purified by phenol/chloroform method (Section 2.8.7) and run on 4% agarose gel (Section 2.7) to check the sonication efficiency, the ChIPped DNA together with the 'input' was purified using the MiniElute DNA Purification Kit (QIAGEN). Concentration of the DNA sample was determined by NanoDrop 2000 spectrophotometer (Thermo). Quality of the ChIP was then checked by quantitative PCR analysis, after which samples achieve satisfactory enrichments were retained to carry on the ChIP-seq experiments performed by the commercial company SourceBioscience (Nottingham, UK).

5.2.6 qChIP-PCR and Data Normalization

1µl of the DNA samples was used as template in a 10µl qChIP-PCR reaction, following the conditions: 95°C for 3mins; 60 cycles of 95°C for 10secs, 50°C for 30secs, 72°C for 20secs; 72°C for 6mins; and then the dissociation programme (1°C per cycle from 55°C to 95°C). Primers were designed based on the sequence of MYB99, covering the promoter, typically 1.8 Kb upstream, and the gene body regions (Figure 5.3, Table 5.2). Each primer pair covered a region of 100-200 bp, which was short enough to amplify fragmented DNA from the ChIP experiments. Data normalization was adopted from Haring et al. (Haring et al., 2007). The '% of input' method was used, which divided the QPCR signals derived from the IP samples by the QPCR signals derived from the input sample to correct for technical variations.

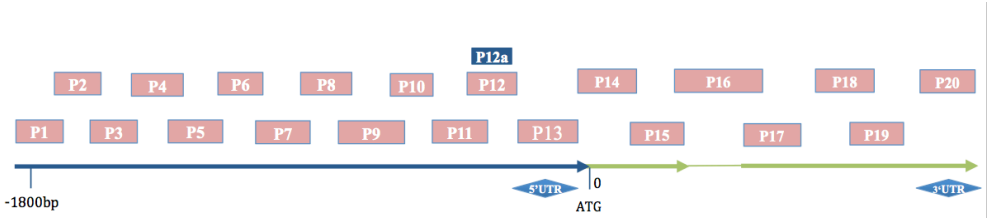


Figure 5.3 Primers specific to *MYB99* gene used for chromatin immunoprecipitation. Features of the genomic structure and DNA fragments covered by 21 pairs of primers (pink) are shown. The blue arrow indicates the promoter and the green ones represent the exons. The region covered by P12a particularly designed within the fragment covered by P12.

Table 5.2 Primers Used for qChIP-PCR Analysis of *MYB99*.

Primer	Sequence 5'-3'
MYB99_P1 F	TGTTTTGTCATGATTGTTGTTTTTTC
MYB99_P1 R	AGTGGAAGATACGTGATTTTTATTTG
MYB99_P2 F	TTCCACTCACTTTCGTTCCA
MYB99_P2 R	AAATTGGCTGTCAACAAAATGTA
MYB99_P3 F	TACATTTTGTTGACAGCCAATTT
MYB99_P3 R	TTCTCATTTTTCTAATCTTTCAAACG
MYB99_P4 F	TCGTTTGAAAGATTAGAAAAATGAGA
MYB99_P4 R	GATAAATATTGGTCGTACTAACGGAAT
MYB99_P5 F	TCAATTCCGTTAGTACGACCAAT
MYB99_P5 R	ACAGAAAGTTAAGCATTTCGCTATGA
MYB99_P6 F	GCGAAATGCTTAACTTTCTGTTT
MYB99_P6 R	TTCGAAACAATCTCAGTTTTATCC
MYB99_P7 F	TTGAAATAATGTGGATAAAACTGAGA
MYB99_P7 R	GAACATGAGACTTGCCTTTGAA
MYB99_P8 F	AACTTCAAAGGCAAGTCTCATGTT
MYB99_P8 R	TCAAATCTTTTATAAACTTGGCTCTTT
MYB99_P9 F	CATACAAAATTCTAAAGAGCCAAGTTT
MYB99_P9 R	TCGTGATTTGATGATATGTCGTT
MYB99_P10 F	ATCAAATCACGAATAATCCAACA
MYB99_P10 R	ACGGTGTTAAATAAGTAATTGTGTGT
MYB99_P11 F	TGAAAGAAGAAAAAGGTTTATATATCG
MYB99_P11 R	TTAAGAAAACTCGATGATCCAAAAG
MYB99_P12 F	GCAAATTCCACAAAAACATCA
MYB99_P12 R	AAGTTTATCTCAACTAACTAACTATATATACACAC
MYB99_P13 F	GGTTACCACTTTTGTGTTGTC
MYB99_P13 R	TATTTATATTATATGGTTTAAAAAAAGC
MYB99_P14 F	ATGGGTGGTCGTAAACCATG
MYB99_P14 R	CAACTTTTGCCACACCTCCT
MYB99_P15 F	GAGGATGGTGCTGGAGAGAC
MYB99_P15 R	TGCCAAGGCGAGCATGAA
MYB99_P16 F	GATCTTCATGCTCGCCTTG
MYB99_P16 R	AAATCGACGATGTGTGTTTGG
MYB99_P17 F	ATGGTCGAAGATTGCAGTGG
MYB99_P17 R	CGGTCTCAGACAGAGGCTTT
MYB99_P18 F	AGTCAACGAGGAGGAAACGA
MYB99_P18 R	TAGAAAGCTCCACGGCTGAT
MYB99_P19 F	CCGTGGAGCTTTCTAATGGA
MYB99_P19 R	CTAAACATCGAAACATCCAAGTTCTA
MYB99_P20 F	AGAACTTGGATGTTTCGATGTTT
MYB99_P20 R	GGGACATTCAAAGTGCAAAG
MYB99_P12a F	CACCGTATTCAATGGTTTTAGCA
MYB99_P12a R	ACACACGTATGGAGTTTCTTGG

5.2.7 Preparation of the ChIP DNA Libraries

ChIP DNA libraries for sequencing were prepared by commercial company SourceBioscience (Nottingham, UK). The libraries were prepared in accordance to the Illumina TruSeq ChIP sample preparation guide (August 2012, rev.A) for Illumina Paired-End Multiplexed Sequencing. Briefly, ends of fragments were repaired by adding a dA base, and Illumina indexing adapters were ligated. Samples were then size selected using Invitrogen 2% agarose E-gels; carried adapter molecules on both ends underwent 17 cycles of PCR to amplify the amount of prepared material. The resulting libraries were diluted 1:20 and validated using Agilent BioAnalyzer High Sensitivity Chip and Qubit High Sensitivity & Broad Range Assays. The results from the Qubit and High Sensitivity chip are used in the following equation to determine the nM concentration of each sample:

nM = ng/ul X (1500/Average bp).

5.3 RESULTS

5.3.1 Comparative Transcriptomic Analysis of MS1 Regulatory Network

5.3.1.1 MS1 regulates pathways of lipid metabolism and pollen exine formation

The FlowerNet correlation work generated in our lab by Dr Simon Pearce (University of Nottingham) contains 10,797 genes, with 605,686 edges between them. Based on all possible connections, genes were grouped into hundreds of clusters, with the largest (Cluster #1) being 171 genes. The network was then mapped with the microarray data of MS1 from previous research (Yang et al., 2007), which had been performed using the *ms1ttg* mutant staged into two developmental groups named young and old:(1) young, from formation of the sporogeneous tissues to pollen mitosis I; (2) old, pollen mitosis II through to dehiscence. Wild type L*er* and the *ttg* mutant at the equivalent stage were both used as the controls, to eliminate influences due related to the *ttg* mutation. The genes that exhibited more than 2-fold change in expression in the mutant were integrated into the FlowerNet correlation networks (Figure 5.4).

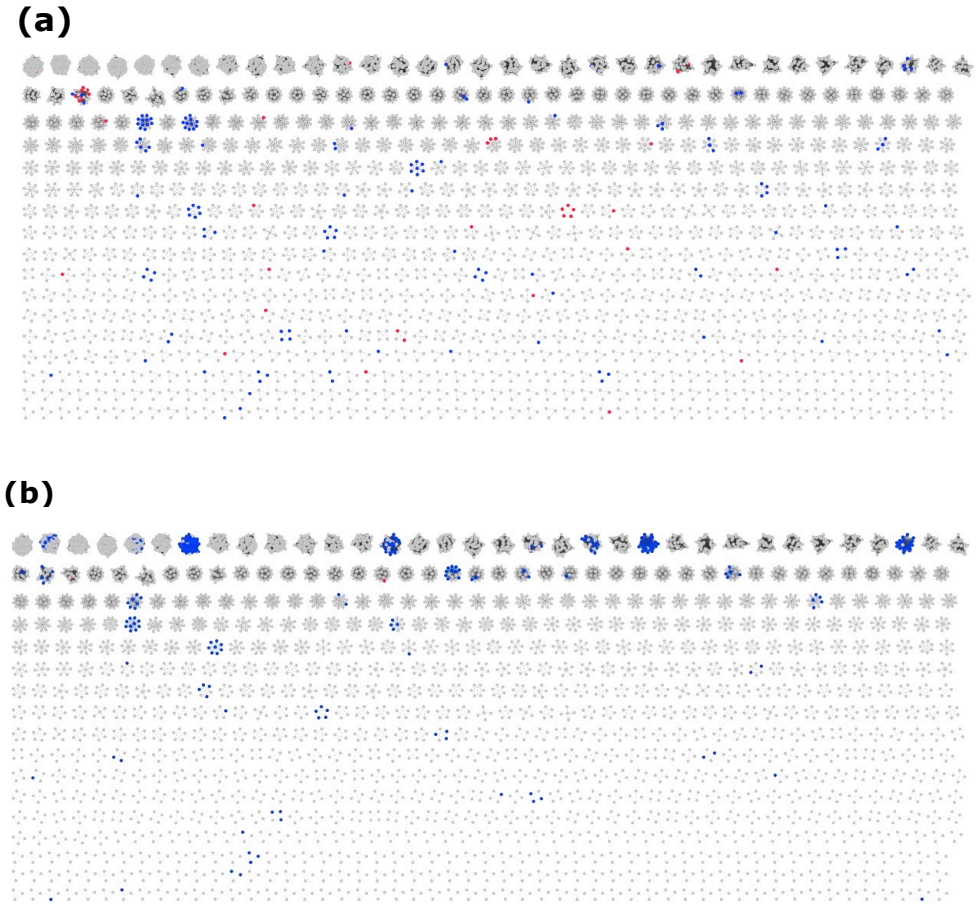


Figure 5.4 Integration of the microarray data of MS1 into the FlowerNet correlation networks. The FlowerNet network was generated as previously described (Section 5.2.1) and clusters comprising no less than three genes are shown. The microarray data was performed using the *ms1ttg* mutant staged into two developmental groups named young and old: young, from formation of the sporogenous tissues to pollen mitosis I; old, pollen mitosis II through to dehiscence (Yang et al., 2007). Genes that exhibited more than 2-fold change in expression as the result of the *ms1* mutation are shown in the clusters, with blue indicating down-regulation and red for up-regulation. (a) Mapping of the microarray data of the young anthers. (b) Mapping of the microarray data of the old anthers.

Clusters, with 4 members at least, with no less than 40% of the members showing expression change were filtered from the FlowerNet correlation networks; 12 clusters for the young anthers prior to Mitosis I stage and 16 for the old ones at Mitosis II stage onwards were identified as putative MS1 regulatory targets using the expression data from the *ms1* mutant (Figure 5.5; Appendix X). Members of four clusters were almost entirely down-regulated in the *ms1* young anthers, including cluster 73, 81, 206 and 263. Particularly, certain members from cluster 37 that showed increased expression in the *ms1* mutant comprised genes that were previously characterised, ACOS5, KNS2, LAP5, DRL1, CYP704B1 and CYP703A2 as introduced in Chapter 1. For the old anthers, reduced expressions of genes from cluster 7, 23, 33, 110 and 156 were discovered in the *ms1*. Those regulated in both the young and old anthers (cluster 73 and 110) may indicate multiple biological functions of the genes, or else as the result of imprecision of the staging method based on anther morphology.

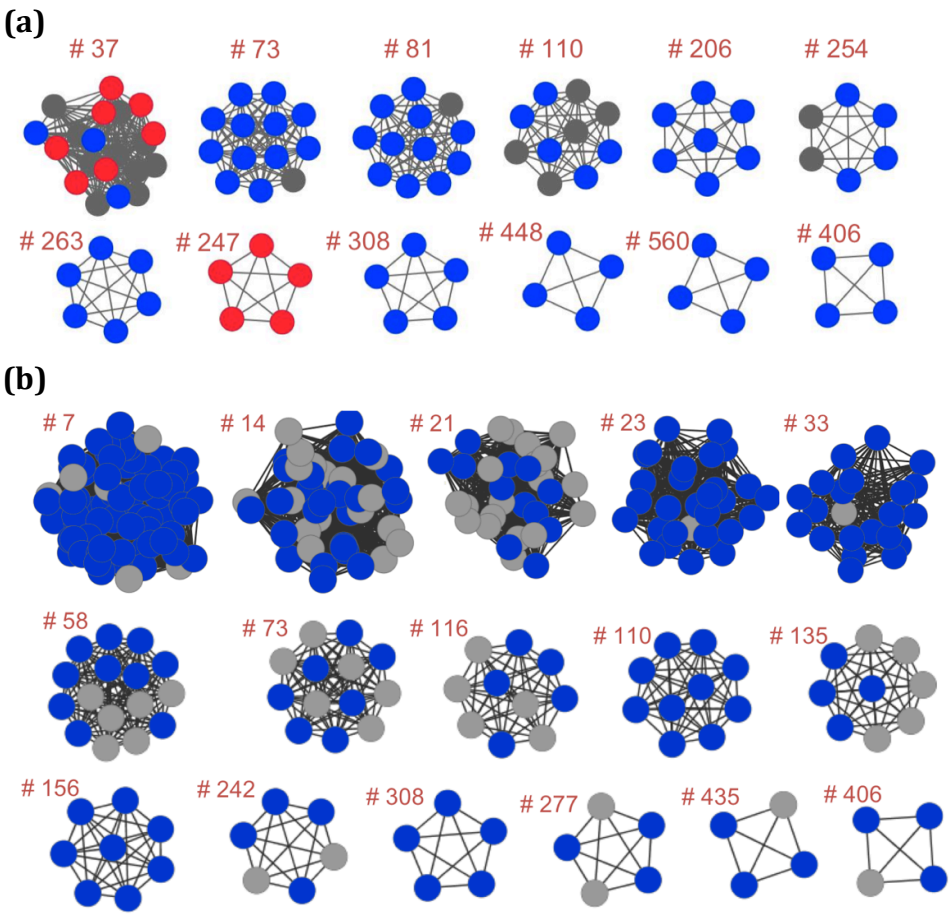


Figure 5.5 Identified MS1 regulatory targets integrated into FlowerNet correlation networks. Each node indicates a certain gene as a member of the numbered clusters**.** Genes down-regulated in the *ms1* mutant are indicated in blue, the up-regulated being marked in red. Those showing no expression change in the *ms1* are shown in grey. (a) Genes identified from the young anthers. (b) Genes identified from the old anthers.

Gene Ontology (GO) overrepresentation analysis was performed on the larger clusters using a GO database set up by Dr Simon Pierce (University of Nottingham). It revealed that MS1 had been critical for normal expression of pathways including lipid mechanism and pollen exine formation in the 'young' anthers, reduced expression in anthers of *ms1* mutant at later stages mainly involving engagement in pathways for cell

wall modification and pollen tube growth (Table 5.3). Altered sucrose transport was also observed in the *ms1*, which might be relevant to the defection in transportation of precursors of sporopollenin. However, it was difficult to define whether the expression change in older anthers had been due to the *ms1* mutation or the absence of normal pollen grains in the anther locules. The noise caused by this factor should be excluded for the purpose of identifying immediate targets of MS1, so that subsequent analysis was focused on the candidates identified from the young anthers.

Table 5.3 Gene Ontology (GO) Overrepresentation Analysis of MS1 Downstream Genes Integrated into FlowerNet Correlation Networks.

	Cluster	Total genes	Proportion regulated	Biological function
Young Anther	37	20	30%	Sporopollenin biosynthesis
	73	13	92%	Lipid storage;
	81	12	92%	Pollen exine formation
	110	10	50%	Actin filament organisation
	206	7	100%	Sexual reproduction
	254	6	67%	Lipid localisation
	247	5	100%	Unknown
	263	6	100%	Pollen exine formation; Lipid transportation
	308	5	100%	Unknown
Old anther	14	43	60%	Cell wall modification; Pollen tube growth
	21	31	39%	Pollen exine formation
	23	29	97%	Cell wall modification; Pollen tube growth
	33	22	95%	ATP catabolic process; Cell wall modification;
	58	16	63%	Cell wall modification; Pollen tube growth
	73	13	92%	Lipid storage
	110	10	100%	Actin filament organisation; Cell wall modification;
	116	10	40%	Pollen tube growth
	135	9	44%	Transmembrane transportation
	156	9	89%	Sucrose transportation
	242	6	67%	Pollination

5.3.1.2 MS1 shares common targets with DYT1 and AMS

The MS1 regulatory targets in the young anthers were mapped alongside the microarray data from the *dyt1* and *ams* mutants compared with wild type (Feng et al., 2012) (Xu et al., 2010). MS1 downstream genes showing more than 2-fold expression change at any anther stage in the *dyt1* or *ams* mutants were identified (Figure 5.6). Some of the genes acted as common targets of DYT1, AMS and MS1, for instance members of cluster 81, while some others appear regulated by AMS independently of DYT1. This indicates that expression of the genes requiring no DYT1, most members of cluster 73 for example, may get involved in alternative regulatory loops.

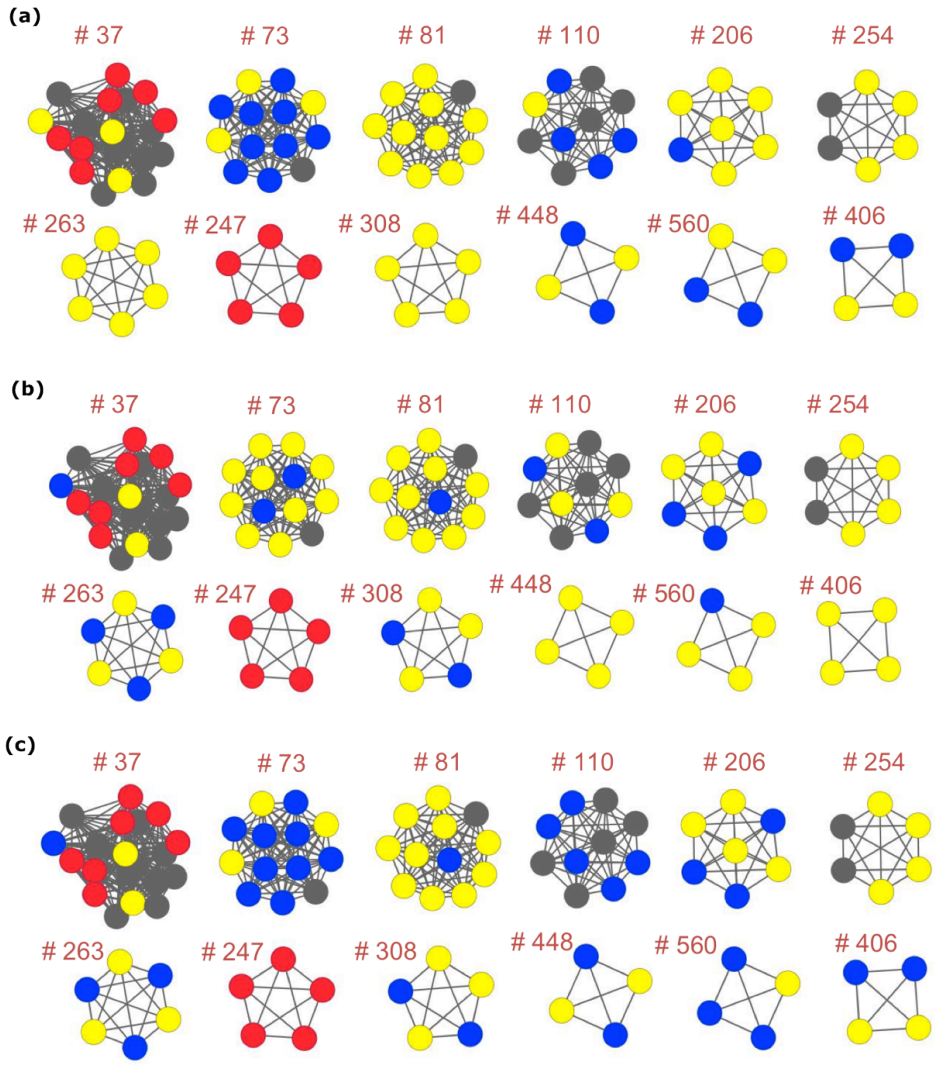


Figure 5.6 Overlapped microarray data in the FlowerNet. Each node represents an individual gene belonging to the cluster, and edges indicate the correlation between the genes. Genes down-regulated in the *ms1* are indicated in blue, with the up-regulated in red. Genes with reduced expression in *ms1* that are simultaneously down-regulated in *dyt1* and/or *ams* are shown as yellow. Grey nodes represent genes with no expression change in the *ms1*. (a) MS1 regulatory targets overlapped with microarray data from *dyt1* (Feng et al., 2012). (b) MS1 regulatory targets overlapped with microarray from *ams* (Xu et al., 2010).(c) MS1 regulatory targets simultaneously regulated by DYT1 and AMS (Table 5.4) .

Many of these genes that simultaneously regulated by MS1, DYT1 and AMS are functionally related (Table 5.4), such as extracellular lipase EXL4 and EXL6; 3-ketoacyl-coa synthase KCS7, KCS15 and KCS21; CYP 450 encoding genes CYP86C3, CYP98A8 and CYP98A9 proteins. Previous studies show that the expression of these genes associated with lipid metabolism and transport (KCS7, KCS15, KCS21, CYP86C3, CYP98A8, CYP98A9), and pollen coat formation (EXL4, EXL6). Some of the candidates in the list were identified as direct targets of AMS, whose promoter regions were enriched by AMS in the ChIP analysis, including EXL4, EXL6, KCS7, KCS15, GRP18, GRP19, CYP98A8, CYP98A9, and At1g06990 (GDSL-like Lipase) (Xu et al., 2010). However, these genes were dramatically down-regulated in the *ms1* mutant, which indicated that AMS alone had not been sufficient to activate the transcription of these genes, with MS1 or the MS1 downstream regulators still being required.

Table 5.4 List of Representative MS1 Downstream Targets Simultaneously Regulated by DYT1 and AMS.

Genes regulated by MS1, DYT1, and AMS			Fold change in *ms1*	Cluster
AT1G76470		NAD(P)-binding Rossmann-fold protein	-11.78	37
AT3G23770		O-Glycosyl hydrolases family 17 protein	-8.68	37
AT1G75930	EXL6	Extracellular lipase 6 (EXL6)	-42.58	73
AT5G07520	GRP18	Glycine-rich protein 18 (GRP18)	-26.94	73
AT5G07560	GRP20	Glycine-rich protein 20 (GRP20)	-118.80	73
AT3G52160	KCS15	3-ketoacyl-coa synthase 15 (KCS15)	-11.19	81
AT5G49070	KCS21	3-ketoacyl-coa synthase 21 (KCS21)	-3.32	81
AT1G71160	KCS7	3-ketoacyl-coa synthase 7 (KCS7)	-7.94	81
AT4G28395	ATA7	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin	-12.85	81
AT5G13380		Auxin-responsive GH3 family protein	-14.59	81
AT5G65205		NAD(P)-binding Rossmann-fold protein	-12.91	81
AT1G67990	TSM1	O-methyltransferase, family 3 protein	-82.62	81
AT5G16960		Zinc-binding dehydrogenase family	-20.37	81
AT1G13140	CYP86C3	Cytochrome CYP86C3	-21.37	81
AT2G19070	SHT	Spermidine hydroxycinnamoyl transferase	-14.67	81
AT1G06260		Cysteine proteinase	-50.31	206
AT1G75920		GDSL-like Lipase/Acylhydrolase	-32.48	206
AT1G06990		GDSL-like Lipase/Acylhydrolase	-7.66	206
AT1G75910	EXL4	Extracellular lipase 4 (EXL4)	-81.15	206
AT1G18280		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin	-6.49	254
AT4G08670		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin	-5.57	254
AT5G20710	BGAL7	Beta-galactosidase 7 (BGAL7)	-6.66	254
AT1G24400	LHT2	Lysine histidine transporter 2 (LHT2)	-4.66	254
AT4G14815		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin	-7.852	263
AT1G74540	CYP98A8	Cytochrome CYP98A8	-19.06	263
AT1G68875		Unknown protein	-133.44	263
AT5G45880		Pollen Ole e 1 allergen and extensin	-17.30	308
AT3G13400	sks13	SKU5 similar 13 (sks13)	-29.54	308
AT1G02790	PGA4	Polygalacturonase 4 (PGA4)	-36.03	308
AT5G59845		Gibberellin-regulated family protein	-7.105	406
AT5G07550	GRP19	Glycine-rich protein 19 (GRP19)	-167.21	406
AT5G44400		FAD-binding Berberine family protein	-27.952	448
AT1G74550	CYP98A9	Cytochrome CYP98A9	-9.911	448
AT2G03850	LEA	Late embryogenesis abundant family	-20.268	560

5.3.1.3 H3K27 trimethylation of MS1 downstream genes

The MS1 regulatory targets were also overlapped with microarray data of genes carrying H3K27me3 marks in *Arabidopsis thaliana* (Zhang et al., 2007). Interestingly, this type of histone modification has been found in many of the down-regulated targets, typically the member genes from cluster 73 (Figure 5.7), indicating that histone modifications have been serving as an important mechanism for regulation of pollen formation. As a transcription factor, MS1 contains a PHD finger motif, which is widely found being involved in the regulation of chromatin structure and dynamics (Wilson et al., 2001; Ito and Shinozaki, 2002; Yang et al., 2007). Yet it is still not sure whether MS1 is responsible for the trimethylation of H3K27 or acting as a 'reader' that recognises these marks. Diverse distribution of the H3K27me3 marks may imply the existence of other types of histone modification.

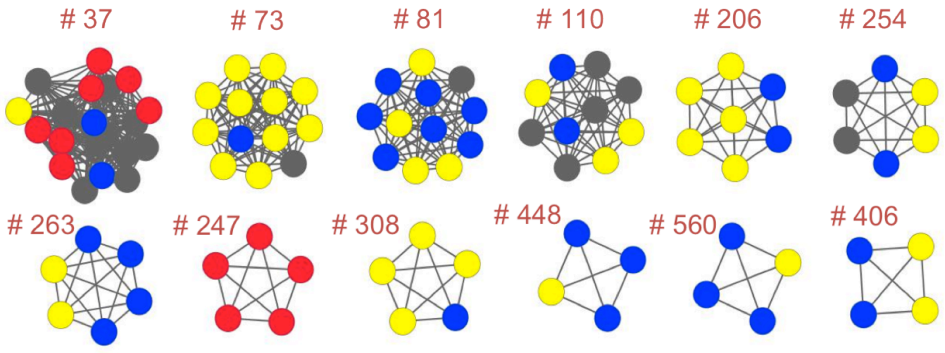


Figure 5.7 MS1 regulatory targets mapping with microarray data of H3K27me3 marks in the *Arabidopsis thaliana* (Zhang et al., 2007). Each node represents an individual gene belonging to the cluster, and edges indicate the correlation between the genes. Genes down-regulated in the *ms1* are indicated in blue, with the up-regulated being in red, and those simultaneously containing H3K27me3 marks are shown as yellow. Grey nodes represent genes with no expression change in the *ms1*.

5.3.2 Identification of MS1 Direct Target by ChIP

5.3.2.1 Identification of the plant materials

ChIP experiments were performed to investigate protein-DNA binding in the rescued *ms1ttg1* mutant, which expressed native MS1-GFP fusion proteins. Considering no antibody specific to the MS1 protein was available, a commercial anti-GFP antibody was used to enable the immunoprecipitaion. To ensure that the plants used were carrying the *MS1-GFP* transgene, expression analysis was performed by semi-quantitative PCR. Native expression of the *MS1-GFP* fusion was detected in the buds, suggesting the existence of the functional MS1-GFP fusion proteins in the rescued *ms1ttg1* mutant (Figure 5.8).

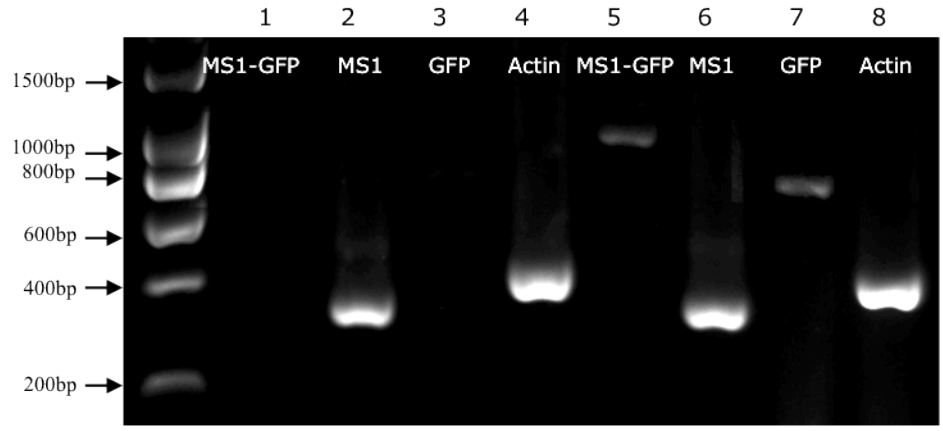


Figure 5.8 Semi-quantitative expression analysis of *MS1-GFP*. Lanes 1-4, wild type; lanes 5-8, rescued *ms1ttg1* line carrying the native MS1-GFP. In the rescued *ms1* mutant, expression of the *MS1* (lane 6), *GFP* (lane 7) and *MS1-GFP* fusion (lane 5) were detected in the whole buds, consistent with the native *MS1* expression in wt L*er* (lane 2); No *GFP* (lane 3) and *MS1-GFP* (lane 1) was detected in L*er* wild type buds; *Actin 7* was used as the control to check the integrity of the cDNA (lane 4 and 8). Primers are shown in Table 5.1.

5.3.2.2 Formaldehyde cross-linking and sonication

After formaldehyde cross-linking and fragmentation by sonication, the chromatin structure of each step was checked after extracted using phenol/chloroform. Cross-linking using 1% (v/v) formaldehyde for 10mins worked effectively to preserve protein-DNA complexes, leaving no free DNA after cross-linking, and no DNA recovery compromised after reverse-linked (Figure 5.9a). For general ChIP analysis, fragmentation was optimized to 200-1000bp by sonication with pulses of 4 times (Figure 5.9a). Regarding the samples used for subsequent ChIP-seq, in which smaller DNA fragments would be required, sonication was performed with pulses of 10 times, generating DNA fragments of 200-600bp (Figure 5.9b).

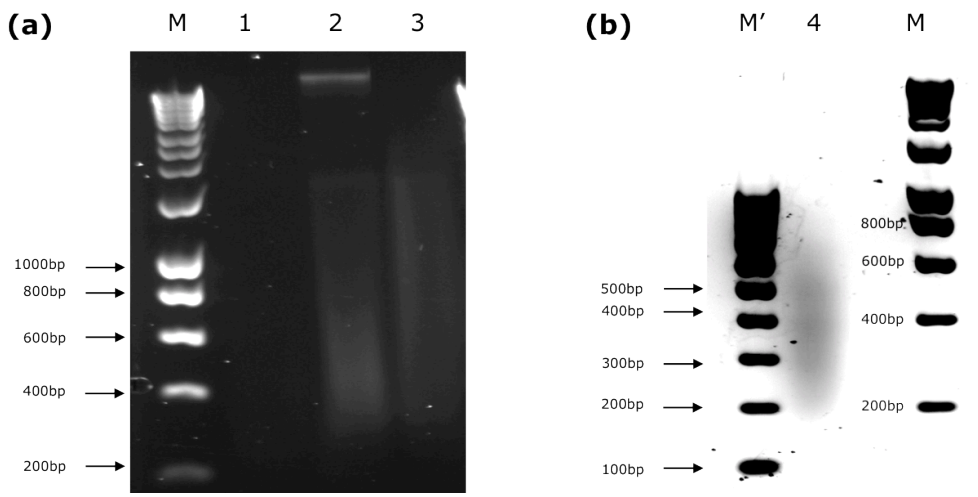


Figure 5.9 ChIP crosslinking and sonication efficiency. Fresh plant materials was cross-linked using 1% (v/v) formaldehyde in Extraction Buffer 1 for 10 min. Chromatin was extracted and fragmented by sonication with pulses of 10 sec at 3µm amplitude, 4 times for general ChIP analysis. Aliquots (10µl) of sheared and un-sheared chromatin DNA were taken and reverse cross-linked. Aliquots (10µl) of the un-sheared chromatin were also taken without reverse cross-linking to test whether cross-linking was sufficient. DNA was extracted from all samples and run on 4% (w/v) agarose gel. (a) Crosslinking and sonication efficiency for general ChIP experiments. M, marker HyperladderI (Bioline). Lane 1, DNA extracted without reverse cross-linking; cross-linking was sufficient to preserve chromatin structure, leaving no free DNA in the chromatin solution. Lane 2, DNA extracted after reverse cross-linking; DNA was efficiently recovered by reverse cross-linking, while small amounts of RNA were seen as well. Lane 3, extracted DNA were fragmented by sonication and purified after reverse cross-linking. The size of the fragmented DNA was predominantly between 200-1000 bp. (b) Sonication efficiency for samples used for ChIP-seq. M, marker HyperladderI (Bioline). M', marker Hyperladder IV (Bioline). Lane 4, extracted DNA were fragmented by sonication and purified after reverse cross-linking. The size of the fragmented DNA was predominantly between 200-600bp.

5.3.2.3 Analysis of chromatin precipitated samples by quantitative PCR

MYB transcription factor MYB99 was chosen to test promoter binding, which had been previously proposed as a direct target of MS1 (Alves-Ferreira et al., 2007; Yang et al., 2007) (Ito et al., 2007). 13 pairs of

primers covering the whole of a 1.8 Kb region upstream of *MYB 99* and 7 primers covering the gene body were designed (Figure 5.3a, Table 5.2), each of which covered a region of 100-200bp. Protein-DNA complexes were precipitated using anti-GFP antibody at two different concentrations. To eliminate the backgrounds, the non-specific antibody polyclonal anti-HA (Sigma-Aldrich) was added to the 'no-antibody' (NoAb) control sample, which had been believed to achieve better results than when no antibody was added.

The ChIPped DNAs were then used as templates to determine the enrichment of the target sequences. Among all the primer pairs, DNA fragments covered by P12, the promoter region adjacent to 5'UTR of *MYB99*, showed enrichment in the quantitative PCR analysis (data not shown). To confirm this, an additional pair of primer P12a covering overlapping but narrower DNA regions (Figure 5.3) was designed to give more precise results. Correspondingly, enrichment was seen in the P12a promoter region in both immunoprecipitates (Figure 5.10 a), demonstrating that MS1 bound to the promoter of *MYB99* at the region adjacent to 5'UTR. ChIP experiments using L*er* wild type were performed in parallel following the same conditions, serving as the negative controls. No significant enrichment of the IP samples were seen, compared with the 'no antibody' control (Figure 5.10 b).

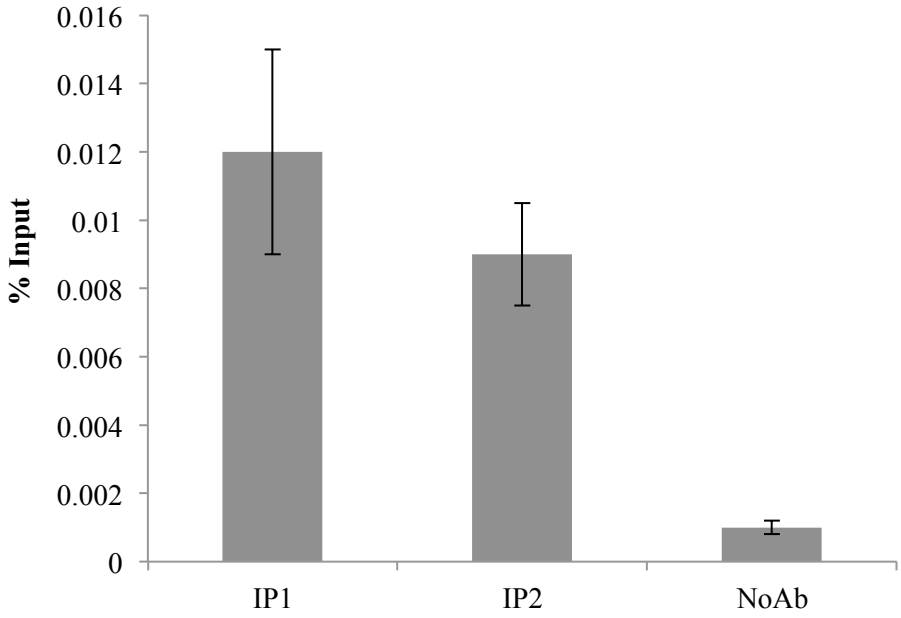


Figure 5.10 qChIP-PCR of ChIP experiment using primer pair P12a on *MYB99* (Figure 5.3, Table 5.2). Enrichments of MYB99 were determined using 13 pairs of primers covering 1.8 Kb upstream region upstream and 7 pair of primers covering the gene body (Figure 5.3a, Table 5.2), among which P12a were showing the highest enrichments. Protein-DNA complexes were precipitated using anti-GFP antibody (IP1 and IP2) at two different concentrations and were compared with samples precipitated with non-specific antibody (NoAb). qPCR reactions were run with four technical replicates, whose mean values were used for data normalization. Normalization was processed with the '% of input' method, dividing the qPCR signals derived from the IP samples by the qPCR signals derived from the input sample. Enrichment was seen only in the P12a promoter region using *ms1ttg1* rescued by native MS1-GFP. IP1, IP2, independent immunoprecipitations using the same plant materials, but antibody of different concentrations were used.

5.3.2.4 Preparation of the ChIP DNA libraries

ChIP samples were sent to SourceBioscience for preparation of the DNA library for sequencing. The 'input' samples as well as two IP samples successfully generated libraries, yet the NoAb sample failed to create a

library, which may be due to insufficient DNA material. This had been consistent with what was expected since no chromatin was expected to be precipitated in the NoAb sample, due to absence of antibody specific to the binding protein. Concentration of the libraries was validated by Qubit High Sensitivity & Broad Range Assays (Table 5.5), and distribution of the DNA sizing was assayed by Agilent BioAnalyzer High Sensitivity Chip (Figure 5.11).

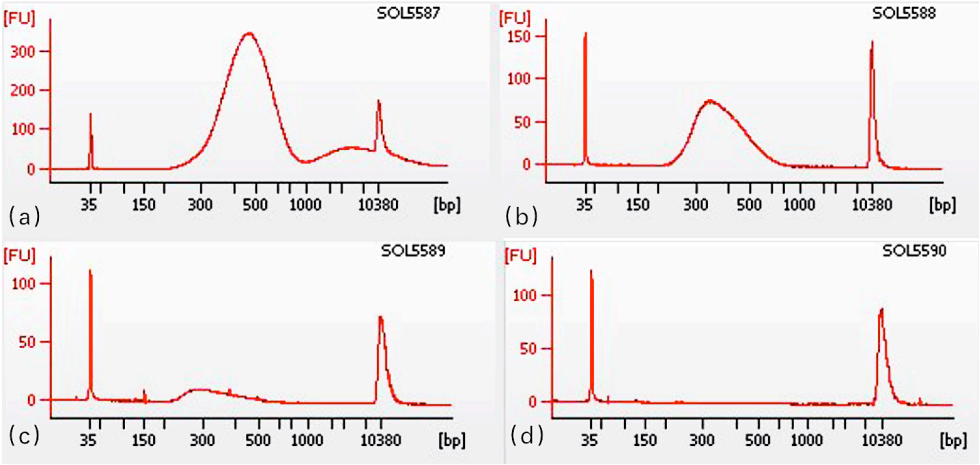


Figure 5.11 ChIP DNA libraries validated by Agilent BioAnalyzer High Sensitivity Chip. X-axis indicates distribution of the DNA sizing, and Y-axis represents concentrations. Libraries were prepared by rounds of amplification. Main peaks in the graphs represent the distribution of DNA sizing after sonication, which should not exceed 1kb. (a) library of 'Input' sample; (b) library of IP1 sample; (c) library of IP2 sample; (d) library of 'NoAb' sample.

Table 5.5 ChIP DNA Libraries for Sequencing Validated by QUIBIT.

External ID	Sample Name	QUBIT ng/ul	Average bp	nM
SOL5587	Ms1_GFP_Input	69.3	467	222.6
SOL5588	Ms1_GFP_IP1	15.1	387	58.5
SOL5589	Ms1_GFP_IP2	2.37	347	10.2
SOL5590	Ms1_GFP_NoAb	0.125	0	

5.3.2.5 Determination of MS1 direct target identified by comparative transcriptome analysis

Previously, putative MS1 directly regulatory targets have been identified using transcriptome analysis (Table 5.4). Gene members of cluster 81 have been found closely correlated with MYB99 in both temporal expression in the wild type and regulatory patterns in the *ms1* mutant (Alves-Ferreira et al., 2007). These genes were then selected for ChIP validation, using both MYB99 active regions (P12a) and inactive regions (P15) as the controls (described in Figure 5.3). Primers were designed at the equivalent position to the region enriched on MYB99 promoters (Table 5.6; Figure 5.12a). Data normalization was adopted from Haring et al. (Haring et al., 2007) using the 'Fold change' method from by dividing the qPCR signals derived from the IP samples by the qPCR signals derived from the NoAb sample to correct for backgrounds.

Table 5.6 Primers Used for qChIP-PCR Analysis of MS1 Regulatory Targets.

Gene	Primer	Sequence 5'-3'
KCS7	KCS7_PRO_F	TGGAGTTGGTGAGATAGAGGT
	KCS7_PRO_R	GCATGTGAAAGGTGTCCAAAAC
ATA7	ATA7_PRO_F	ACAATGAAACAAACCGATACGT
	ATA7_PRO_R	TGAGCCGATATCATTGCCTA
CYP86C3	CYP86C3_PRO_F	AGCCATGAACATTAACGTTTCG
	CYP86C3_PRO_R	AGAACTCCACTTTTGTTGCA
SHT	SHT _PRO_F	AGAGAAGAAGGGTAATACGC
	SHT _PRO_R	CGTGGTTTAGTTTGTTTCCTTCG
TSM1	TSM1_PRO_F	TGGTCATATCATTTGGGTGGA
	TSM1_PRO_R	TTTCTGGTCGACACATTATTATTAC
KCS15	KCS15_PRO_F	CGGAGATGATGGATTTCAGTTGT
	KCS15_PRO_R	ACCAATGCTCGGTTCAGGTA
KCS21	KCS21_PRO_F	CAGTGGCTGGTTTAACGTGT
	KCS21_PRO_R	CAGAGGATTTCACTGCGTTAGG

MYB99 promoter showing 4.2-fold enrichment in the P12a region served as the positive control, with P15 region exhibiting no enrichment used as the negative one. For the other genes tested, KCS7, KCS15, SHT and TSM1 showed enrichment in the ChIP analysis (Figure 5.12b), and are proposed as direct regulatory targets of MS1. In contrast, enrichments failed to occur on the sequences tested derived from ATA7, CYP86C3 and KC721. However, this cannot exclude the possibility of unchecked promoter regions acting as the binding sites for MS1.

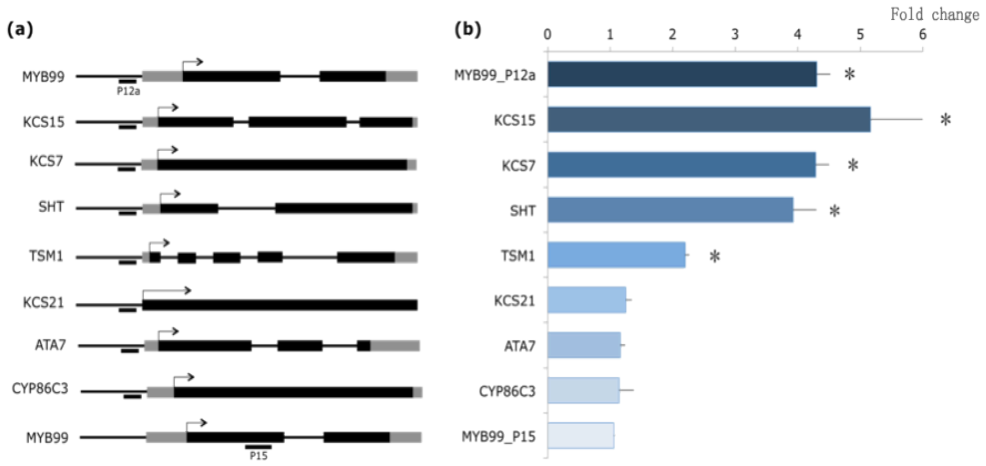


Figure 5.12 qChIP-PCR analysis of the enrichment of MS1 regulatory targets and the predicted regions tested in their promoter regions. (a) Diagram representing genomic structure and the tested regions of various genes identified as downregulated in *ms1* anthers. Black boxes represent exons; grey boxes indicate UTRs; arrows indicate the ATG start codon sites; bars represent regions amplified by qChIP-PCR analysis. (b) Fold enrichment calculations from qPCR assays in two independent ChIP experiments. qPCR reactions were run with four technical replicates, whose mean values were used for data normalization. Fold enrichment data were analysed to calculate the fold change between the IP samples and no-antibody control. Lines represent technical replicate variations, and asterisks represent statistically significant fold changes compared with the MY99_P15 negative control, indicated by *t*-test.

5.4 DISCUSSION

Normal male reproductive development in *Arabidopsis thaliana* requires proper development of the microspores and the surrounding tapetum, under the regulation by a number of transcription factors. The *ms1* mutant is defective in tapetum development giving rise to no viable pollen grains produced in the anther locule (Wilson et al., 2001; Ito et al., 2007). By studying the transcriptome of anthers prior to mitosis I stage in the *ms1* mutant, it provides us a system to investigate pathways directly regulated by MS1. In this chapter, comparative transcriptome analysis of anthers from *ms1* and wild type were performed based on the FlowerNet correlation network generated by Transclust analysis. By focusing on transcriptome of anthers prior to mitosis I stage, noise from later stages have been excluded, which shows that MS1 directly regulates normal expression of lipid biosynthesis and transport, and other pollen exine formation pathways (Table 5.3).

For young anthers before mitosis I stage, pathways involving lipid metabolism and transport (cluster 73, 254 and 263) and pollen exine formation (cluster 37, 81 and 263) show change in expression levels due to the *ms1* mutation. Other pathways include a group of endomembrane proteins (cluster 206) related to sexual reproduction, with detailed biological functions as yet undefined. Those regulated in both the young and old

anthers (cluster 73 and 110) may indicate multiple biological functions of the gene members, or else as the result of imprecision in the staging method based on anther morphology. For instance, cluster 73 is entirely down-regulated in the young anthers of *ms1*, yet partially in the old ones. Conversely, cluster 110 engaged in actin filament organisation pathways is mainly regulated in the old anthers of *ms1*. In wild type anthers, gene members of cluster 110 show prolonged expression until mature pollen grain stage, which peaks at the mitosis II stage.

MS1 belongs to a PHD finger transcription factor, a motif that is found in a wide variety of eukaryotic proteins involved in the regulation of chromatin structure and dynamics (Wilson et al., 2001; Ito and Shinozaki, 2002; Yang et al., 2007). Interestingly, it has been discovered that many of the genes down-regulated in the *ms1* mutants are carrying trimethyl-H3K27 marks in wild type plants, typically the member genes from cluster 73 (Figure 5.7), which indicates that modifications of chromatin structure are critical mechanisms for regulating pollen formation. However, as a PHD finger transcription factor, it is still unknown whether MS1 acts as a 'writer' responsible for this modification, or the 'reader' that recognises this marks to facilitate transcription. Previously, MYB99 has been identified as a MS1 direct target, which has been confirmed by our chromatin immunoprecipitation (ChIP) analysis. The MS1 protein binds to the promoter region of MYB99, adjacent to the 5'UTR, but it's still unknown whether this association is through physically connection or with the assistance of other proteins like histone tails. The transcriptional regulation

process may involve multiple chromatin mechanisms, with the assistance of various TFs. Correspondingly, MYB99 has also been found to be regulated by ASHH2 via H3K36 trimethylation, a SET domain transcription factor, whose homolog ASHR3 is an interaction partner of the AMS protein (Grini et al., 2009; Ma et al., 2012). ASSH2 is assumed to participate in genetic networks operating earlier than, or in parallel to MS1.

Certain putative direct targets of MS1 that are identified by comparative analysis show closely correlated with MYB99 in both temporal expression patterns and regulatory behaviors (Alves-Ferreira et al., 2007). These selective candidates were examined by ChIP analysis, using both MYB99 active and inactive regions as the controls. Primers were designed at the equivalent position to the region enriched on MYB99 promoters. Noticeably, KCS7, KCS15, SHT and TSM1 show enrichments in the ChIP analysis, proposed as direct regulatory targets of MS1. Sequences from ATA7, CYP86C3 and KC721 failed to be enriched by MS1, but cannot be excluded as being MS1 putative direct targets considering that only selected regions were examined. Previously, KCS7 and KCS15 were also identified as direct targets of AMS, whose promoter regions were enriched by AMS in the ChIP analysis (Xu et al., 2010), indicating that MS1 and AMS may collaborate in promoting the transcription of the two genes in a certain way.

Previous studies on the regulatory networks of pollen development suggest that DYT1 regulates the normal expression of MS1 directly by binding to its promoter region, or else via the DYT1-AMS-MS1 forward loops (Xu et al.,

2010; Feng et al., 2012; Ma et al., 2012). It is also revealed that the DYT1 and AMS proteins, as bHLH transcription factors, potentially interact with each other to form heterodimers (Xu et al., 2010; Feng et al., 2012; Ma et al., 2012). As shown in Figure 5.6, some genes are common downstream targets of DYT1, AMS and MS1, for instance members of cluster 81. Though some of these candidates were identified as direct targets of AMS by the ChIP analysis, including EXL4, EXL6, KCS7, KCS15, GRP18, GRP19, CYP98A8, CYP98A9, and At1g06990 (GDSL-like Lipase) (Xu et al., 2010), they are severely down-regulated in the *ms1* mutant indicating that AMS alone is sufficient to activate of the transcription of these genes, with MS1 or the MS1 downstream regulators still being required.

A proportion of MS1 downstream genes are regulated by AMS independently of DYT1. In other words, expression of these genes can be activated in the absence of DYT1, which may indicate the existence of intermediate regulators that act as alternative activators apart from MS1, yet repressed by DYT1 (Figure 5.13). Interestingly, unlike most of the genes defining pollen exine formation, increased expression of some previously characterised genes controlling sporopollenin pathways are identified, including ACOS5, KNS2, LAP5, DRL1, CYP704B1 and CYP703A2 (Chapter 1), all of which are downstream genes of DYT1 and AMS (Xu et al., 2010; Feng et al., 2012).

In summary, this study highlights the function of MS1 as a key transcriptional regulator for normal pollen formation. A proposed regulatory

network exhibiting the correlation of DYT1, AMS and MS1 are shown as Figure 5.13. Further investigations such as sequencing of the ChIP library generated from the ChIP analysis will help to gain insights into the MS1 regulatory mechanism in controlling anther development and pollen formation.

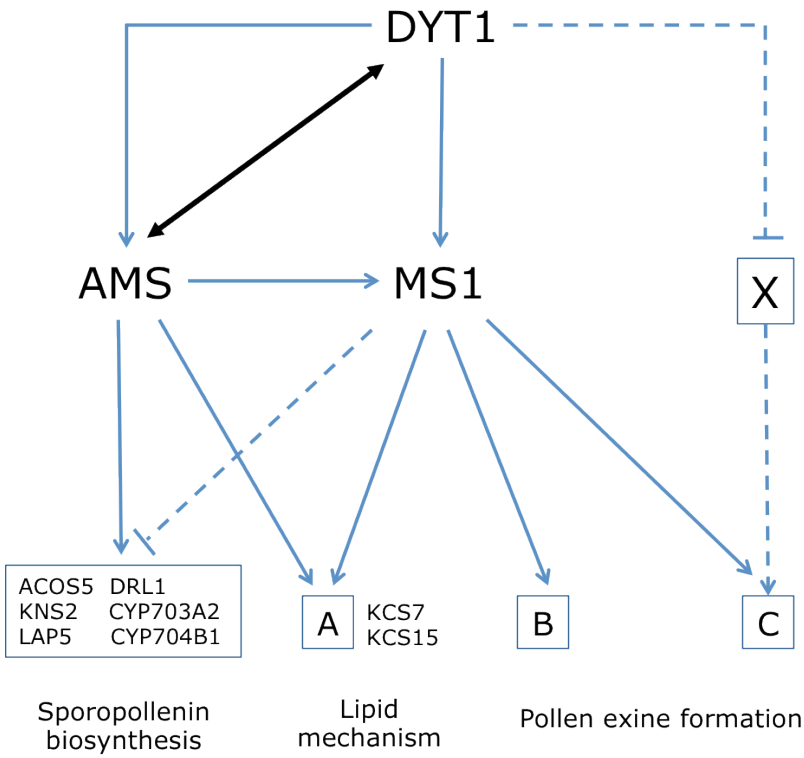


Figure 5.13 Model of gene regulatory network controlled by MS1. DYT1, AMS and other genes reported to be essential for anther development are shown. Arrows represent positive regulations and T-bars indicate negative regulation. The double arrow represents a physical interaction. Regulations confirmed by experiments are shown in bold line. Group A represents putative direct targets of both MS1 and AMS. Group B indicates other common downstream genes of DYT1, AMS and MS1. Group C indicates those regulated by AMS independently of DYT1. X, repressed by DYT1, represents intermediate protein that activate group C proteins.

CHAPTER 6 MS1 REGULATES POLLEN FORMATION PATHWAYS VIA HISTONE MODIFICATIONS

6.1 INTRODUCTION

The MS1 protein contains a PHD-finger motif that is found in a wide variety of eukaryotic proteins involved in the regulation of chromatin structure and dynamics (Figure 6.1a) (Wilson et al., 2001; Ito and Shinozaki, 2002; Yang et al., 2007). The PHD finger is a small protein domain of 50–80 amino acid containing a zinc-binding motif (Aasland et al., 1995). Common features of the PHD finger consist of the two-strand anti-parallel β-sheet and a C terminal α-helix (not present in all PHDs), which are stabilized by the Zinc-coordinating residues (Figure 6.1b) (Sanchez and Zhou, 2011). This structure is critical for chromatin association to allow the PHD fingers to read the N terminal tail of histone H3 (Sanchez and Zhou, 2011).

This finger motif was originally discovered in the *Arabidopsis thaliana* protein HAT3.1, and has been found from a range of organisms from human to yeast (Schindler et al., 1993; Gong et al., 2005; Shi et al., 2007). Previous studies in humans have proposed that the large family of PHD fingers contained proteins are involved in facilitating DNA transcription, repair, recombination and replication by binding to mainly the methylation state of H3K4 (K4me0 and K4me3/2) (Li et al., 2006; Taverna et al., 2006; Ramón-Maiques et al., 2007; van Ingen et al., 2008; Hung et al., 2009; Wang et al., 2009; Wen et al., 2010), and to a lesser degree the methylation

state of H3R2 (R2me0 and R2me2) (Lan et al., 2007; Chignola et al., 2008; Chakravarty et al., 2009; Chignola et al., 2009) and the acetylation state of H3K14 (Zeng et al., 2010). In addition, some PHD fingers proteins in the yeast shows the preference to histone H3 trimethylated at Lys36 (H3K36me3) (Shi et al., 2007; Musselman and Kutateladze, 2009) .

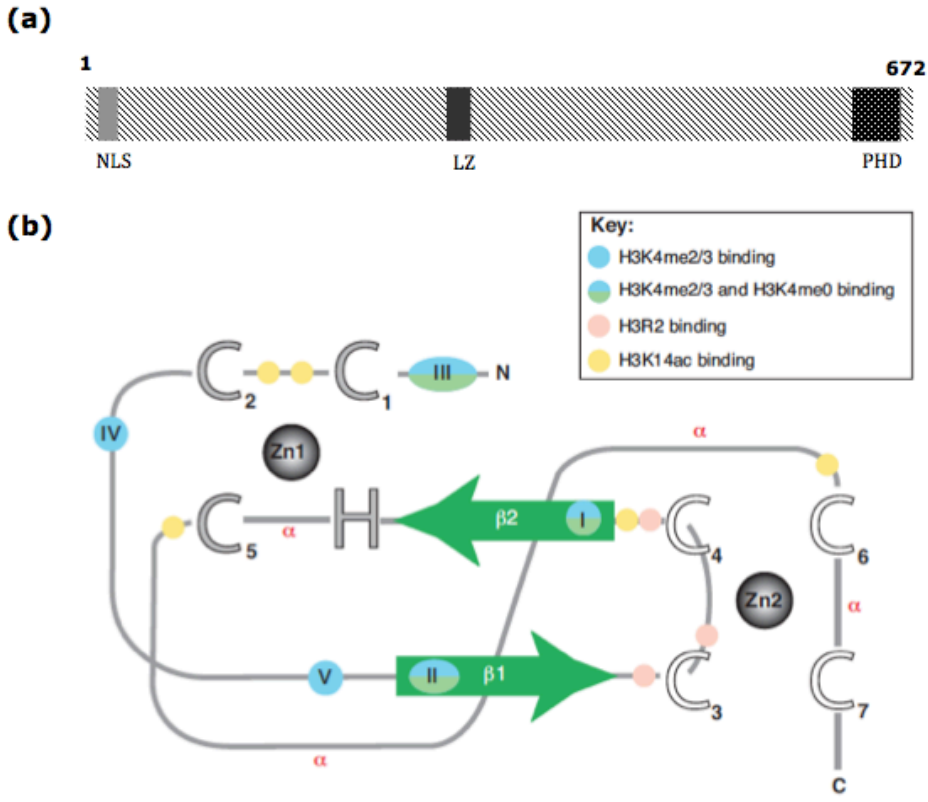


Figure 6.1 Structures of MS1 protein and PHD finger motif. (a) Schematic representation of the MS1 protein (Ito and Shinozaki, 2002). NLS, LZ, and PHD shown by dark gray indicate nuclear localization signal, Leu zipper–like region, and PHD region, respectively. (b) Schematic of the PHD fold (Sanchez and Zhou, 2011). All PHD fingers adopt the same basic topology. Zinc atoms are shown as gray spheres, with Zinc-coordinating residues indicated as large uppercase letters. The two coreβ-strands are shown in green. Regions that adopt anα-helical conformation in some PHD fingers are indicated (α).

Arabidopsis thaliana HAT3.1 was the first characterised PHD-finger protein in plants, which was demonstrated to be capable of interacting with any DNA fragment larger than 100 bp. While a deletion of the N-terminal PHD-finger domain gave rise to completely abolished DNA binding. This suggested that the plant PHD finger motif might play an important functional role in protein—protein or protein—DNA interaction (Schindler et al., 1993). Histone modification can play a central role in regulating development (Aasland et al., 1995). In *Arabidopsis thaliana*, the vernalization mechanism through epigenetic repression of FLOWERING LOCUS C (FLC) expression involves three PHD-finger proteins VERNALIZATION 2 (VRN2), VERNALIZATION INSENSITIVE 3 (VIN3), and VERNALIZATION 5 (VRN5) (Gendall et al., 2001; Bastow et al., 2004; Wood et al., 2006; Greb et al., 2007). VRN5 and VIN3 form a heterodimer to establish the vernalization-induced trimethylation of H3K27 and histone deacetylation, which is required for the epigenetic silencing of FLC (Gendall et al., 2001; Bastow et al., 2004; Wood et al., 2006; Greb et al., 2007). VERNALIZATION 2 (VRN2) is essential for maintenance of this modification (Gendall et al., 2001). Another example is the *Arabidopsis thaliana* SET Domain Group (SDG4) gene that contains a PHD-finger motif required for the maintenance of methylated histone H3K4 and K36 levels in the mature pollen grain (Cartagena et al., 2008).

Previous research on the *MS1* gene strongly supports that MS1 is a transcription factor that plays a key role in supporting pollen development by regulating late tapetal gene expression. Large numbers of genes are

down-regulated in the *ms1* mutant, and many transcription factors were identified as downstream of MS1 (Alves-Ferreira et al., 2007; Ito et al., 2007; Yang et al., 2007). The proposed MS1 target MYB99 (Ito et al., 2007; Yang et al., 2007) was confirmed as a MS1 direct target by Chromatin Immnoprecipitation (ChIP) analysis (Chapter 5). It has been found to carry the H3K27me3 and H3K36me3 marks (Zhang et al., 2007; Grini et al., 2009). In addition, levels of H3K36me3 marks on MYB99 are reduced in the *ashh2* mutant, suggesting that MYB99 is regulated by ASHH2 via H3K36me3, considered as a mechanism in parallel to MS1 regulation (Grini et al., 2009; Grini et al., 2009).

By characterisation of the histone marks on *MYB99* and investigation on the MS1-mediated histone modification will give us valuable insights into the molecular mechanism of the regulation role of MS1.

6.2 MATERIALS AND METHODS

6.2.1 Plant Materials

Seeds of *Arabidopsis thaliana* Landsberg *erecta* (L*er*) and *ms1ttg* mutant (L*er* background, NASC ID N1298) were obtained from the Nottingham *Arabidopsis* Stock Center (NASC). Overexpression lines (L*er* background) of *Y2H54* under the control of CaMV35S promoter from a Gateway destination vector pGWB14 (Appendix VII) (Nakagawa et al., 2007) were generated and maintained in our lab by Dr. Caiyun-Yang (University of Nottingham).

Seeds were sterilised (Section 2.2) and screened on half MS medium (Section 2.10.2), with Kanamcyin added for MS1 overexpression line screening. Plates were placed in a cold room for 3 days at 4°C to synchronize germination, and then cultured under full light (140μmol·m^{-2}·sec^{-1}) at 22-24°C. After 10 days, surviving seedlings were transformed into Levington M3: John Innes No.3: vermiculite: perlite (6:6:1:1) compost mix supplemented with 2% (w/v) Intercept® (Scotts) and placed in the glasshouse (Section 2.1). Rosette leaves from overexpression lines were harvested at 20 days after sowing, whole inflorescences of the *ms1ttg* mutant collected at 28 days after sowing.

6.2.2 Histone Extraction and Western Blotting Assay

Arabidopsis thaliana histones were extracted from 3-week-old rosette leaves of MS1 overexpression lines and 4-week-old buds of *ms1* mutants, with tissues from wild type L*er* used as controls. 10g of materials collected from each line were respectively ground in liquid N_2 into fine powder. Nuclear precipitation was performed as previously described in the chromatin immunoprecipitation analysis (Section 5.2.2). The pelleted nuclei were then treated twice for 45mins with 0.4M H_2SO_4 and centrifuged at 12,000g for 10mins after each treatment, supernatant transferred to a new tube. The proteins retained in the supernatant were precipitated with 10 volumes of pre-cold acetone overnight at −20°C, with 0.2% (w/v) DTT and 10% (v/v) TCA added. After centrifugation at 12,000g for 10mins, the pellet was washed with pre-cold acetone twice, dried and resuspended in 0.01 N HCl. Quantification of extracted histones was carried out by NanoDrop 2000 spectrophotometer (Thermo) at OD280. Equal amount of each sample was separated by electrophoresis on SDS-PAGE gel (Section 2.19).

The covalent modification status of H3 tails was subsequently analysed by western blotting following the instructions previously described (Section 2.16), using specific antibodies (Table 6.1) diluted to a concentration of 1 μg/ml.

Table 6.1 Antibodies Used for Western Blotting Assay.

Antibody	Target	Manufacture	Catalog No.
anti-trimethyl-H3K4	H3K4me3	Abcam	ab8580
anti-trimethyl-H3K27	H3K27me3	Millipore	Upstate 07-449
anti-trimethyl-H3K36	H3K36me3	Abcam	ab9050
anti-H3	H3 core	Abcam	ab1791

6.2.3 Chromatin Immunoprecipitation Analysis

To determine whether histone modifications were altered by *ms1* mutation, chromatin immunoprecipitation (ChIP) analysis was performed in the *ms1ttg* mutant compared with wild type L*er*, using the specific histone antibodies that were used for the western blotting (Table 6.1). The ChIP experiments were established as previously described (Section 5.2), using 5μg or 10μg of each antibody in the two independent IP samples, respectively.

Quantitative PCR analysis (qChIP-PCR) was performed on selective genes with the FLC gene serving as a positive control (Table 6.2). Data normalization was adopted from Haring et al. (Haring et al., 2007). The '% of input' method was used to determine the enrichments of FLC, with the qPCR signals derived from the IP samples divided by the qPCR signals derived from the input sample to correct for technical variations. The 'relative to control sequences' method was applied to normalisation of other

genes, the qPCR signal derived from the genes examined divided by the signal derived from the FLC positive control

Table 6.2 Primers Used for ChIP Analysis.

Gene	Primer	Sequence 5'-3'
FLC	FLC F	GCTTGTGGGATCAAATGTCAA
	FLC R	TAGTCACGGAGAGGGCAGTC
MYB99	MYB99 F	CACCGTATTCAATGGTTTTAGCA
	MYB99 R	ACACACGTATGGAGTTTCTTGG
CYP703A2	CYP703A2 F	TCCCTCTTCGCTGTTCTCAT
	CYP703A2 R	TAGGCAATCTTGGTGGACCT
CYP704B1	CYP704B1 F	AAAAATGTCGTTGTGTTTGGTT
	CYP704B1 R	CTCAACGAGCCAATCATGC
ACOS5	ACOS5 F	TCATTTCCGGTTTCGGTTTA
	ACOS5 R	AGCCTCTTTGTGTCCCTCAC
LAP5	LAP5 F	GAGTTTTGCTTGCCACGTCT
	LAP5 R	ATAGCTCCGGCTCCATCAC
DRL1	DRL1 F	TTGCACGAACCCACATAGTC
	DRL1 R	CAAACCTCTTGGGGATAGGG

6.3 RESULTS

6.3.1 Conserved Residues at the MS1 PHD Finger

PHD finger motif is a common structural motif that is widely found in eukaryotic genomes. Alignment of the sequences from MS1 with those from other PHD fingers revealed that MS1 include conserved residues predicted critical for binding to the di/tri-methylation state of H3 tails (Figure 6.2). This indicates that MS1 potentially binds to di/tri-methylated H3 tails, however, exact substrates for MS1 PHD fingers cannot be confirmed until further experiments are conducted.

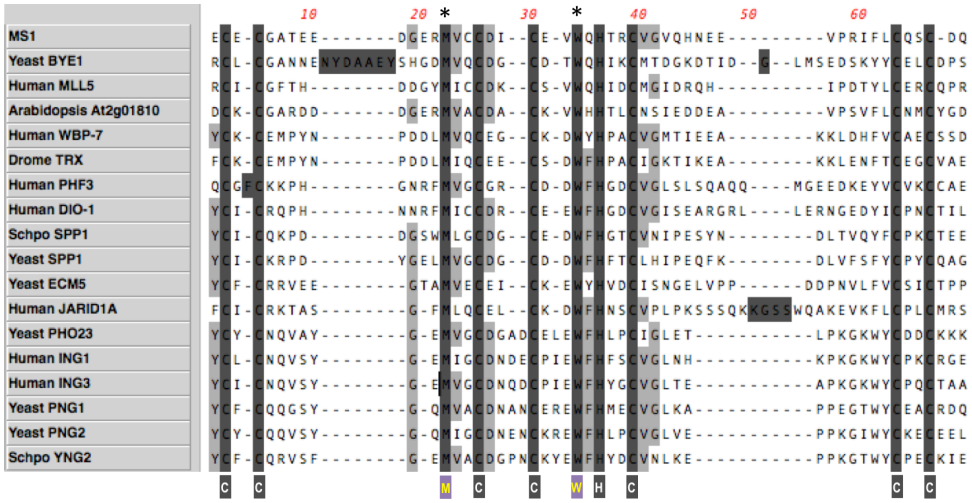


Figure 6.2 Alignment of amino acid sequences between PHD-finger motifs, including MS1 protein. Zinc-coordinating residues (grey) and conserved residues predicted critical for histone binding (purple) are shown.

6.3.2 Global Levels of Trimethyl-H3 Altered by Misoverexpression of MS1

To obtain insight into the molecular mechanism of MS1-mediated regulation of gene expression, trimethyl-histone 3 levels were examined in the MS1 overexpression plants and the *ms1ttg* mutant, compared with those in the wild type L*er*. Western blotting analysis (Figure 6.3) showed that the overexpression of MS1 caused dramatically reduced levels of H3K4me3, H3K36me3 and H3K27me3, in the rosette leaves where endogenous MS1 is not expressed. Correspondingly, levels of H3K4me3 and H3K27me3 were significantly enhanced in the buds of *ms1ttg* mutant, compared with wild type L*er*. However, it was difficult to determine the change in levels of H3K36me3, which had been barely detected in neither the buds of the *ms1ttg* mutant or those of wild type L*er*.

The western blotting analysis indicated that MS1 is able to facilitate the detrimethylation of H3K4, H3K27 and H3K36 marks. It is not surprising that the histone lysine demethylation reactions are nonspecific, since previous researches show that one kind of demethylase can demethylate many methylated histone states (Tan et al., 2008). However, it has not been clear whether MS1 itself acted as a demethylase to erase the marks or as a linkage that anchors the marked histone to enable other proteins to assemble to them.

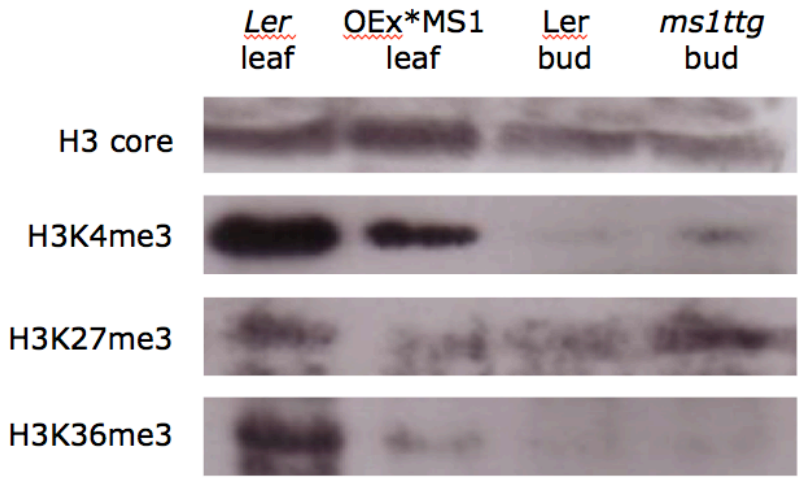


Figure 6.3 Comparison of Histone Methylation in MS1 overexpression lines and *ms1ttg* mutants with wild type L*er*. Global levels of H3K4, H3K27 and H3K36 tri-methylation in the leaf tissue are curtailed by overexpression of MS1. In the *ms1ttg* mutant enhanced levels of H3K4me3 and H3K27me3 are detected in the buds, compared with wt L*er*. Histone-enriched protein extracts were analyzed by protein immunoblots using specific antibodies (Table 6.1) that recognize different histone methylation forms as indicated.

6.3.3 *MYB99* Show Enhanced H3K36 Trimethylation in *ms1ttg*

To figure out more precisely the molecular mechanisms of MS1 regulation, trimethyl-H3 was investigated at specific genes by ChIP assays in the inflorescences of *ms1ttg*, compared with wild type L*er*. ChIP analysis was performed using the *FLOWERING LOCUS C (FLC)* gene as a positive control, which simultaneously carries H3K4me3, H3K27me3 and H3K36me3 marks on its chromatin structure (Xu et al., 2008). As an earlier transcription factor that regulates the induction of flowering by vernalisation, the chromatin structure of FLC should not be altered by the *ms1* mutation (Michaels and Amasino, 1999; Sheldon et al., 2000).

ChIP analysis revealed that relative high levels of H3K4me3, H3K27me3 and H3K36me3 at *MYB99* were all detected in the wild type L*er*. It was also shown that level of H3K36me3 was increased on *MYB99* in the *ms1ttg,* with the remaining two being reduced (Figure 6.4). Previous research suggests that tri-methylation of histone H3 on lysine 36 is repressive when found in the promoter region (Pfluger and Wagner, 2007); this is the region where the enrichment of *MYB99* by MS1 occurred (see Section 5.3.2). It appears that loss of the MS1 function may maintain the repressive H3K36me3 marks on *MYB99*, with chromatin structure of *MYB99* staying inactivated. Taken together, this ChIP data suggests that MS1 mediates demethylation of H3K36me3 selectively at *MYB99.*

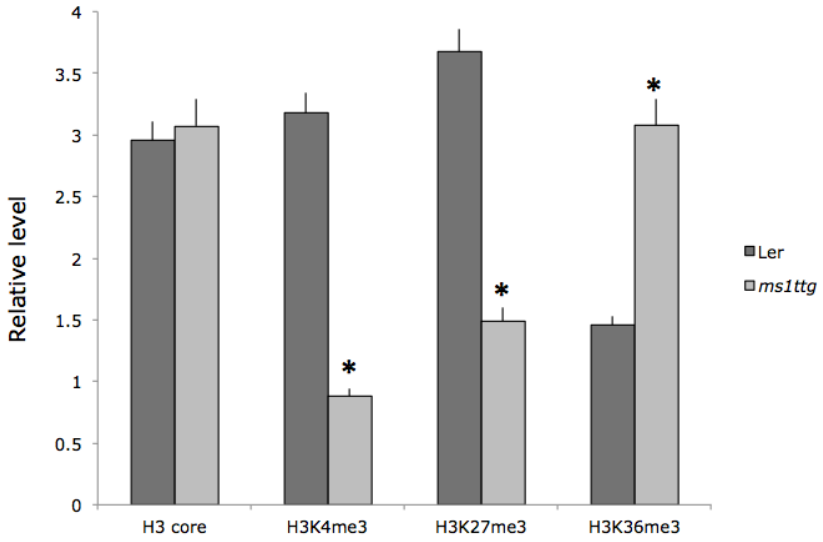


Figure 6.4 ChIP analysis of H3K4me3, H3K27me3 and H3K36me3 at *MYB99* in L*er* and *ms1ttg* in inflorescences. Samples of H3 core were used as the control to make sure equal amount of input histones. ChIP samples were analyzed by quantitative PCR on each gene. Relative levels are calculated from mean values of four replicates; error bars show SD. The asterisk indicates a significant difference between *ms1ttg* and L*er* ($P < 0.01$).

6.3.4 Altered Histone Modification in *ms1* on Genes Involving Sporopollenin Biosynthesis

Trimethylation of H3 marks were examined specifically on genes representing sporopollenin biosynthesis, including *CYP703A2, CYP704B1, ACOS5, DRL1* and *LAP5* (Figure 6.5). In wild type plants, relative high levels of the H3K4me3, H3K27me3 and H3K36me3 marks are widely found (Zhang et al., 2007). Some genes are carrying bivalent marks of both active H3K4me3 and repressive H3K27me3 marks (*CY703A2, ACOS5* and *DRL1*), whose transcription activations require complex regulating mechanisms by recruitment of methylases and demethylases, to resolve bivalent domain into monovalent domain (Lan et al., 2008).

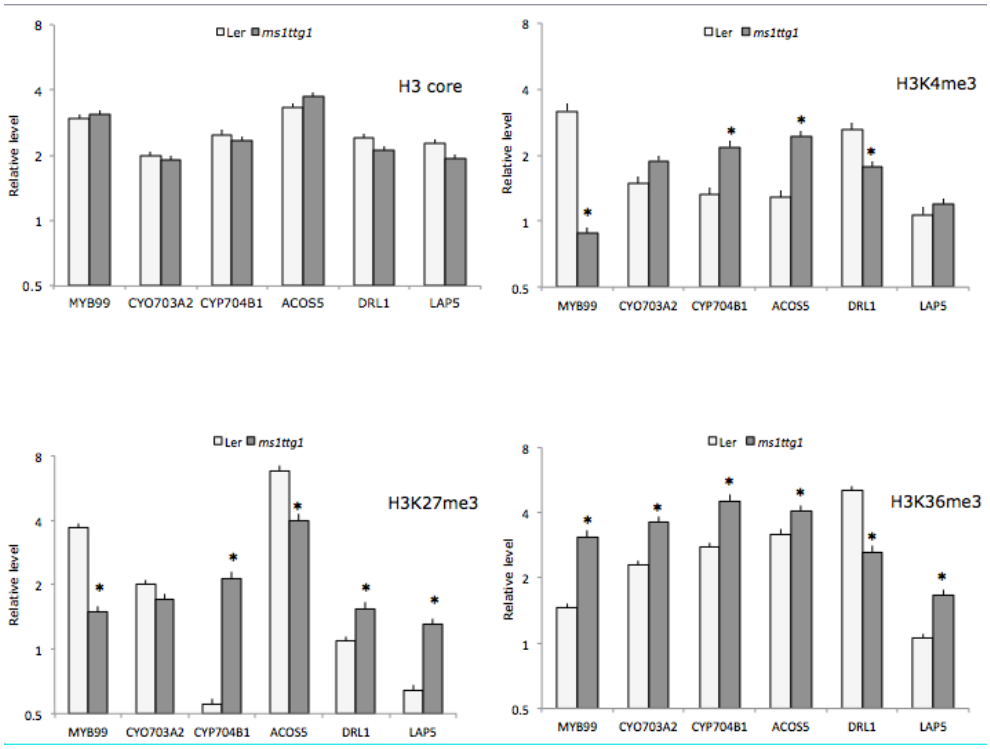


Figure 6.5 ChIP analysis of H3K4me3, H3K27me3 and H3K36me3 at specific genes in in inflorescences from L*er* (white columns) and *ms1ttg* (gray columns). Samples of H3 core were used as the control to ensure equal amount of input histones. ChIP samples were analyzed by quantitative PCR on each gene. Relative levels are calculated from mean values of four replicates; error bars show SD. The asterisk indicates a significant difference between *ms1ttg* and L*er* ($P < 0.01$).

In the *ms1* mutant, levels of trimethylation at histone Lys were dramatically changed (Figure 6.5). Almost all levels of the trimethyl-H3 marks were altered in *ms1ttg*, except for the H3K4me3 at *LAP5* and H3K27me3 at *CYP703A2*. A typical example is the H3K27me3 mark on *CYP704B1*, which was barely detected in wild type, but considerably increased in the *ms1ttg*. In theory, increasing the repressive H3K27me3 marks should cause a reduction in gene expression level, but conversely, *CYP704B1* is

up-regulated in *ms1* as indicated by microarray analysis (see Section 5.3.1) (Yang et al., 2007), probably as the consequence of increased levels of active H3K4me3 or H3K36me3 marks.

This demonstrates that the final regulation effects on a specific gene depend on the overprint change of various histone marks. In summary, the results indicate that histone methylation plays a central role in regulating pollen formation.

6.4 DISCUSSION

In the context of chromatin remodification, the amino acid sequence from the MS1 PHD domain showed high similarity to H3 Lys methyltransferases in yeast, Drosophila and mammals (Figure 6.3). However, previous research argues that the conserved sequence information provides only limited value to predict potential histone methyl-lysine binding activity for PHD fingers (Aasland et al., 1995). For example, in yeast the ING2 homologues and $SPP1_{(PHD)}$ containing the full complement of conserved residues bind to H3K4me3 (Shi et al., 2007). Whereas $ECM5_{(PHD)}$, despite having virtually all the known conserved residues required for H3K4me3 binding, recognizes H3K36me3 rather than H3K4me3 (Cartagena et al., 2008). The exact substrate and product specificities of MS1 have not been demonstrated *in vitro*. The use of histone peptide microarrays containing various methylated histone peptides probed with MS1 would help to trace the molecular mechanism of the interaction between MS1 and methylated histone tails.

It seems that MS1 is of general importance for the demethylation of trimethylated H3 marks, since the overexpression of MS1 caused dramatically reduction in H3K4me3, H3K36me3 and H3K27me3 (Figure 6.3). Over-expression of MS1 was found to have a harmful effect on plant

development, suggesting that a tightly regulated system of expression is required to moderate the effects of missexpression of MS1(Yang et al., 2007). This might be explained as the consequence of altered histone modification on a general level in the overexpression lines.

The corresponding enhanced levels of H3K4me3 and H3K27me3 in *ms1ttg* mutant indicates that MS1 is able to facilitate the detrimethylation of H3 marks. It appears that the role of MS1 was not specific to a particular histone modification state, consistent with previous studies that suggest that histone lysine demethylation reactions are nonspecific, since one kind of demethylase can demethylate many methylated histone states (Tan et al., 2008). Another interesting question raised is whether MS1 itself acted as a demethylase to erase the marks or as a help hand that anchored the marked histone to enable other proteins to assemble to. Many of the demehylases are found to contain the PHD finger motif or at least to associate with a PHD finger protein (Lan et al., 2008).

A number of genes were identified as downregulated in *ms1* (Yang et al., 2007), among which the MYB99 transcription factor gene is of particular interest, proposed as a direct target of MS1 (Yang et al., 2007) (Ito et al., 2007). This was also confirmed by ChIP analysis, showing that MS1 directly regulate MYB99 by binding its promoter (see Section 5.3.2). However, this association has been determined as via direct DNA binding or

by an indirect way. In this chapter, the increased H3K36me3 levels found for MYB99 in *ms1* mutant inflorescences (Figure 6.4) are consistent with a role for MS1 in H3K36 de-methylation. It is likely that MS1 can activate the expression of MYB99 by removing H3K36me3 marks on the promoter. Loss of the MS1 function may maintain the repressive H3K36me3 marks on *MYB99*, resulting in chromatin structure of *MYB99* staying inactivated.

However, it is noticed that only H3K36me3 is increased on *MYB99*, while H3K4me3 and H3K27me3 are reduced. This give rise to questions such as, what is the key factor determine the preference of the recognition activity for PHD finger? And what is the coordination mechanism for all the proteins involving transcription regulation of pollen formation genes? As a growing number of MS1 regulatory targets are now becoming identified, it is extremely important to have deeper insights into the molecular mechanism of the regulation role of MS1.

CHAPTER 7 GENERAL DISCUSSION AND CONCLUSIONS

Male sterility in flowering plants, a phenomenon first observed by Kölreuter in 1763, has valuable significance in selective breeding, by greatly facilitating the production of hybrids via cross-pollination (Kaul, 1988). In *Arabidopsis thaliana,* molecular and genetic studies have identified numerous genes regulating stamen and pollen development, which are crucial for normal male reproductive development (Ma, 2005). One example is the *Arabidopsis MALE STERILITY1 (MS1)* gene (Wilson et al., 2001). It encodes a PHD finger motif transcriptional factor which plays a key role in tapetum development and pollen wall formation, that is highly conserved and has also been characterised in rice and barley (Li et al., 2011; Gómez and Wilson, 2012).

MS1 is essential for viable pollen formation (Wilson et al., 2001; Alves-Ferreira et al., 2007; Ito et al., 2007; Yang et al., 2007); its expression is tightly regulated in the tapetum, from the callose breakdown until the free microspore stage (Yang et al., 2007). In the *ms1* mutant, there are alterations in the production of pollen wall materials (Ariizumi et al., 2005; Vizcay-Barrena and Wilson, 2006; Yang et al., 2007), as well as a failure of tapetal programmed cell death (PCD) (Vizcay-Barrena and Wilson, 2006). This ultimately results in the failure to produce viable pollen (Wilson et al., 2001; Ito and Shinozaki, 2002). Large numbers of genes are down-regulated in the *ms1* mutant indicating that MS1 plays a key role in regulating late tapetal gene expression and pollen wall formation. Moreover,

over-expression of MS1 result in a deleterious influence on plant development, suggesting that moderation of the effects of missexpression of MS1 requires a tightly regulated system of expression (Yang et al., 2007).

7.1 MS1 PROTEIN INTERACTIONS

Eight proteins showed interaction in previous yeast-2-hybrid screens (Caiyun Yang and Z. A. Wilson, unpublished). One issue with the Y2H assay is producing false positive results (Lalonde et al., 2008), therefore potential interactions need to be confirmed by another *in vitro* method and also *in planta*. Two of the clones Y2H54 and Y2H19 (termed as POB2) showed strongest interaction in the Y2H screens, which were then tested by transient expression in tobacco epidermal cells and protein pull-down assays.

When co-expressed with MS1, the POB2 protein is co-localised with MS1 in the nucleus (see Section 3.3.1), yet Y2H54 appears localised on the cell membrane in the presence of MS1 (see Section 3.3.1), and thus was excluded from the subsequent analysis. Förster resonance energy transfer (FRET) assay was set up to demonstrate physical interaction between POB2 and MS1 (see Section 3.1.3) in tobacco epidermal cells. However, the FRET efficiency between MS1-eGFP and POB2-mRFP has not been significantly different from those of the negative controls, implying that physical interaction may not occur between MS1 and POB2.

Interestingly, the association between MS1 and POB2 is demonstrated by protein pull-down analysis, providing additional evidence apart from Y2H assay for the existence of the interaction *in vitro.* If not considering the technical issues that could cause the discrepancy between the *in vivo* and the *in vitro* results, it suggests that POB2 contains particular structures capable of assembling with MS1. However, preliminary analysis of newly developed POB2-GUS lines suggest that POB2 expression may be delayed compared with MS1, however further analysis of these lines is required (Simpson and Wilson, unpublished). This suggests that there may be a discrepancy in the temporal expression pattern of MS1 and POB2, raising further questions as to whether POB2 interacts with MS1 *in Arabidopsis thaliana*.

7.2 AMPLIFICATION AND CHARACTERIZATION OF Y2H54

Although Y2H54 failed to prove itself as a nuclear-localised protein when co-expressed with MS1 in the tobacco epidermal cells, making it less likely that it acts as a MS1 direct interacting protein (Figure 3.7), the association of Y2H54 with pollen development pathways is still evident. It has been shown specifically expressed in the pollen grains as well as at the plasma membrane of the shank zone in the growing pollen tubes (Deeks et al., 2012), with expression being changed throughout pollen grain germination and pollen tube growth (Wang et al., 2008). This may suggest a role for Y2H54 in regulating plant fertility.

7.2.1 Y2H54 Transcription Profile

In this project, the 5' end RACE PCR result suggests that the coding sequence of Y2H54 is 2888 bp in length, with two exons separated by one intron, which is different from the publically available data. Interestingly, various transcription initiations have been found by sequence alignment, but all generate a common functional polypeptide of 951aa. A highly homologous gene *At1g09720* termed as *Y2H54_Like* was identified by BLAST analysis, with 73% similarity at the cDNA level and 77% at the amino-acid sequence.

Quantitative expression analysis revealed that Y2H54 is expressed predominantly in young buds, old buds, open flowers and leaves, with maximal expression in the old buds (see Section 4.3.2). The Y2H54-GFP fusions under the Y2H54 endogenous promoter stably expressed in *Arabidopsis thaliana* reveals the gene expression is seen predominantly at tapetum membrane during anther stage 7~8, and slightly in the microspores at anther stage 12~13 (see Section 4.3.2). This is in agreement with the quantitative expression analysis, considering that the amount of microspores is preponderant in the anthers compared with tapetal cells, though the gene expression in individual microspores is more faint.

7.2.2 *Y2H54* RNAi Silencing

RNAi of *Y2H54* was performed in *Arabidopsis thaliana*, which resulted in various levels of gene silencing, with respectively 10%, 50%, 70% and 90% of expression levels decreased in different lines (See section 4.3.4). However, no alteration in fertility or reduction in pollen viability was observed even in T1 and T2 RNAi plants with maximal expression reduction, indicating that the silencing was not sufficient to impair overall function. RNAi gene silencing is frequently associated with knocking down gene expression rather than knocking it out (Baulcombe, 2004). Therefore, the residual amount of transcript may be sufficient to maintain wild type function. Nevertheless, it doesn't exclude the possibility that alternative proteins act redundantly with Y2H54. However, further confirmation of the role of Y2H54 could be obtained using stable null mutant generated by other methods, for example artificial micro RNA (Eulalio et al., 2008).

7.2.3 Y2H54 Overexpression

Abnormal plant morphologies in many organs have been observed in Y2H54 overexpression lines. These transgenic lines grew faster and stronger with enlarged size of rosette leaves, visibly sturdier stems and more branches if compared with wild type (see Section 4.3.5). It is interesting to see enlarged inflorescences as a common phenotype for the Y2H54

overexpression lines and the *ms1ttg* mutants, which has been even intensified in the hybrids of the two.

Y2H54 serves as the actin-membrane nexus, whose NAB motifs have been proved capable of binding F-actin (Deeks et al., 2012). Coordinated actin-binding proteins have been hypothesized to be essential for maintaining proper cell morphology by affecting cell growth (Hussey et al., 2006). The altered morphologies of Y2H54 showed some similarities with previous research on actin pathways (Ramachandran et al., 2000; Chen et al., 2002; Gu et al., 2005), which may imply that actin organisation is changed in the overexpression lines.

It appears that knocking out of MS1 also lead to stimulated cell growth, for instance inflorescences are enhanced in size, with an increased number of buds. Additionally, numerous genes involving actin organisation are found down-regulated in the old buds of the *ms1ttg* mutant (see section 5.3.1), suggesting that loss of MS1 may have an effect on the controlled actin activities, probably in an indirect way.

7.3 MS1 REGULATES POLLEN EXINE FORMATION THROUGH MULTIPLE MECHANISMS

7.3.1 MS1 Directly Regulates MYB99 via H3K36me3

Previously, the MYB transcription factor MYB99 has been proposed as a direct target of MS1 (Alves-Ferreira et al., 2007; Ito et al., 2007; Yang et al., 2007). However, evidence based on direct molecular experiments has never

been reported. Here, we confirmed that MS1 binds to the promoter of MYB99 *in vivo* by a robust ChIP assay, with the binding site being closed to 5'UTR (See section 5.3.2). Considering that MS1 belongs to a PHD finger transcription factor assembling modified H3 tails, and that no additional DNA-binding domain has been identified on *MS1*, it appears MS1 associates with the promoter of MYB99 through methylated histone tails, rather than direct DNA-binding.

In the overexpression lines, MS1 contributes to demethylate the trimethyl-H3 marks nonspecifically (see Section 6.3.2), with global levels of H3K4me3, H3K27me3 and H3K36me3 all enhanced. However, it's still not clear whether MS1 itself acts as a demethylase to erase the marks, or in a supporting role to anchor the marked histones to enable other proteins to function. The trimethyl-H3 marks were examined specifically on *MYB99* in the *ms1ttg* mutant (see section 6.3.3), this showed that level of H3K36me3 was increased on *MYB99,* whereas the other two were reduced. It appears that loss of the MS1 function may mean that H3K36me3 cannot be demethylated into other states, with the chromatin structure of MYB99 being maintained inactivated. Consistent with this, previous researches show that tri-methylation of histone H3 on lysine 36 is repressive when found in the promoter region (Pfluger and Wagner, 2007), where the enrichment of *MYB99* by MS1 occurred. Taken together, these results suggest that MS1 directly regulates MYB99 through detrimethylation of H3K36me3, which is required for activation of the *MYB99* chromatin.

7.3.2 MS1 Collaborates with AMS to Regulate KCS7 and KCS15

Other candidates showing close association with *MYB99* were examined by ChIP analysis, including *KCS7*, *KCS15*, *KC721, ACOS5, SHT* and *TSM1* (see section 5.3.2.) Noticeably, *KCS7, KCS15, SHT* and *TSM1* showed enrichments, these are proposed as direct regulatory targets of MS1, while *ATA7, CYP86C3* and *KC721* failed to be enriched. In these experiments, only a small region of each gene sequence were tested, but the enrichments of particular genes can be very specific, for instance that seen for *MYB99.* Even though the examined regions of *ATA7*, *CYP86C3* and *KC721* failed to be enriched by MS1, it cannot entirely be excluded that these genes also act as direct targets of MS1.

Previously, KCS7 and KCS15 were also identified as direct targets of AMS, whose promoter regions were enriched by AMS in the ChIP analysis (Xu et al., 2010). These genes are also found to be dramatically reduced in the *ms1* mutants (Yang et al., 2007), indicating that AMS alone is insufficient to activate their transcription, with the MS1 regulator still required. This means that MS1 and AMS may specifically collaborate in promoting the transcription of the two genes.

7.3.3 Other Potential Regulatory Pathways

We established an *Arabidopsis thaliana* correlation network (FlowerNet) of global transcriptional interactions associated with anther development (Simon Pierce and Zoe Wilson, unpublished). Gene Ontology (GO)

overrepresentation analysis was performed on the microarray dataset for *ms1* mutant mapped with the FlowerNet correlation network (see Section 5.3.1). It suggests that MS1 acts as a critical regulator for the normal expression of pathways including lipid mechanism and pollen exine formation in the young anthers, and pathways of cell wall modification and pollen tube growth in the old ones. Altered sucrose transport was also observed in the *ms1*, which may be relevant to the defection in transportation of precursors of sporopollenin.

Genes regulated in the young anthers are functionally related, involving lipid metabolism and transport (*KCS7, KCS15, KCS21, CYP86C3, CYP98A8* and *CYP98A9*), and pollen coat formation (*EXL4* and *EXL6*). Some of these candidates have previously been identified as direct targets of AMS by ChIP analysis, including *EXL4, EXL6, KCS7, KCS15, GRP18, GRP19, CYP98A8, CYP98A9*, and *At1g06990* (GDSL-like Lipase) (Xu et al., 2010). There are genes also down-regulated by MS1, DYT1 and AMS, while conversely, certain DYT1 and AMS common downstream genes were up-regulated in *ms1*, for instance *ACOS5, KNS2, LAP5, DRL1, CYP704B1* and *CYP703A2*. This implies that various mechanisms of transcriptional control exist in the process of pollen formation. Regulators may control the pathways indirectly, i.e. through multiple feedback loops and feed-forward loops.

7.4 HISTONE METHYLATION PLAYS A CENTRAL ROLE IN REGULATING POLLEN FORMATION

Trimethylation of H3 marks were examined specifically on genes correlated to sporopollenin biosynthesis, including *CYP703A2, CYP704B1, ACOS5, DRL1* and *LAP5* (see Section 6.3.4). In wild type plants, relatively high levels of the H3K4me3, H3K27me3 and H3K36me3 marks are widely found. Some genes carry bivalent marks of both active H3K4me3 and repressive H3K27me3 marks, whose transcription activations require complex regulating mechanisms by recruitment of methylases and demethylases, to resolve bivalent domain into monovalent domain (Lan et al., 2008).

In the *ms1* mutant, levels of trimethylation at histone lysines are dramatically changed. For instance, levels of H3K27me3 marks were barely detected on *CYP704B1* in wild type, but considerably increased in the *ms1ttg*, with H3K4me3 and H3K27me3 simultaneously enhanced. The overprint effects of various histone marks eventually caused the up-regulation of *CYP704B1* in *ms1ttg*, as indicated by microarray analysis.

Additionally, many of the MS1 regulatory targets identified are found to carry repressive H3K27me3 marks (see Section 5.3.1), which might act as a key mechanism for the transcription activation of these genes. In summary, these studies indicate that histone methylation plays a central role in regulating pollen formation.

7.5 FUTURE PERSPECTIVES

Previous research showed that MS1 plays a key role in regulating late tapetum gene expression and pollen wall deposition (Alves-Ferreira et al., 2007; Ito et al., 2007; Yang et al., 2007). In this thesis, multiple regulatory factors have been investigated, including protein interaction, DNA binding, gene regulatory network, and histone modification.

Wider and deeper research focusing on the regulatory system of MS1 will be needed in the future. For instance, the global map of the MS1 binding sites across the entire *Arabidopsis thaliana* genome, achieved by ChIP-seq, will facilitate an understanding of the integrated regulatory networks of pollen formation. Additionally histone peptide microarrays containing various methylated histone peptides probed with MS1 would be worthwhile to trace the molecular mechanism of the interaction between MS1 and methylated histone tails. As all the PHD finger transcription factor identified to date act as a structural unit of the large complex (Sanchez and Zhou, 2011), continuing characterisation of the MS1 putative interacting proteins would also give us valuable insights into the roles of MS1 in transcriptional regulation.

REFERENCES

Aasland R, Gibson TJ, Stewart AF (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem Sci 20: 56-59

Aexander MP (1969) Differential staining of aborted and non-aborted pollen. Stain Technol 44: 117-122

Ahlers F, Lambert J, Wiermann R (1999) Structural Elements of Sporopollenin from the Pollen of Torreya californica Torr.(Gymnospermae): Using the^ 1H-NMR Technique. ZEITSCHRIFT FUR NATURFORSCHUNG C 54: 492-495

Albrecht C, Russinova E, Hecht V, Baaijens E, de Vries S (2005) The *Arabidopsis thaliana* SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1 and 2 control male sporogenesis. Plant Cell 17: 3337-3349

Alves-Ferreira M, Wellmer F, Banhara A, Kumar V, Riechmann JL, Meyerowitz EM (2007) Global expression profiling applied to the analysis of *Arabidopsis* stamen development. Plant Physiol 145: 747-762

***Arabidopsis* Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796

Ariizumi T, Hatakeyama K, Hinata K, Inatsugi R, Nishida I, Sato S, Kato T, Tabata S, Toriyama K (2004) Disruption of the novel plant

protein NEF1 affects lipid accumulation in the plastids of the tapetum and exine formation of pollen, resulting in male sterility in *Arabidopsis thaliana*. Plant J 39: 170-181

Ariizumi T, Hatakeyama K, Hinata K, Sato S, Kato T, Tabata S, Toriyama K (2003) A novel male-sterile mutant of *Arabidopsis thaliana*, faceless pollen-1, produces pollen with a smooth surface and an acetolysis-sensitive exine. Plant Mol Biol 53: 107-116

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

Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C (2004) Vernalization requires epigenetic silencing of FLC by histone methylation. Nature 427: 164-167

Baulcombe D (2004) RNA silencing in plants. Nature 431: 356-363

Berendzen K, Searle I, Ravenscroft D, Koncz C, Batschauer A, Coupland G, Somssich IE, ‹lker B (2005) A rapid and versatile combined DNA/RNA extraction protocol and its application to the analysis of a novel DNA marker set polymorphic between *Arabidopsis thaliana* ecotypes Col-0 and Landsberg erecta. Plant Methods 1: 1-15

Berr A, McCallum EJ, M¨¦nard R, Meyer D, Fuchs J, Dong A, Shen WH (2010) *Arabidopsis* SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte

development. The Plant Cell 22: 3232

Bhat RA, Lahaye T, Panstruga R (2006) The visible touch: *in planta* visualization of protein-protein interactions by fluorophore-based methods. Plant Methods 2: 12

Blackmore S, Wortley AH, Skvarla JJ, Rowley JR (2007) Pollen wall development in flowering plants. New Phytol 174: 483-498

Borg M, Brownfield L, Twell D (2009) Male gametophyte development: a molecular perspective. J Exp Bot 60**:** 1465-1478

Bowler C, Benvenuto G, Laflamme P, Molino D, Probst AV, Tariq M, Paszkowski J (2004) Chromatin techniques for plant cells. Plant J 39**:** 776-789

Bubert H, Lambert J, Steuernagel S, Ahlers F, Wiermann R (2002) Continuous decomposition of sporopollenin from pollen of Typha angustifolia L. by acidic methanolysis. Z Naturforsch C 57: 1035-1041

Canales C, Bhatt AM, Scott R, Dickinson H (2002) A putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in Arabidopsis. Current Biology 12: 1718-1727

Cartagena JA, Matsunaga S, Seki M, Kurihara D, Yokoyama M, Shinozaki K, Fujimoto S, Azumi Y, Uchiyama S, Fukui K (2008) The

Arabidopsis SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen. Developmental biology 315: 355-368

Cartagena JA, Matsunaga S, Seki M, Kurihara D, Yokoyama M, Shinozaki K, Fujimoto S, Azumi Y, Uchiyama S, Fukui K (2008) The *Arabidopsis* SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen. Developmental biology 315: 355-368

Chakravarty S, Zeng L, Zhou MM (2009) Structure and site-specific recognition of histone H3 by the PHD finger of human autoimmune regulator. Structure 17: 670-679

Chen CY, Wong EI, Vidali L, Estavillo A, Hepler PK, Wu HM, Cheung AY (2002) The regulation of actin organization by actin-depolymerizing factor in elongating pollen tubes. Plant Cell 14: 2175-2190

Chien CT, Bartel PL, Sternglanz R, Fields S (1991) The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proc Natl Acad Sci U S A 88: 9578-9582

Chignola F, Gaetani M, Rebane A, Org T, Mollica L, Zucchelli C, Spitaleri A, Mannella V, Peterson P, Musco G (2009) The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation. Nucleic Acids Res 37: 2951-2961

Chignola F, Hetenyi C, Gaetani M, Rebane A, Liiv I, Maran U, Mollica L, Bottomley MJ, Musco G, Peterson P (2008) The autoimmune regulator PHD finger binds to non‐methylated histone H3K4 to activate gene expression. EMBO reports 9: 370-376

Choi H, Jin JY, Choi S, Hwang JU, Kim YY, Suh MC, Lee Y (2011) An ABCG/WBC‐type ABC transporter is essential for transport of sporopollenin precursors for exine formation in developing pollen. The Plant Journal 65: 181-193

Christians MJ, Gingerich DJ, Hua Z, Lauer TD, Vierstra RD (2012) The light-response BTB1 and BTB2 proteins assemble nuclear ubiquitin ligases that modify phytochrome B and D signaling in Arabidopsis. Plant Physiology 160: 118-134

Chuang CF, Meyerowitz EM (2000) Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 97: 4985-4990

Colcombet J, Boisson-Dernier A, Ros-Palau R, Vera CE, Schroeder JI (2005) *Arabidopsis* SOMATIC EMBRYOGENESIS RECEPTOR KINASES1 and 2 are essential for tapetum development and microspore maturation. Plant Cell 17: 3350-3361

de Azevedo Souza C, Kim SS, Koch S, Kienow L, Schneider K, McKim SM, Haughn GW, Kombrink E, Douglas CJ (2009) A novel fatty

Acyl-CoA Synthetase is required for pollen development and sporopollenin biosynthesis in Arabidopsis. The Plant Cell 21: 507-525

Deeks MJ, Calcutt JR, Ingle EK, Hawkins TJ, Chapman S, Richardson AC, Mentlak DA, Dixon MR, Cartwright F, Smertenko AP (2012) A superfamily of actin-binding proteins at the actin-membrane nexus of higher plants. Current Biology 22: 1595-1600

Dobritsa AA, Lei Z, Nishikawa S, Urbanczyk-Wochniak E, Huhman DV, Preuss D, Sumner LW (2010) LAP5 and LAP6 encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in Arabidopsis. Plant Physiol 153: 937-955

Dobritsa AA, Nishikawa S-I, Preuss D, Urbanczyk-Wochniak E, Sumner LW, Hammond A, Carlson AL, Swanson RJ (2009) LAP3, a novel plant protein required for pollen development, is essential for proper exine formation. Sexual plant reproduction 22: 167-177

Dobritsa AA, Shrestha J, Morant M, Pinot F, Matsuno M, Swanson R, Møller BL, Preuss D (2009) CYP704B1 is a long-chain fatty acid ω-hydroxylase essential for sporopollenin synthesis in pollen of Arabidopsis. Plant Physiology 151: 574-589

Dong X, Hong Z, Sivaramakrishnan M, Mahfouz M, Verma DPS (2005) Callose synthase (CalS5) is required for exine formation during microgametogenesis and for pollen viability in Arabidopsis. The Plant Journal 42: 315-328

Ecker RC, de Martin R, Steiner GE, Schmid JA (2004) Application of spectral imaging microscopy in cytomics and fluorescence resonance energy transfer (FRET) analysis. Cytometry Part A 59: 172-181

Eulalio A, Huntzinger E, Izaurralde E (2008) Getting to the root of miRNA-mediated gene silencing. Cell 132: 9-14

Fang RX, Nagy F, Sivasubramaniam S, Chua NH (1989) Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. Plant Cell 1: 141-150

Feng B, Lu D, Ma X, Peng Y, Sun Y, Ning G, Ma H (2012) Regulation of the *Arabidopsis* anther transcriptome by DYT1 for pollen development. Plant J 72: 612-624

Feng X, Dickinson HG (2010) Tapetal cell fate, lineage and proliferation in the *Arabidopsis* anther. Development 137: 2409-2416

Frankel R, Izhar S, Nitsan J (1969) Timing of callase activity and cytoplasmic male sterility in Petunia. Biochem Genet 3: 451-455

Gabarayeva N, Grigorjeva V, Rowley JR, Hemsley AR (2009) Sporoderm development in Trevesia burckii (Araliaceae). I. Tetrad period: Further evidence for the participation of self-assembly processes. Review of Palaeobotany and Palynology 156: 211-232

Gadella Jr TW, van der Krogt GN, Bisseling T (1999) GFP-based FRET

microscopy in living plant cells. Trends in plant science 4: 287-291

Gendall AR, Levy YY, Wilson A, Dean C (2001) The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. Cell 107: 525-535

Gingerich DJ, Gagne JM, Salter DW, Hellmann H, Estelle M, Vierstra RD (2005) Cullin 3A and B assemble with members of the broad complex/tramtrack/bric-A-brac (BTB). Journal of Biological Chemistry

Goldberg RB, Beals TP, Sanders PM (1993) Anther development: basic principles and practical applications. Plant Cell 5: 1217-1229

Gómez JF, Wilson ZA (2012) Non-destructive staging of barley reproductive development for molecular analysis based upon external morphology. Journal of experimental botany 63: 4085-4094

Gong W, Suzuki K, Russell M, Riabowol K (2005) Function of the ING family of PHD proteins in cancer. The international journal of biochemistry & cell biology 37: 1054-1065

Greb T, Mylne JS, Crevillen P, Geraldo N, An H, Gendall AR, Dean C (2007) The PHD finger protein VRN5 functions in the epigenetic silencing of *Arabidopsis* FLC. Current biology 17: 73-78

Green MR, Sambrook J (2012) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press New York

Grefen C, Donald N, Hashimoto K, Kudla J, Schumacher K, Blatt MR (2010) A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. Plant J 64: 355-365

Grini PE, Thorstensen T, Alm V, Vizcay-Barrena G, Windju SS, Jorstad TS, Wilson ZA, Aalen RB (2009) The ASH1 HOMOLOG 2 (ASHH2) histone H3 methyltransferase is required for ovule and anther development in Arabidopsis. PLoS One 4: e7817

Grini PE, Thorstensen T, Alm V, Vizcay-Barrena G, Windju SS, Jørstad TS, Wilson ZA, Aalen RB (2009) The ASH1 HOMOLOG 2 (ASHH2) histone H3 methyltransferase is required for ovule and anther development in Arabidopsis. PLoS One 4: e7817

Gu Y, Fu Y, Dowd P, Li S, Vernoud V, Gilroy S, Yang Z (2005) A Rho family GTPase controls actin dynamics and tip growth via two counteracting downstream pathways in pollen tubes. J Cell Biol 169: 127-138

Guan Y-F, Huang X-Y, Zhu J, Gao J-F, Zhang H-X, Yang Z-N (2008) RUPTURED POLLEN GRAIN1, a member of the MtN3/saliva gene family, is crucial for exine pattern formation and cell integrity of microspores in Arabidopsis. Plant Physiology 147: 852-863

Haring M, Offermann S, Danker T, Horst I, Peterhansel C, Stam M (2007) Chromatin immunoprecipitation: optimization, quantitative analysis

and data normalization. Plant Methods **3:** 11

Haring M, Offermann S, Danker T, Horst I, Peterhansel C, Stam M (2007) Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. Plant Methods 3: 11

Hartley JL, Temple GF, Brasch MA (2000) DNA cloning using *in vitro* site-specific recombination. Genome research 10: 1788-1795

Hawkins TJ, Deeks MJ, Wang P, Hussey PJ (2014) The evolution of the actin binding NET superfamily. Front Plant Sci 5: 254

Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM (2000) pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. Plant Mol Biol 42: 819-832

Higginson T, Li SF, Parish RW (2003) AtMYB103 regulates tapetum and trichome development in *Arabidopsis thaliana*. Plant J 35: 177-192

Hord CL, Chen C, Deyoung BJ, Clark SE, Ma H (2006) The BAM1/BAM2 receptor-like kinases are important regulators of *Arabidopsis* early anther development. Plant Cell 18: 1667-1680

Hu CD, Chinenov Y, Kerppola TK (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. Molecular cell 9: 789-798

Huang X-Y, Niu J, Sun M-X, Zhu J, Gao J-F, Yang J, Zhou Q, Yang

Z-N (2013) CYCLIN-DEPENDENT KINASE G1 is associated with the spliceosome to regulate CALLOSE SYNTHASE5 splicing and pollen wall formation in Arabidopsis. The Plant Cell 25: 637-648

Hung T, Binda O, Champagne KS, Kuo AJ, Johnson K, Chang HY, Simon MD, Kutateladze TG, Gozani O (2009) ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. Molecular cell 33: 248-256

Hussey PJ, Ketelaar T, Deeks MJ (2006) Control of the actin cytoskeleton in plant cell growth. Annu Rev Plant Biol 57: 109-125

Ito T, Nagata N, Yoshiba Y, Ohme-Takagi M, Ma H, 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, 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 and cell physiology 43: 1285

Jiang J, Zhang Z, Cao J (2013) Pollen wall development: the associated enzymes and metabolic pathways. Plant Biol (Stuttg) 15: 249-263

Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of *in vivo* protein-DNA interactions. Science 316: 1497-1502

Joung JK, Ramm EI, Pabo CO (2000) A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. Proceedings of the National Academy of Sciences 97: 7382

Karimi M, Inzé D, Depicker A (2002) GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. Trends in plant science 7: 193-195

Kaufmann K, Muino JM, Osteras M, Farinelli L, Krajewski P, Angenent GC (2010) Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nat Protoc 5: 457-472

Kaul ML (1988) Male sterility in higher plants. Springer-Verlag

Kim SS, Grienenberger E, Lallemand B, Colpitts CC, Kim SY, Souza Cde A, Geoffroy P, Heintz D, Krahn D, Kaiser M, Kombrink E, Heitz T, Suh DY, Legrand M, Douglas CJ (2010) LAP6/POLYKETIDE SYNTHASE A and LAP5/POLYKETIDE SYNTHASE B encode hydroxyalkyl alpha-pyrone synthases required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana*. Plant Cell 22: 4045-4066

Klee H, Horsch R, Rogers S (1987) Agrobacterium-mediated plant transformation and its further applications to plant biology. Annual Review of Plant Physiology 38**:** 467-486

Knox RB, Heslop-Harrison J (1971) Pollen-wall proteins: the fate of intine-held antigens on the stigma in compatible and incompatible pollinations of Phalaris tuberosa L. J Cell Sci 9: 239-251

Koltunow AM, Truettner J, Cox KH, Wallroth M, Goldberg RB (1990) Different Temporal and Spatial Gene Expression Patterns Occur during Anther Development. Plant Cell 2: 1201-1224

Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an insertional mutagen in Arabidopsis. The Plant Cell 11: 2283-2290

Lalonde S, Ehrhardt DW, Loque D, Chen J, Rhee SY, Frommer WB (2008) Molecular and cellular approaches for the detection of protein-protein interactions: latest techniques and current limitations. Plant J 53: 610-635

Lan F, Collins RE, De Cegli R, Alpatov R, Horton JR, Shi X, Gozani O, Cheng X, Shi Y (2007) Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature 448: 718-722

Lan F, Nottke AC, Shi Y (2008) Mechanisms involved in the regulation of histone lysine demethylases. Curr Opin Cell Biol 20: 316-325

Landy A (1989) Dynamic, structural, and regulatory aspects of lambda site-specific recombination. Annu Rev Biochem 58: 913-949

Langenkamper G, Fung RW, Newcomb RD, Atkinson RG, Gardner

RC, MacRae EA (2002) Sucrose phosphate synthase genes in plants belong to three different families. J Mol Evol 54: 322-332

Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, Patel DJ (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature 442: 91-95

Li H, Pinot F, Sauveplane V, Werck-Reichhart D, Diehl P, Schreiber L, Franke R, Zhang P, Chen L, Gao Y (2010) Cytochrome P450 family member CYP704B2 catalyzes the ω-hydroxylation of fatty acids and is required for anther cutin biosynthesis and pollen exine formation in rice. The Plant Cell 22: 173-190

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

Li SF, Iacuone S, Parish RW (2007) Suppression and restoration of male fertility using a transcription factor. Plant biotechnology journal 5: 297-312

Livak KJ, Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. methods 25: 402-408

Lutfiyya LL, Xu N, D'Ordine RL, Morrell JA, Miller PW, Duff SM (2007) Phylogenetic and expression analysis of sucrose phosphate synthase

isozymes in plants. J Plant Physiol 164: 923-933

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

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

Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. Nature 469: 343-349

Mariani C, Beuckeleer MD, Truettner J, Leemans J, Goldberg RB (1990) Induction of male sterility in plants by a chimaeric ribonuclease gene. Nature 347: 737-741

Massie CE, Mills IG (2009) Chromatin immunoprecipitation (ChIP) methodology and readouts. Methods Mol Biol 505: 123-137

Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M (1998) *Arabidopsis thaliana*: a model plant for genome analysis. Science 282: 662, 679-682

Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. The Plant Cell 11: 949-956

Mizuno S, Osakabe Y, Maruyama K, Ito T, Osakabe K, Sato T,

Shinozaki K, Yamaguchi‐Shinozaki K (2007) Receptor‐like protein kinase 2 (RPK 2) is a novel factor controlling anther development in *Arabidopsis thaliana*. The Plant Journal 50: 751-766

Morant M, Jørgensen K, Schaller H, Pinot F, Møller BL, Werck-Reichhart D, 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. The Plant Cell 19: 1473-1487

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiologia plantarum 15: 473-497

Musselman CA, Kutateladze TG (2009) PHD fingers: epigenetic effectors and potential drug targets. Molecular Interventions 9: 314

Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng 104: 34-41

Odell JT, Nagy F, Chua NH (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature **313:** 810-812

Osthoff KS, Wiermann R (1987) Phenols as integrated compounds of

sporopollenen from Pinus pollen. Journal of plant physiology 131: 5-15

Owen H, Makaroff C (1995) Ultrastructure of microsporogenesis and microgametogenesis in *Arabidopsis thaliana* (L.). Heynh. ecotype Wassilewskija (Brassicaceae). Protoplasma**:** 7-21

Pacini E, Franchi G, Hesse M (1985) The tapetum: Its form, function, and possible phylogeny inEmbryophyta. Plant Systematics and Evolution 149: 155-185

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

Parish RW, Li SF (2010) Death of a tapetum: a programme of developmental altruism. Plant Science 178: 73-89

Paxson-Sowders D, Owen H, Makaroff C (1997) A comparative ultrastructural analysis of exine pattern development in wild-type*Arabidopsis* and a mutant defective in pattern formation. Protoplasma 198: 53-65

Paxson-Sowders DM, Dodrill CH, Owen HA, Makaroff CA (2001) DEX1, a novel plant protein, is required for exine pattern formation during pollen development in Arabidopsis. Plant Physiology 127: 1739-1749

Pfluger J, Wagner D (2007) Histone modifications and dynamic regulation of genome accessibility in plants. Curr Opin Plant Biol 10: 645-652

Phan HA, Li SF, Parish RW (2012) MYB80, a regulator of tapetal and pollen development, is functionally conserved in crops. Plant molecular biology 78: 171-183

Piffanelli P, Ross JHE, Murphy D (1998) Biogenesis and function of the lipidic structures of pollen grains. Sexual Plant Reproduction 11: 65-80

Promega Corporation (2009) Characterizing Protein Interactions Using GST Pull-Down Assays and Cell-Free Expression. BioTechniques

Qin P, Tu B, Wang Y, Deng L, Quilichini TD, Li T, Wang H, Ma B, Li S (2013) ABCG15 encodes an ABC transporter protein, and is essential for post-meiotic anther and pollen exine development in Rice. Plant and Cell Physiology 54: 138-154

Quilichini TD, Friedmann MC, Samuels AL, Douglas CJ (2010) ATP-binding cassette transporter G26 is required for male fertility and pollen exine formation in *Arabidopsis*. Plant Physiology 154: 678-690

Ramachandran S, Christensen HE, Ishimaru Y, Dong CH, Chao-Ming W, Cleary AL, Chua NH (2000) Profilin plays a role in cell elongation, cell shape maintenance, and flowering in *Arabidopsis*. Plant Physiol 124: 1637-1647

Ramón-Maiques S, Kuo AJ, Carney D, Matthews AG, Oettinger MA, Gozani O, Yang W (2007) The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2.

Proceedings of the National Academy of Sciences 104: 18993-18998

Saiga S, Möller B, Watanabe-Taneda A, Abe M, Weijers D, Komeda Y (2012) Control of embryonic meristem initiation in *Arabidopsis* by PHD-finger protein complexes. Development 139: 1391-1398

Sambrook J, Russell DW (2006) Detection of Protein-Protein Interactions Using the GST Fusion Protein Pulldown Technique. Cold Spring Harbor Protocols 2006: pdb. prot3757

Sanchez R, Zhou MM (2011) The PHD finger: a versatile epigenome reader. Trends Biochem Sci 36: 364-372

Sanders PM, Bui AQ, Weterings K, McIntire K, Hsu YC, Lee PY, Truong MT, Beals T, Goldberg R (1999) Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. Sexual Plant Reproduction 11: 297-322

Schindler U, Beckmann H, Cashmore AR (1993) HAT3. 1, a novel *Arabidopsis* homeodomain protein containing a conserved cysteine‐rich region. The Plant Journal 4: 137-150

Scott R (1994) Pollen exine: the sporopollenin enigma and the physics of pattern. Molecular and cellular aspects of plant reproduction 81

Scott RJ, Spielman M, Dickinson HG (2004) Stamen structure and function. Plant Cell 16 Suppl: S46-60

Sekar RB, Periasamy A (2003) Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. J Cell Biol 160: 629-633

Shah A (2009) Chromatin immunoprecipitation sequencing (ChIP-Seq) on the SOLiD™ system. Nature Methods 6

Shah K, Russinova E, Gadella TW, Jr., Willemse J, De Vries SC (2002) The *Arabidopsis* kinase-associated protein phosphatase controls internalization of the somatic embryogenesis receptor kinase 1. Genes Dev 16: 1707-1720

Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES (2000) The molecular basis of vernalization: the central role of FLOWERING LOCUS C (FLC). Proc Natl Acad Sci U S A 97: 3753-3758

Shi X, Kachirskaia I, Walter KL, Kuo J-HA, Lake A, Davrazou F, Chan SM, Martin DG, Fingerman IM, Briggs SD (2007) Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. Journal of Biological Chemistry 282: 2450-2455

Shi X, Kachirskaia I, Walter KL, Kuo JHA, Lake A, Davrazou F, Chan SM, Martin DGE, Fingerman IM, Briggs SD (2007) Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. Journal of Biological Chemistry 282: 2450-2455

Shukla A, Vijayaraghavan M, Chaudhry B (1998) Biology of pollen. APH Publishing

Skirpan AL, McCubbin AG, Ishimizu T, Wang X, Hu Y, Dowd PE, Ma H, Kao T-h (2001) Isolation and characterization of kinase interacting protein 1, a pollen protein that interacts with the kinase domain of PRK1, a receptor-like kinase of petunia. Plant Physiology 126: 1480-1492

Smith NA, Singh SP, Wang M-B, Stoutjesdijk PA, Green AG, Waterhouse PM (2000) Gene expression: Total silencing by intron-spliced hairpin RNAs. Nature 407: 319-320

Smyth DR, Bowman JL, Meyerowitz EM (1990) Early flower development in Arabidopsis. Plant Cell 2: 755-767

Sorensen AM, Krober S, Unte US, Huijser P, Dekker K, Saedler H (2003) The *Arabidopsis* ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. Plant J 33: 413-423

Stevens VA, Murray BG (1981) Studies on heteromorphic self-incompatibility systems: the cytochemistry and ultrastructure of the tapetum of Primula obconica. J Cell Sci 50: 419-431

Stieglitz H (1977) Role of beta-1, 3-glucanase in postmeiotic microspore release. Developmental biology 57: 87-97

Stoutjesdijk PA, Singh SP, Liu Q, Hurlstone CJ, Waterhouse PA,

Green AG (2002) hpRNA-mediated targeting of the *Arabidopsis* FAD2 gene gives highly efficient and stable silencing. Plant Physiol 129: 1723-1731

Tan H, Wu S, Wang J, Zhao ZK (2008) The JMJD2 members of histone demethylase revisited. Mol Biol Rep 35: 551-556

Taverna SD, Ilin S, Rogers RS, Tanny JC, Lavender H, Li H, Baker L, Boyle J, Blair LP, Chait BT (2006) Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. Molecular cell 24: 785-796

Thorstensen T, Grini PE, Mercy IS, Alm V, Erdal S, Aasland R, Aalen RB (2008) The *Arabidopsis* SET-domain protein ASHR3 is involved in stamen development and interacts with the bHLH transcription factor ABORTED MICROSPORES (AMS). Plant molecular biology 66: 47-59

Tzafrir I, Pena-Muralla R, Dickerman A, Berg M, Rogers R, Hutchens S, Sweeney TC, McElver J, Aux G, Patton D, Meinke D (2004) Identification of genes required for embryo development in *Arabidopsis*. Plant Physiol 135: 1206-1220

Valster AH, Pierson ES, Valenta R, Hepler PK, Emons A (1997) Probing the Plant Actin Cytoskeleton during Cytokinesis and Interphase by Profilin Microinjection. Plant Cell **9:** 1815-1824

Valvekens D, Van Montagu M, Van Lijsebettens M (1988)

Agrobacterium tumefaciens-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. Proceedings of the National Academy of Sciences 85: 5536-5540

van Ingen H, van Schaik F, Wienk H, Ballering J, Rehmann H, Dechesne AC, Kruijzer JA, Liskamp RM, Timmers H, Boelens R (2008) Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure 16: 1245-1256

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

Voinnet O, Rivas S, Mestre P, Baulcombe D (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J 33: 949-956

Wang GG, Song J, Wang Z, Dormann HL, Casadio F, Li H, Luo JL, Patel DJ, Allis CD (2009) Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature 459: 847-851

Wang Y, Zhang W-Z, Song L-F, Zou J-J, Su Z, Wu W-H (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. Plant Physiology 148: 1201-1211

Waterhouse PM, Helliwell CA (2003) Exploring plant genomes by RNA-induced gene silencing. Nat Rev Genet 4**:** 29-38

Waterhouse PM, Wang M-B, Lough T (2001) Gene silencing as an adaptive defence against viruses. Nature 411: 834-842

Wen H, Li J, Song T, Lu M, Kan PY, Lee MG, Sha B, Shi X (2010) Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. J Biol Chem 285: 9322-9326

Wilson ZA, Morroll SM, Dawson J, Swarup R, Tighe PJ (2001) The *Arabidopsis* MALE STERILITY1 (MS1) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD©\finger family of transcription factors. The Plant Journal 28: 27-39

Wilson ZA, Song J, Taylor B, Yang C (2011) The final split: the regulation of anther dehiscence. J Exp Bot 62: 1633-1649

Wilson ZA, Zhang DB (2009) From *Arabidopsis* to rice: pathways in pollen development. J Exp Bot 60: 1479-1492

Wittkop T, Emig D, Truss A, Albrecht M, Bocker S, Baumbach J (2011) Comprehensive cluster analysis with Transitivity Clustering. Nat Protoc 6: 285-295

Wood CC, Robertson M, Tanner G, Peacock WJ, Dennis ES, Helliwell CA (2006) The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. Proceedings of the National Academy of Sciences 103: 14631-14636

Worrall D, Hird DL, Hodge R, Paul W, Draper J, Scott R (1992) Premature dissolution of the microsporocyte callose wall causes male sterility in transgenic tobacco. Plant Cell 4: 759-771

Xu J, Ding Z, Vizcay-Barrena G, Shi J, Liang W, Yuan Z, Werck-Reichhart D, Schreiber L, Wilson ZA, Zhang D (2014) ABORTED MICROSPORES Acts as a Master Regulator of Pollen Wall Formation in *Arabidopsis*. The Plant Cell 26: 1544-1556

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

Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou J-P, Steinmetz A, Shen W-H (2008) Di-and tri-but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. Molecular and cellular biology 28: 1348-1360

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

Zeng L, Zhang Q, Li S, Plotnikov AN, Walsh MJ, Zhou MM (2010) Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. Nature 466: 258-262

Zhang W, Sun Y, Timofejeva L, Chen C, Grossniklaus U, Ma H (2006) Regulation of *Arabidopsis* tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. Development 133: 3085-3095

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

Zhang ZB, Zhu J, Gao JF, Wang C, Li H, Zhang HQ, Zhang S, Wang DM, Wang QX (2007) Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. The Plant Journal 52: 528-538

Zhao D-Z, Wang G-F, Speal B, Ma H (2002) The EXCESS MICROSPOROCYTES1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the *Arabidopsis* anther. Genes & development 16: 2021-2031

Zhong R, Richardson EA, Ye ZH (2007) The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. The Plant Cell 19: 2776-2792

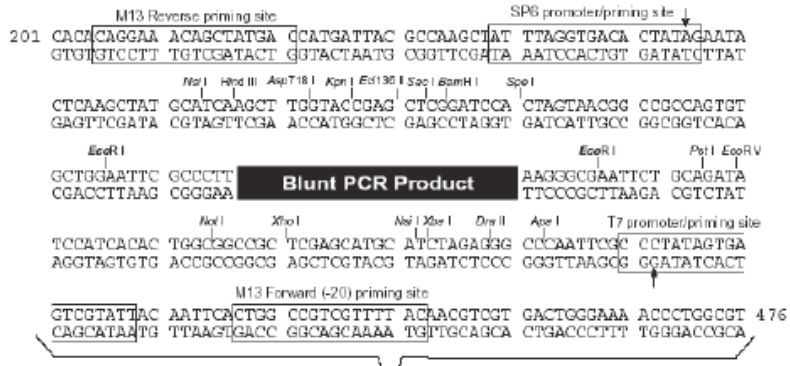
Zhu J, Chen H, Li H, Gao JF, Jiang H, Wang C, Guan YF, Yang ZN (2008) Defective in Tapetal Development and Function 1 is essential for anther development and tapetal function for microspore maturation in *Arabidopsis*. The Plant Journal 55: 266-277

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

Zhu J, Zhang G, Chang Y, Li X, Yang J, Huang X, Yu Q, Chen H, Wu T, Yang Z (2010) AtMYB103 is a crucial regulator of several pathways affecting *Arabidopsis* anther development. Science China Life Sciences 53: 1112-1122

APPENDICES

APPENDIX I PCR-BLUNT II-TOPO VECTOR (INVITROGEN)



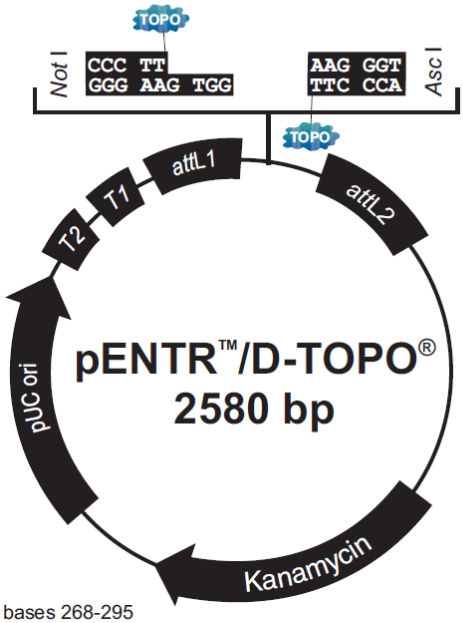
Comments for pCR®-Blunt II-TOPO®
3519 nucleotides

lac promoter/operator region: bases 95-216
M13 Reverse priming site: bases 205-221
LacZ-alpha ORF: bases 217-576
SP6 promoter priming site: bases 239-256
Multiple Cloning Site: bases 269-399
TOPO®-Cloning site: bases 336-337
T7 promoter priming site: bases 406-425
M13 (-20) Forward priming site: bases 433-448
Fusion joint: bases 577-585
ccdB lethal gene ORF: bases 586-888
kan gene: bases 1099-2031
kan promoter: bases 1099-1236
Kanamycin resistance gene ORF: bases 1237-2031
Zeocin resistance ORF: bases 2238-2612
pUC origin: bases 2724-3397

APPENDIX II PENTRTM/D-TOPO (INVITROGEN)

pENTR™/D-TOPO® Map

The figure below shows the features of pENTR™/D-TOPO® vector. **The complete sequence of pENTR™/D-TOPO® is available for downloading from www.invitrogen.com)or by contacting Technical Service (see page 35).**

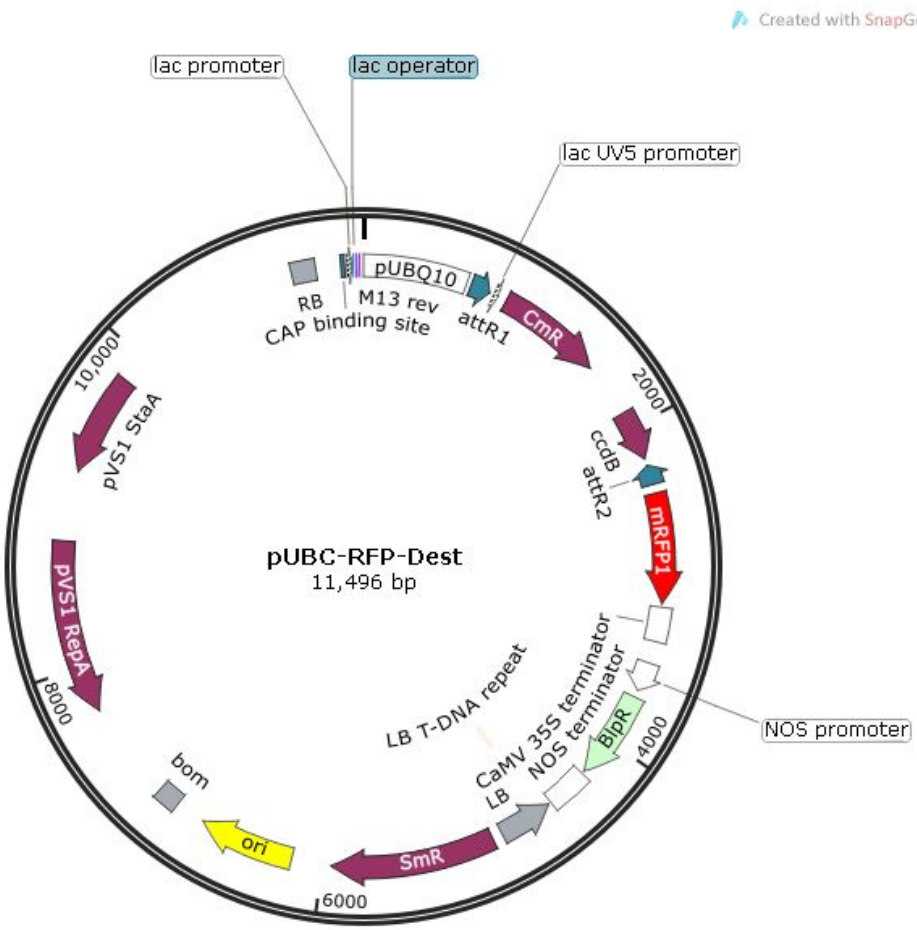


Comments for pENTR™/D-TOPO®
2580 nucleotides

*rrn*B T2 transcription termination sequence: bases 268-295
*rrn*B T1 transcription termination sequence: bases 427-470
M13 forward (-20) priming site: bases 537-552
*att*L1: bases 569-668 (c)
TOPO® recognition site 1: bases 680-684
Overhang: bases 685-688
TOPO® recognition site 2: bases 689-693
*att*L2: bases 705-804
T7 Promoter/priming site: bases 821-840 (c)
M13 reverse priming site: bases 845-861
Kanamycin resistance gene: bases 974-1783
pUC origin: bases 1904-2577

(c) = complementary sequence

APPENDIX III PUBC-RFP-DEST CONSTRUCT

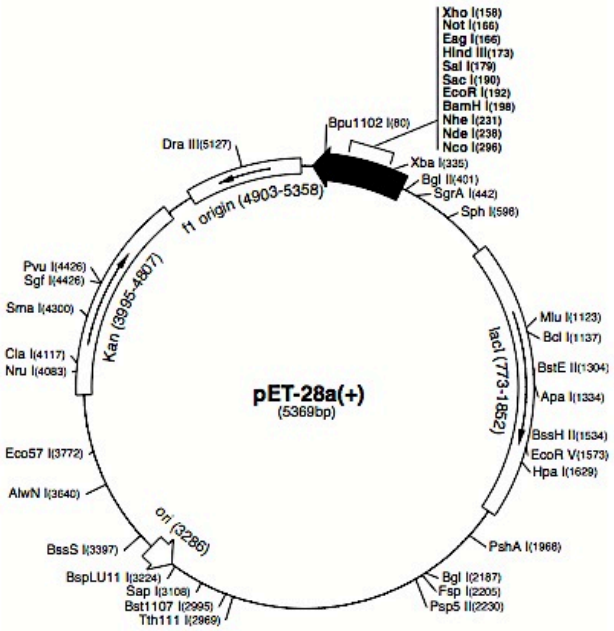


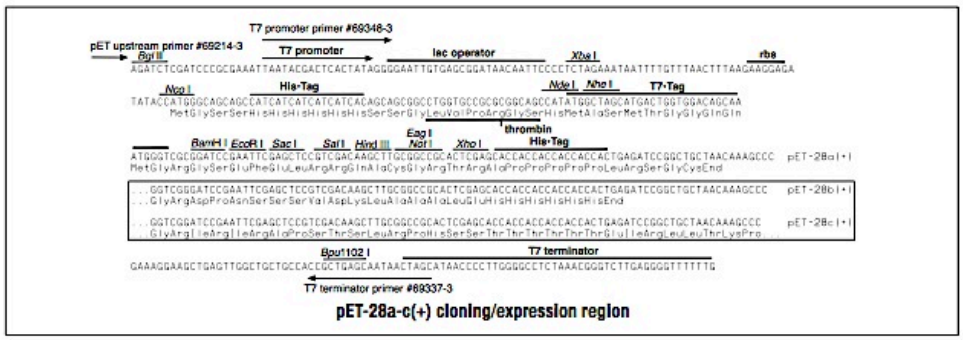
APPENDIX IV PETS VECTORS (NOVAGEN)

a. PET28a

pET-28a(+) sequence landmarks	
T7 promoter	370-386
T7 transcription start	369
His•Tag coding sequence	270-287
T7•Tag coding sequence	207-239
Multiple cloning sites	
(*Bam*H I - *Xho* I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	773-1852
pBR322 origin	3286
Kan coding sequence	3995-4807
f1 origin	4903-5358

The maps for pET-28b(+) and pET-28c(+) are the same as pET-28a(+) (shown) with the following exceptions: pET-28b(+) is a 5368bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-28c(+) is a 5367bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198.

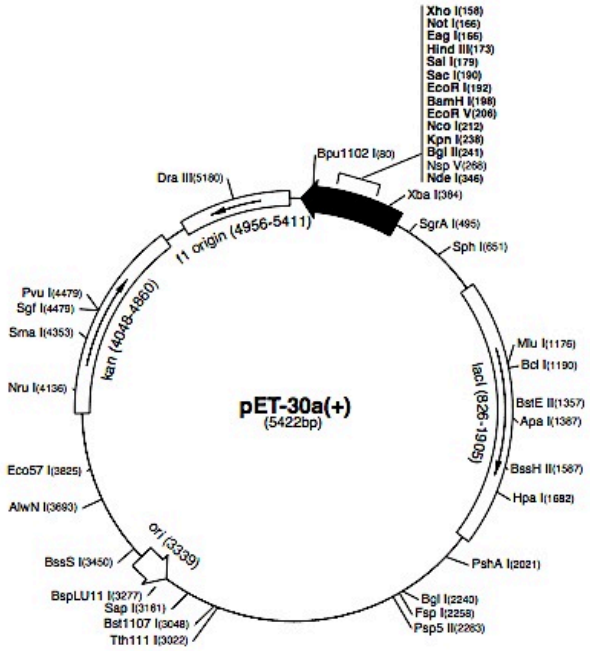


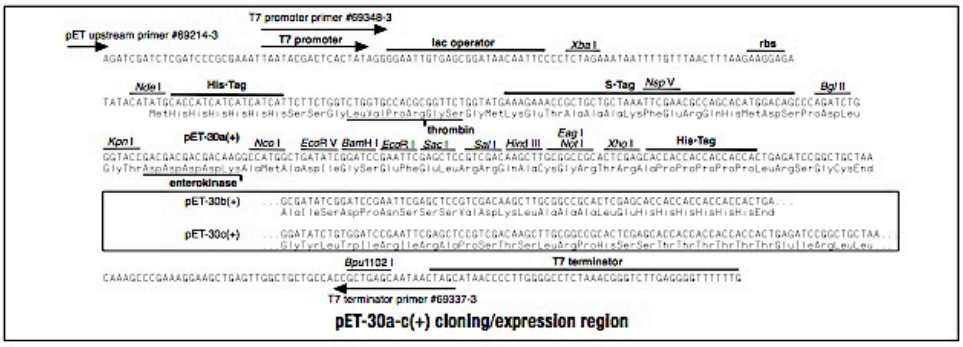


b. PET30a

pET-30a(+) sequence landmarks	
T7 promoter	419-435
T7 transcription start	418
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	826-1905
pBR322 origin	3339
Kan coding sequence	4048-4860
f1 origin	4956-5411

The maps for pET-30b(+) and pET-30c(+) are the same as pET-30a(+) (shown) with the following exceptions: pET-30b(+) is a 5421bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-30c(+) is a 5423bp plasmid; add 1bp to each site beyond *Bam*H I at 198.

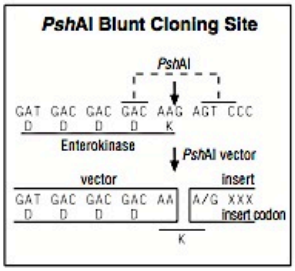


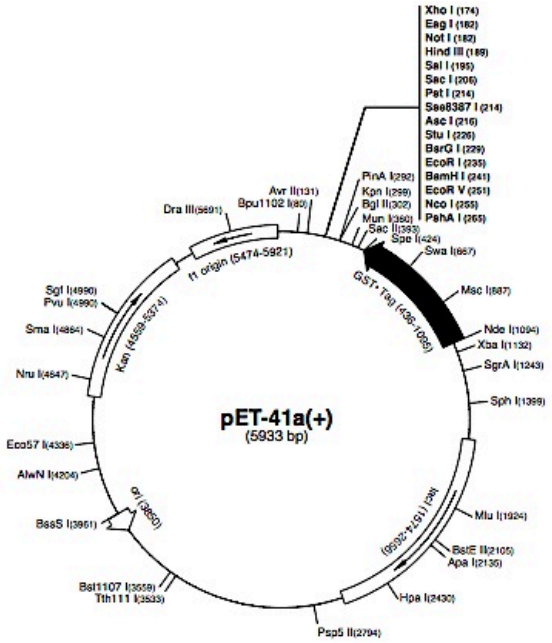


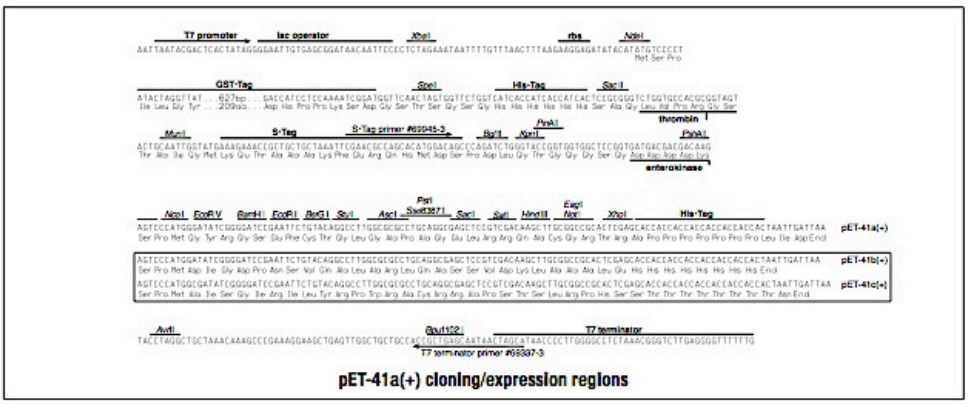
pET-30a-c(+) cloning/expression region

c. PET41a

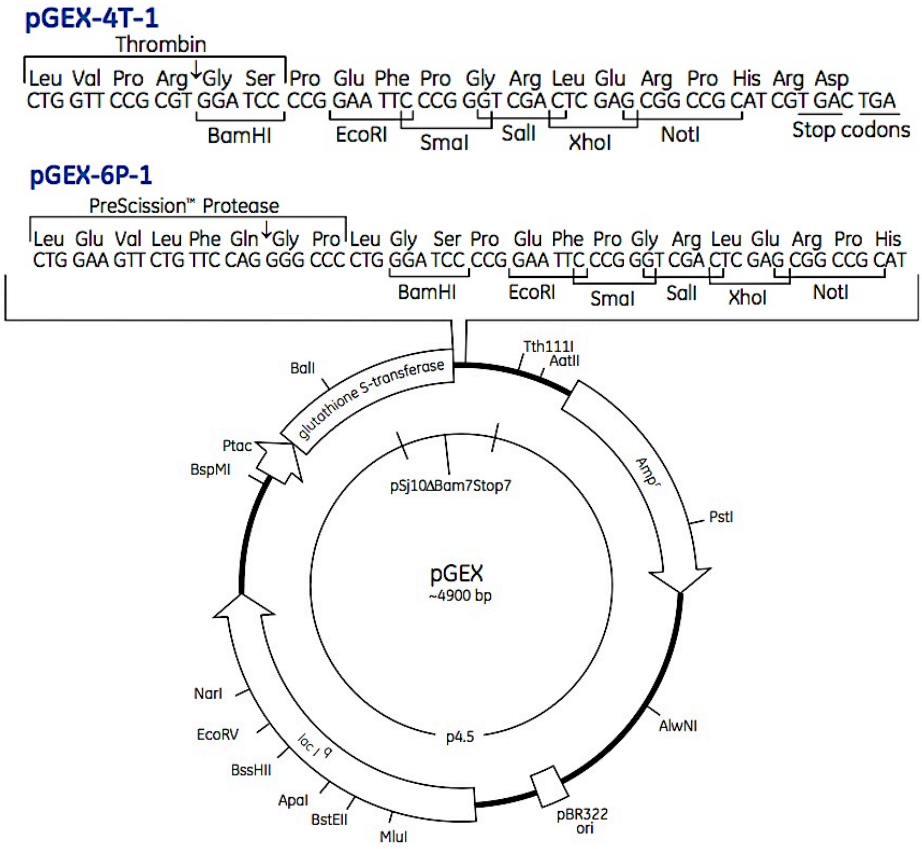
pET-41a(+) sequence landmarks	
T7 promoter	1167–1183
T7 transcription start	1166
GST•Tag coding sequence	436–1095
His•Tag coding sequence	397–414
S•Tag coding sequence	310–354
Multiple cloning sites (*Psh*AI-*Xho*I)	174–265
His•Tag coding sequence	150–173
T7 terminator	26–72
*lac*I coding sequence	1574–2656
pBR322 origin	3850
Kan coding sequence	4559–5374
F1 origin	5474–5921







APPENDIX V PGEX VECTORS (GE LIFESCIENCE)



APPENDIX VI PGWB5 VECTOR

pGWB5 : [(35S promoter, C-sGFP) (--35Spromoter–R1-ccdB-R2–sGFP–)]

(AAGCTT) -- 35S promoter --//

(TCT AA)T CAA ACA AGT TTG TAC AAA AAA --(CmR, ccdB)–TTC TTG TAC AAA GTG GTT CGA TCT AGA GGA TCC ATG --(*GFP* sequence) --(GAGCTC)

(1) As a result of a mistake during the fill-in reaction while constructing pGWB5, the XbaI site TCTAGA was changed to TCTAA. However, I think it would not affect the result of your experiments.
(2) The initiation codon ATG (double-underlined) is from the *GFP* coding sequence.

After LR reaction, the following sequence will be obtained.

35S promoter –(TCT AA)T CAA ACA AGT TTG TAC AAA AAA GCA GGC TNN (*Your Clone*)
NAC CCA GCT TTC TTG TAC AAA GTG GTT CGA TCT AGA GGA TCC ATG --
-X- -P- -A- -F- -L- -Y- -K- -V- -V- -R- -S- -R- -G- -S- -M-
(*GFP sequence*) -- (GAGCTC)

(1) Include an initiation codon ATG in *Your Clone*. Protein fusion with the C-terminal GFP will be linked by the peptide sequence derived from the *att*B2 and linker region as shown in the above sequence.
(2) N indicates the nucleotide (and X the amino acid) that is needed to be included in *Your Clone*.
(3) The DNA sequence downstream of *GFP* is the same ase in pGWB4.
(4) Hyphenations flanking the amino acid residues are used for proper spacing.

APPENDIX VII PGWB14 VECTOR

pGWB7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23

(no promoter or 35S promoter, C-tag)

No promoter --- 7, 10, 13, 16, 19, 22

35S promoter --- 8, 11, 14, 17, 20, 23

```
(AAGCTT) -- (no promoter or 35S promoter) --//
(TCTAGA)GTTA TCA ACA AGT TTG TAC AAA AAA --(CmR,ccdB)--TTC TTG TAC AAA GTG
GTT GAT AAC AGC tag GCT TA(G AGCTC)
```

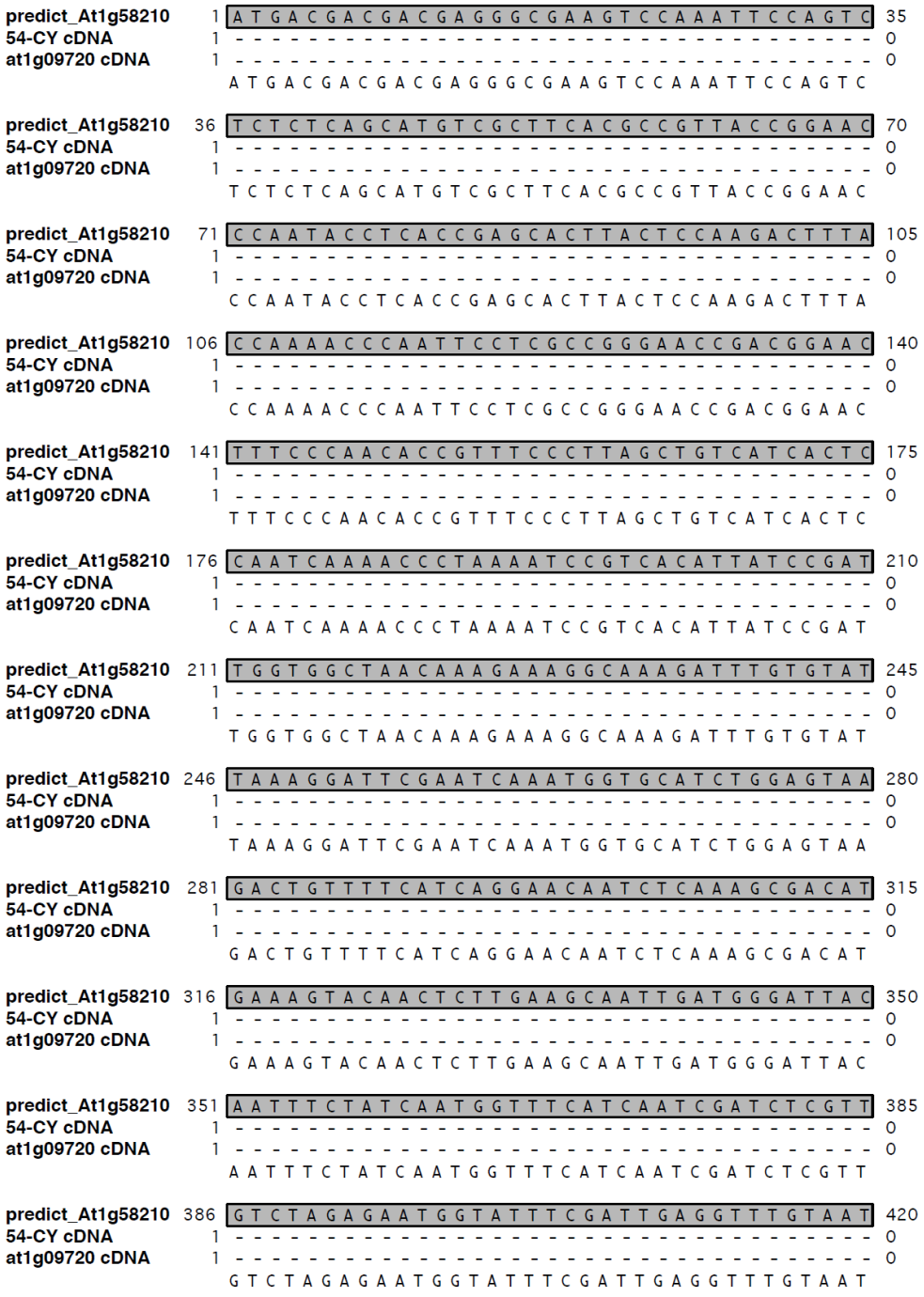
(1) The *Hin*dIII (AAGCTT), *Xba*I (TCTAGA) and *Sac*I (GAGCTC) sites are shown.
(2) The *att*R1-CmR-*ccd*B-*att*R2 cassette is underlined.
(3) The sequence upstream of the *Xba*I site is the same as in pBI101 or pBI121.
(4) The sequence of each tag is indicated in the "Tags" section.

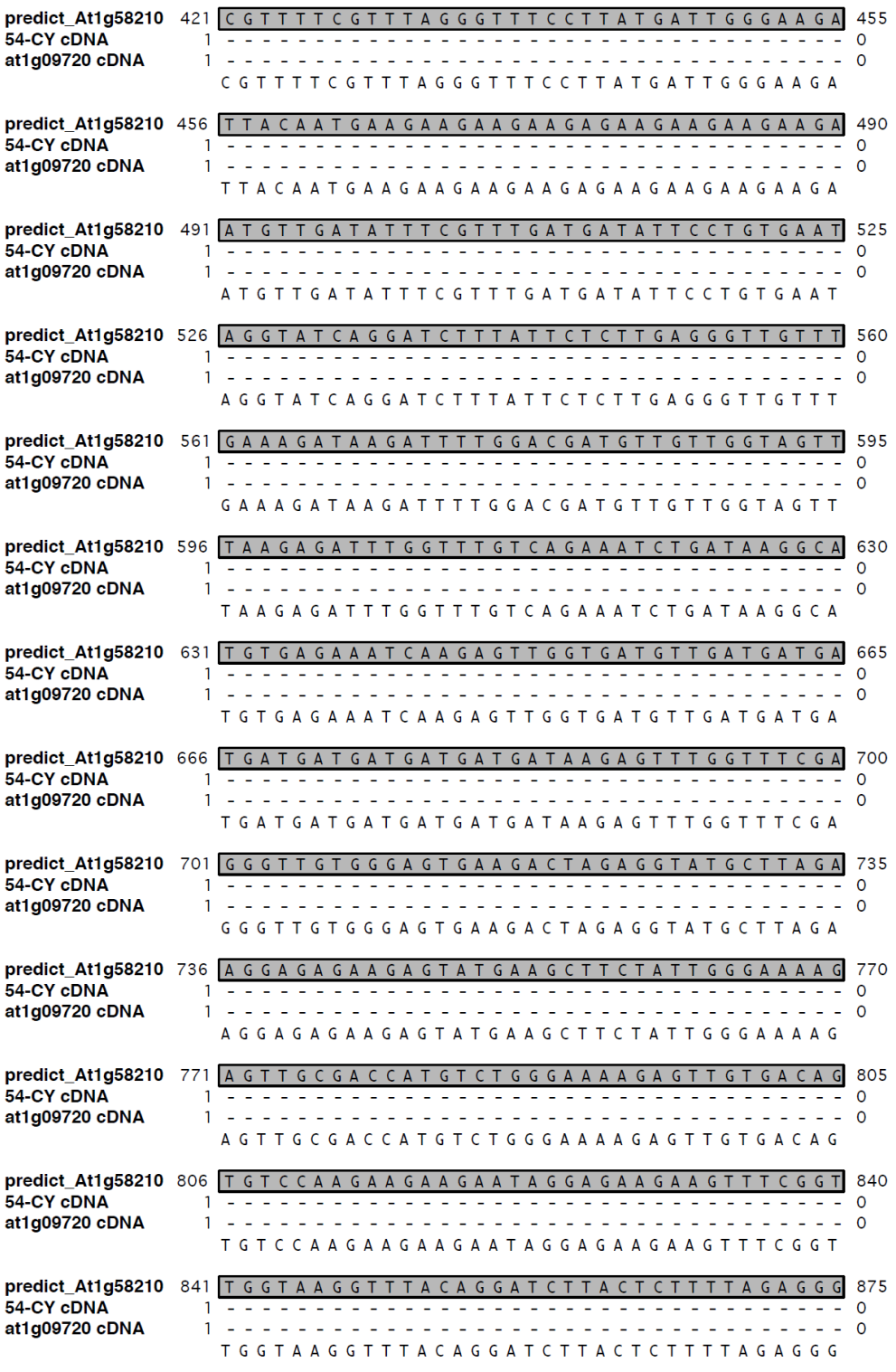
After LR reaction, the following sequence will be obtained.

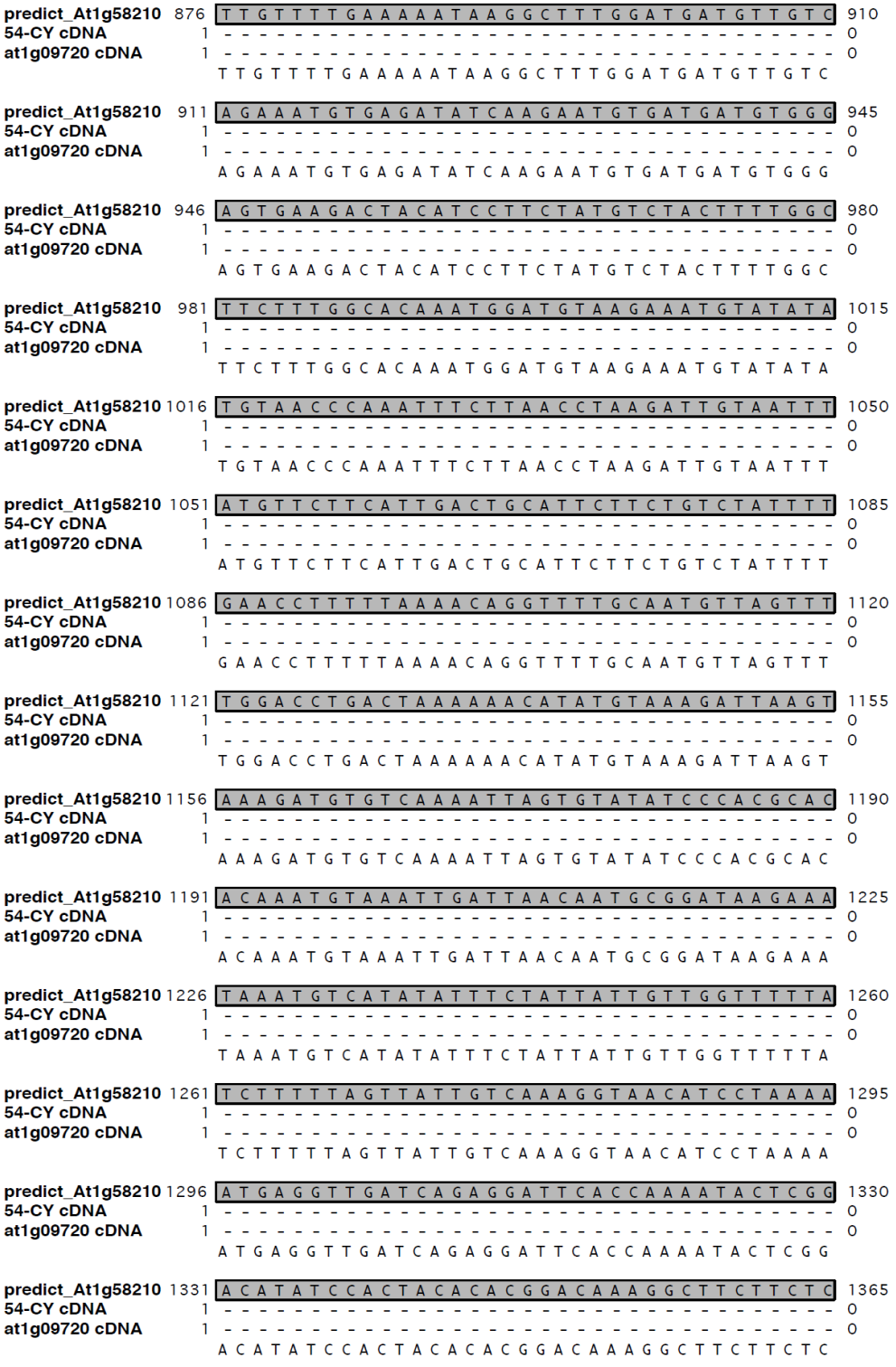
```
(AAGCTT) -- (no promoter or 35S promoter) --//
(TCTAGA)GTTA TCA ACA AGT TTG TAC AAA AAA GCA GGC TNN (Your Clone) NAC CCA
                                                    M -------   -X- -P-
GCT TTC TTG TAC AAA GTG GTT GAT AAC AGC tag GCT TA(G AGCTC)
-A- -F- -L- -Y- -K- -V- -V- -D- -N- -S- tag -A-   *
```

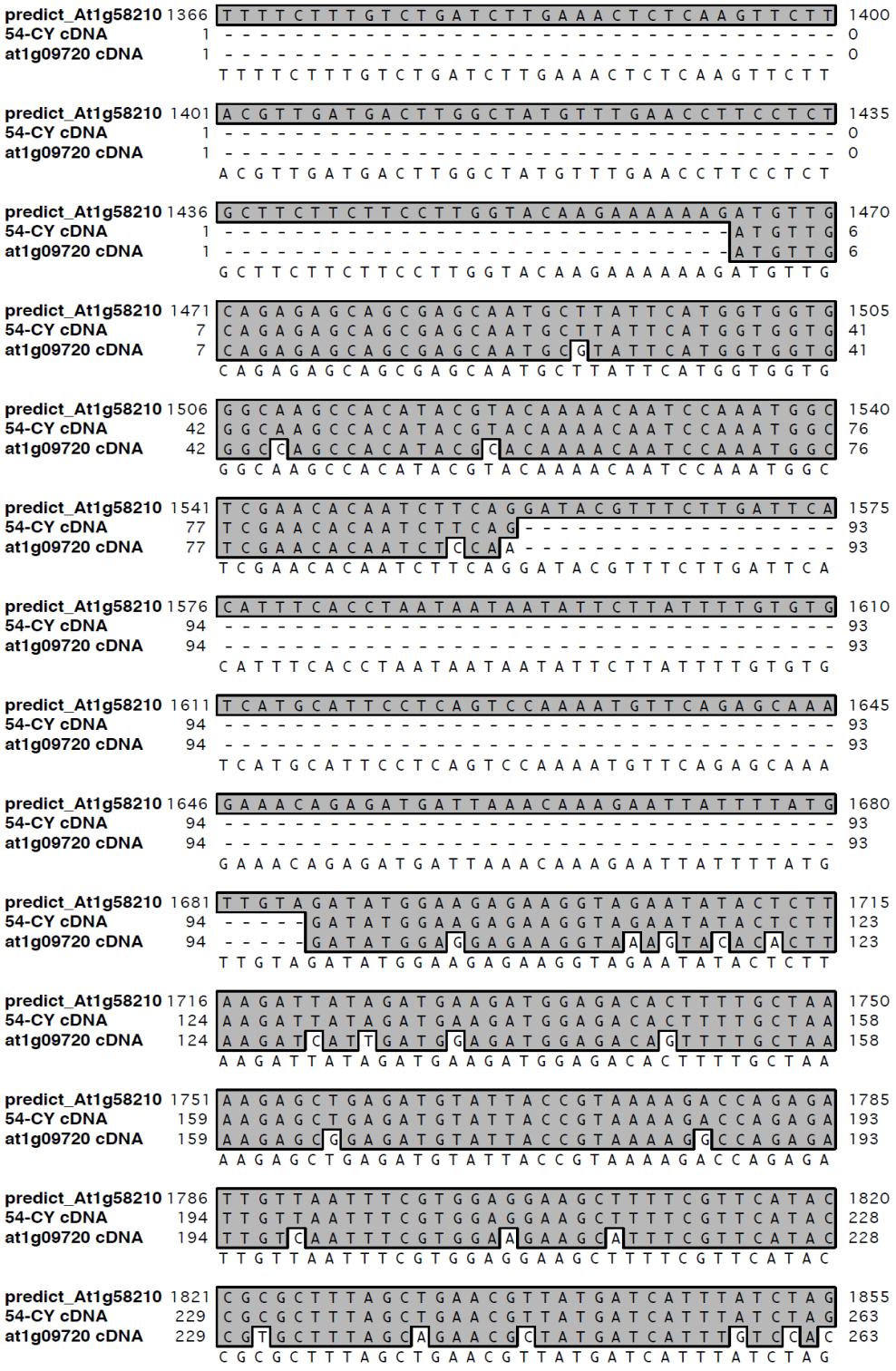
(1) Include an initiation codon ATG in *Your Clone*. Protein fusion with the C-terminal tag will be linked by the peptide sequence derived from the *att*B2 and linker region as shown in the above sequence.
(2) Translation will stop at the termination codon TAG (double-underlined) after the tag.

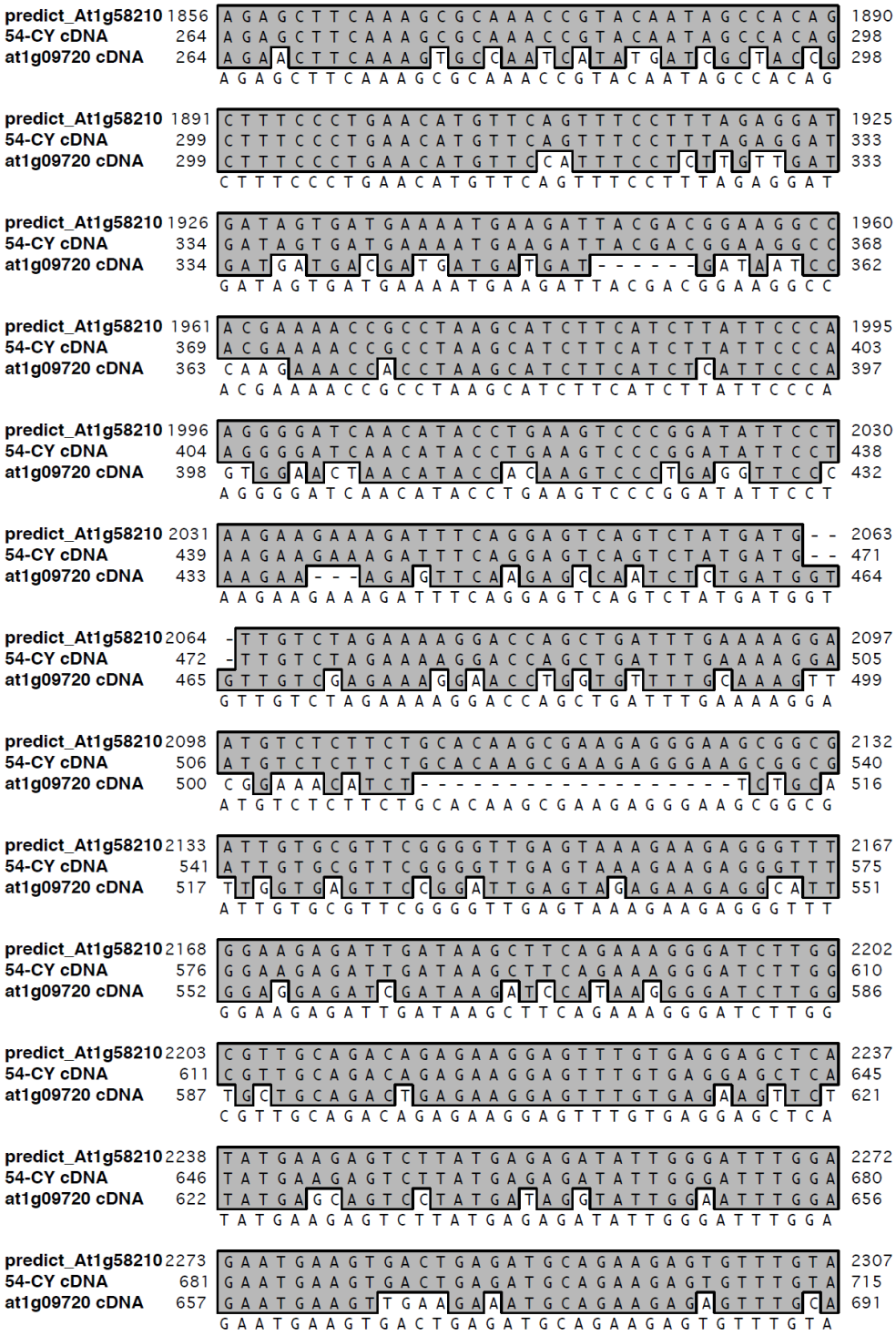
APPENDIX VIII CDNA ALIGNMENT OF AT1G58210 AND AT1G09720

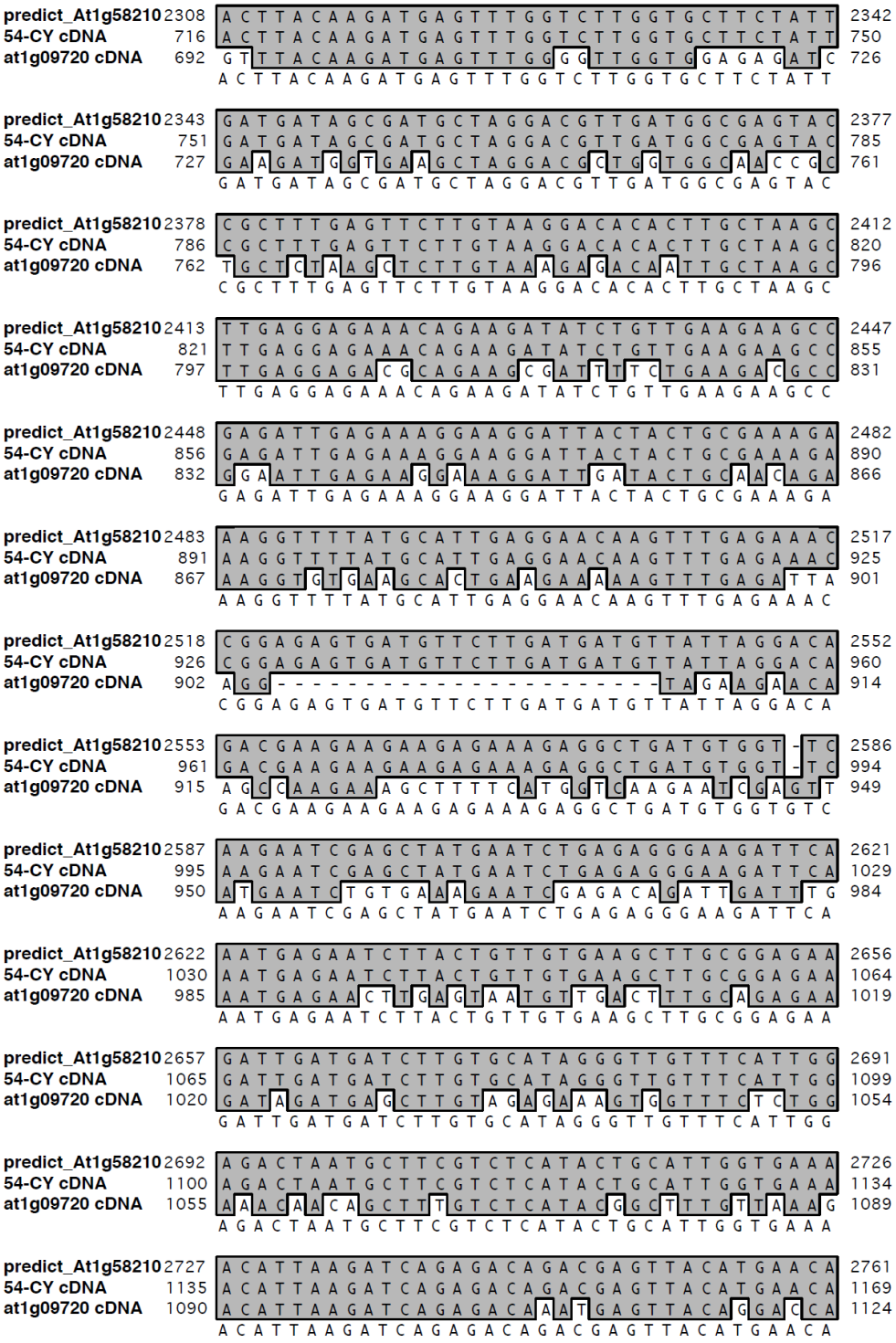


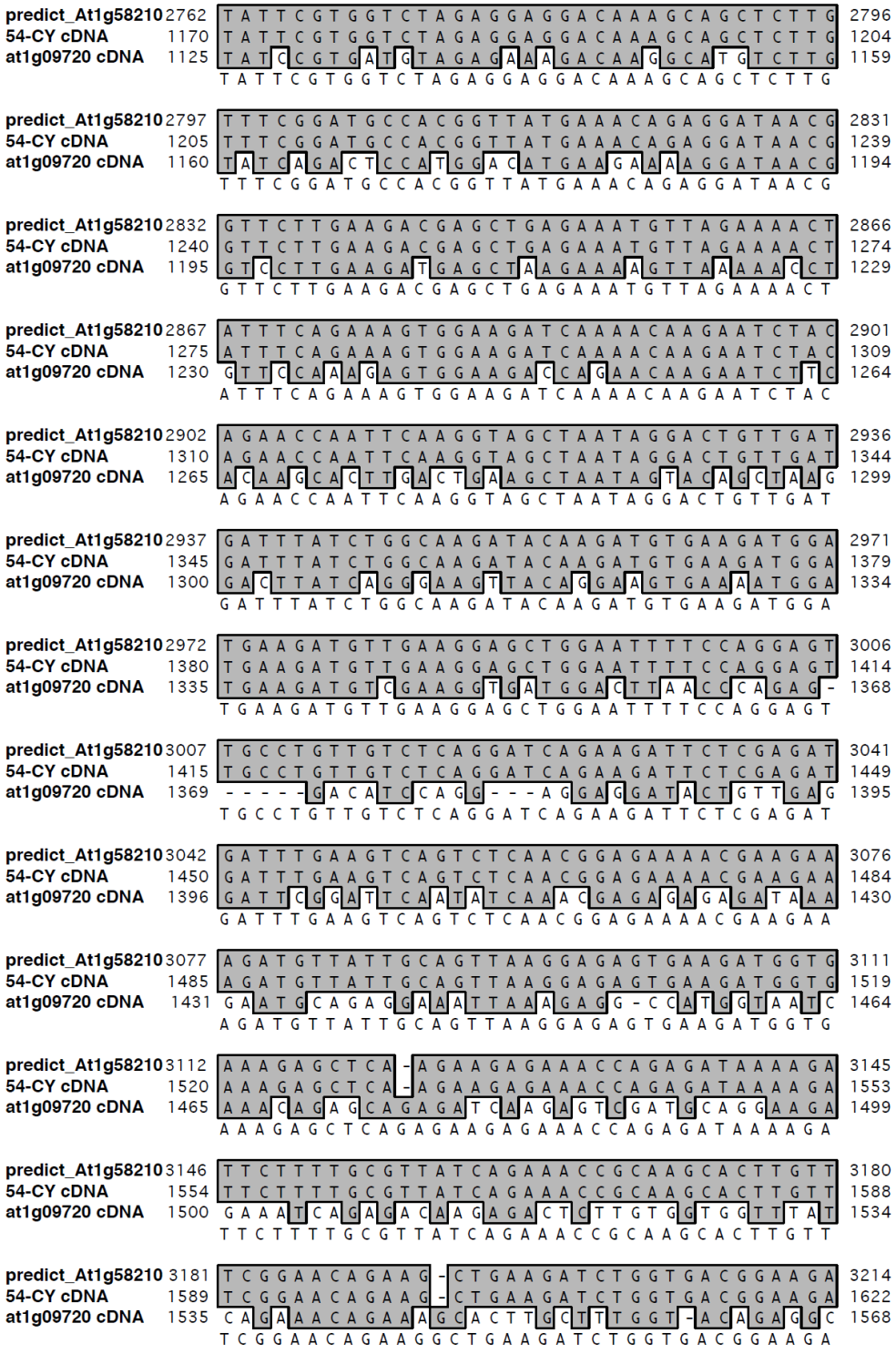

predict_At1g58210 421 CGTTTTCGTTTAGGGTTTCCTTATGATTGGGAAGA 455
54-CY cDNA 1 0
at1g09720 cDNA 1 0
CGTTTTCGTTTAGGGTTTCCTTATGATTGGGAAGA
predict_At1g58210 456 TTACAATGAAGAAGAAGAAGAGAAGAAGAAGAAGA 490
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TTACAATGAAGAAGAAGAAGAGAAGAAGAAGAAGA
predict_At1g58210 491 ATGTTGATATTTCGTTTGATGATATTCCTGTGAAT 525
54-CY cDNA 1 0
at1g09720 cDNA 1 0
ATGTTGATATTTCGTTTGATGATATTCCTGTGAAT
predict_At1g58210 526 AGGTATCAGGATCTTTATTCTCTTGAGGGTTGTTT 560
54-CY cDNA 1 0
at1g09720 cDNA 1 0
AGGTATCAGGATCTTTATTCTCTTGAGGGTTGTTT
predict_At1g58210 561 GAAAGATAAGATTTTGGACGATGTTGTTGGTAGTT 595
54-CY cDNA 1 0
at1g09720 cDNA 1 0
GAAAGATAAGATTTTGGACGATGTTGTTGGTAGTT
predict_At1g58210 596 TAAGAGATTTGGTTTGTCAGAAATCTGATAAGGCA 630
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TAAGAGATTTGGTTTGTCAGAAATCTGATAAGGCA
predict_At1g58210 631 TGTGAGAAATCAAGAGTTGGTGATGTTGATGATGA 665
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TGTGAGAAATCAAGAGTTGGTGATGTTGATGATGA
predict_At1g58210 666 TGATGATGATGATGATGATAAGAGTTTGGTTTCGA 700
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TGATGATGATGATGATGATAAGAGTTTGGTTTCGA
predict_At1g58210 701 GGGTTGTGGGAGTGAAGACTAGAGGTATGCTTAGA 735
54-CY cDNA 1 0
at1g09720 cDNA 1 0
GGGTTGTGGGAGTGAAGACTAGAGGTATGCTTAGA
predict_At1g58210 736 AGGAGAGAAGAGTATGAAGCTTCTATTGGGAAAAG 770
54-CY cDNA 1 0
at1g09720 cDNA 1 0
AGGAGAGAAGAGTATGAAGCTTCTATTGGGAAAAG
predict_At1g58210 771 AGTTGCGACCATGTCTGGGAAAAGAGTTGTGACAG 805
54-CY cDNA 1 0
at1g09720 cDNA 1 0
AGTTGCGACCATGTCTGGGAAAAGAGTTGTGACAG
predict_At1g58210 806 TGTCCAAGAAGAAGAATAGGAGAAGAAGTTTCGGT 840
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TGTCCAAGAAGAAGAATAGGAGAAGAAGTTTCGGT
predict_At1g58210 841 TGGTAAGGTTTACAGGATCTTACTCTTTTAGAGGG 875
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TGGTAAGGTTTACAGGATCTTACTCTTTTAGAGGG

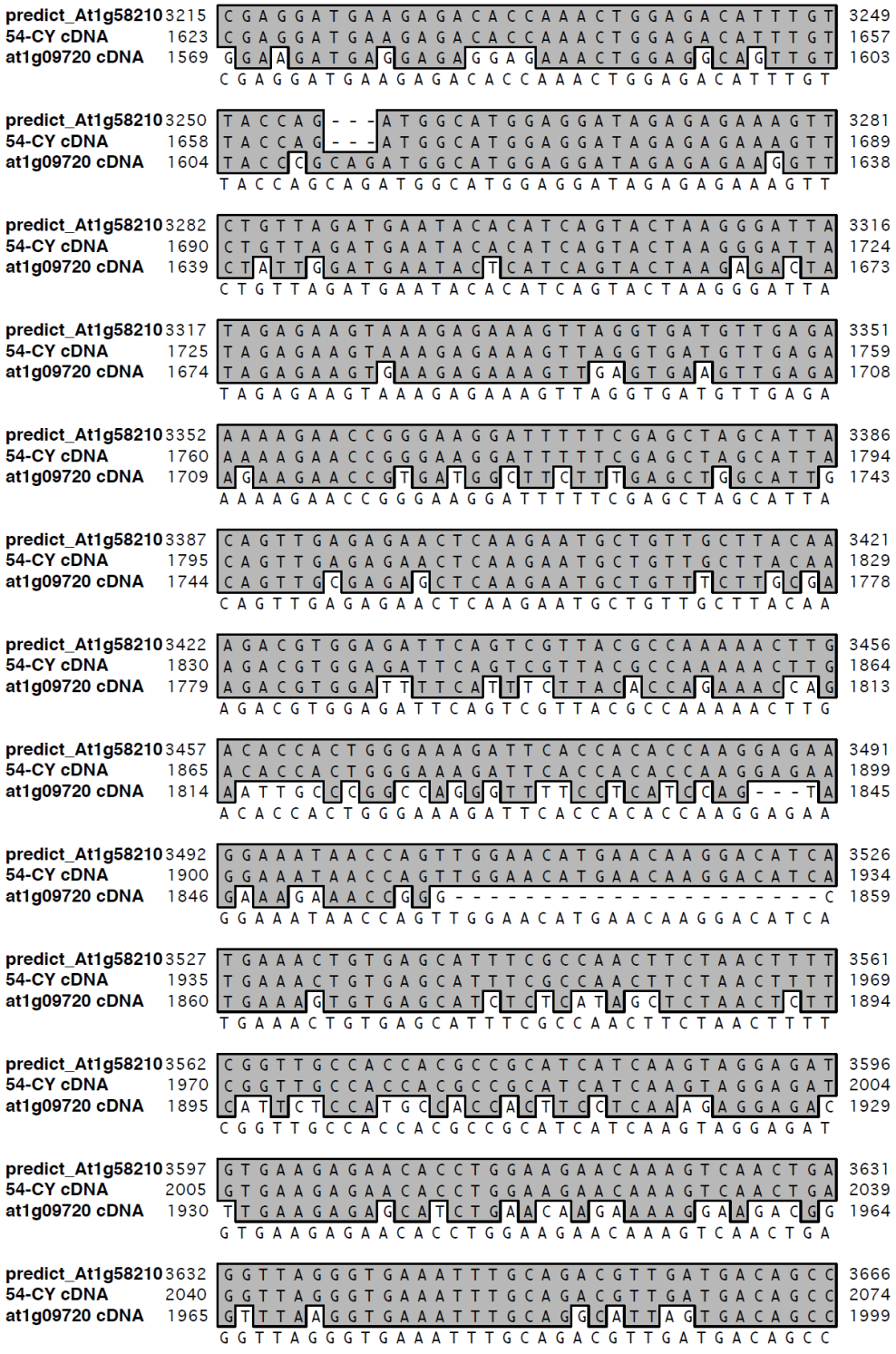

predict_At1g58210 876 TTGTTTTGAAAAATAAGGCTTTGGATGATGTTGTC 910
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TTGTTTTGAAAAATAAGGCTTTGGATGATGTTGTC
predict_At1g58210 911 AGAAATGTGAGATATCAAGAATGTGATGATGTGGG 945
54-CY cDNA 1 0
at1g09720 cDNA 1 0
AGAAATGTGAGATATCAAGAATGTGATGATGTGGG
predict_At1g58210 946 AGTGAAGACTACATCCTTCTATGTCTACTTTTGGC 980
54-CY cDNA 1 0
at1g09720 cDNA 1 0
AGTGAAGACTACATCCTTCTATGTCTACTTTTGGC
predict_At1g58210 981 TTCTTTGGCACAAATGGATGTAAGAAATGTATATA 1015
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TTCTTTGGCACAAATGGATGTAAGAAATGTATATA
predict_At1g58210 1016 TGTAACCCAAATTTCTTAACCTAAGATTGTAATTT 1050
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TGTAACCCAAATTTCTTAACCTAAGATTGTAATTT
predict_At1g58210 1051 ATGTTCTTCATTGACTGCATTCTTCTGTCTATTTT 1085
54-CY cDNA 1 0
at1g09720 cDNA 1 0
ATGTTCTTCATTGACTGCATTCTTCTGTCTATTTT
predict_At1g58210 1086 GAACCTTTTTAAAACAGGTTTTGCAATGTTAGTTT 1120
54-CY cDNA 1 0
at1g09720 cDNA 1 0
GAACCTTTTTAAAACAGGTTTTGCAATGTTAGTTT
predict_At1g58210 1121 TGGACCTGACTAAAAAACATATGTAAAGATTAAGT 1155
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TGGACCTGACTAAAAAACATATGTAAAGATTAAGT
predict_At1g58210 1156 AAAGATGTGTCAAAATTAGTGTATATCCCACGCAC 1190
54-CY cDNA 1 0
at1g09720 cDNA 1 0
AAAGATGTGTCAAAATTAGTGTATATCCCACGCAC
predict_At1g58210 1191 ACAAATGTAAATTGATTAACAATGCGGATAAGAAA 1225
54-CY cDNA 1 0
at1g09720 cDNA 1 0
ACAAATGTAAATTGATTAACAATGCGGATAAGAAA
predict_At1g58210 1226 TAAATGTCATATATTTCTATTATTGTTGGTTTTTA 1260
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TAAATGTCATATATTTCTATTATTGTTGGTTTTTA
predict_At1g58210 1261 TCTTTTTAGTTATTGTCAAAGGTAACATCCTAAAA 1295
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TCTTTTTAGTTATTGTCAAAGGTAACATCCTAAAA
predict_At1g58210 1296 ATGAGGTTGATCAGAGGATTCACCAAAATACTCGG 1330
54-CY cDNA 1 0
at1g09720 cDNA 1 0
ATGAGGTTGATCAGAGGATTCACCAAAATACTCGG
predict_At1g58210 1331 ACATATCCACTACACACGGACAAAGGCTTCTTCTC 1365
54-CY cDNA 1 0
at1g09720 cDNA 1 0
ACATATCCACTACACACGGACAAAGGCTTCTTCTC

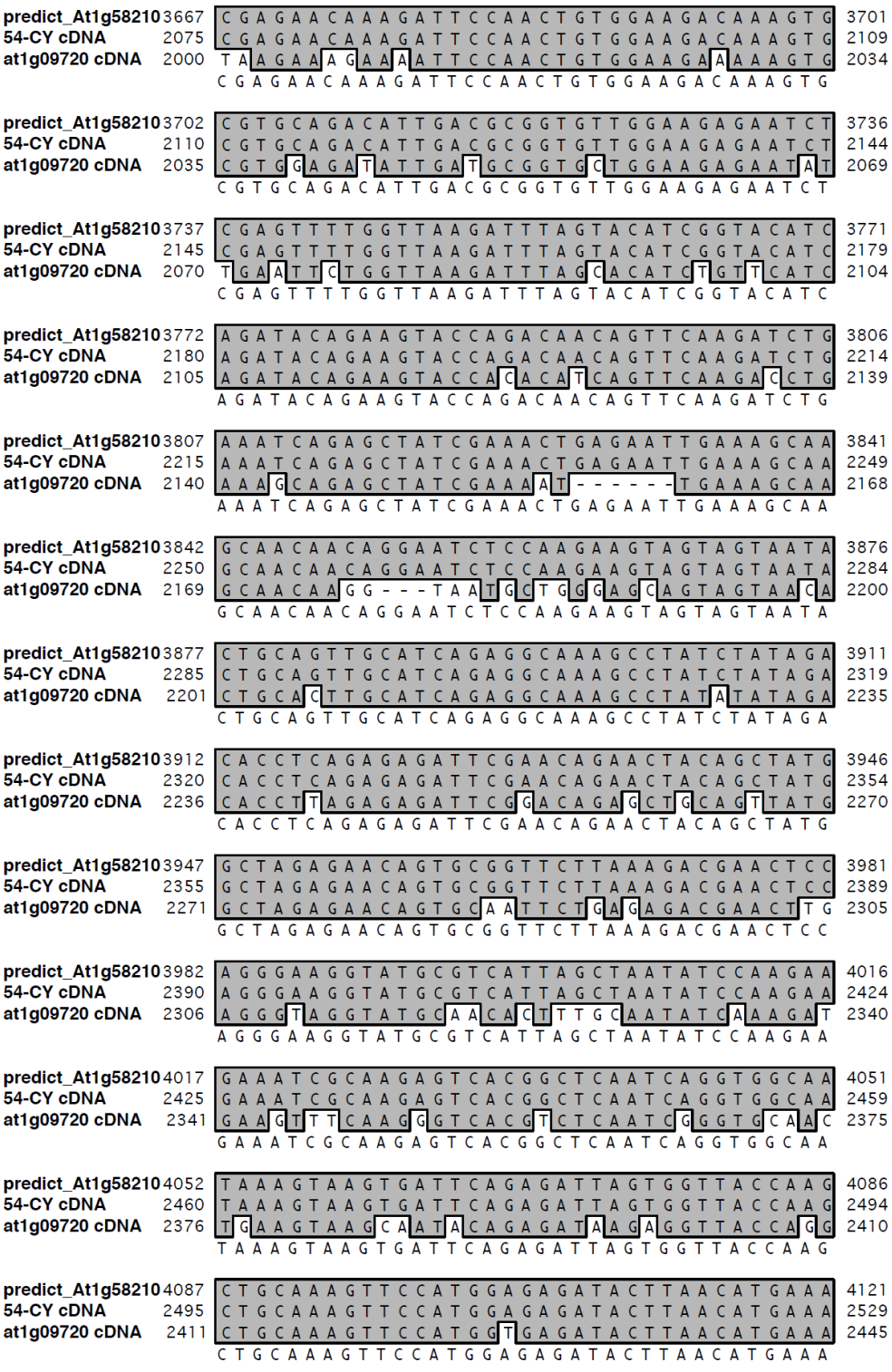

predict_At1g58210 1366 TTTTCTTTGTCTGATCTTGAAACTCTCAAGTTCTT 1400
54-CY cDNA 1 0
at1g09720 cDNA 1 0
TTTTCTTTGTCTGATCTTGAAACTCTCAAGTTCTT
predict_At1g58210 1401 ACGTTGATGACTTGGCTATGTTTGAACCTTCCTCT 1435
54-CY cDNA 1 0
at1g09720 cDNA 1 0
ACGTTGATGACTTGGCTATGTTTGAACCTTCCTCT
predict_At1g58210 1436 GCTTCTTCTTCCTTGGTACAAGAAAAAAGATGTTG 1470
54-CY cDNA 1 ATGTTG 6
at1g09720 cDNA 1 ATGTTG 6
GCTTCTTCTTCCTTGGTACAAGAAAAAAGATGTTG
predict_At1g58210 1471 CAGAGAGCAGCGAGCAATGCTTATTCATGGTGGTG 1505
54-CY cDNA 7 CAGAGAGCAGCGAGCAATGCTTATTCATGGTGGTG 41
at1g09720 cDNA 7 CAGAGAGCAGCGAGCAATGCGTATTCATGGTGGTG 41
CAGAGAGCAGCGAGCAATGCTTATTCATGGTGGTG
predict_At1g58210 1506 GGCAAGCCACATACGTACAAAACAATCCAAATGGC 1540
54-CY cDNA 42 GGCAAGCCACATACGTACAAAACAATCCAAATGGC 76
at1g09720 cDNA 42 GGCCAGCCACATACGCACAAAACAATCCAAATGGC 76
GGCAAGCCACATACGTACAAAACAATCCAAATGGC
predict_At1g58210 1541 TCGAACACAATCTTCAGGATACGTTTCTTGATTCA 1575
54-CY cDNA 77 TCGAACACAATCTTCAG 93
at1g09720 cDNA 77 TCGAACACAATCTCCAA 93
TCGAACACAATCTTCAGGATACGTTTCTTGATTCA
predict_At1g58210 1576 CATTTCACCTAATAATAATATTCTTATTTTGTGTG 1610
54-CY cDNA 94 93
at1g09720 cDNA 94 93
CATTTCACCTAATAATAATATTCTTATTTTGTGTG
predict_At1g58210 1611 TCATGCATTCCTCAGTCCAAAATGTTCAGAGCAAA 1645
54-CY cDNA 94 93
at1g09720 cDNA 94 93
TCATGCATTCCTCAGTCCAAAATGTTCAGAGCAAA
predict_At1g58210 1646 GAAACAGAGATGATTAAACAAAGAATTATTTTATG 1680
54-CY cDNA 94 93
at1g09720 cDNA 94 93
GAAACAGAGATGATTAAACAAAGAATTATTTTATG
predict_At1g58210 1681 TTGTAGATATGGAAGAGAAGGTAGAATATACTCTT 1715
54-CY cDNA 94 GATATGGAAGAGAAGGTAGAATATACTCTT 123
at1g09720 cDNA 94 GATATGGAGGAGAAGGTAAAGTACACACTT 123
TTGTAGATATGGAAGAGAAGGTAGAATATACTCTT
predict_At1g58210 1716 AAGATTATAGATGAAGATGGAGACACTTTTGCTAA 1750
54-CY cDNA 124 AAGATTATAGATGAAGATGGAGACACTTTTGCTAA 158
at1g09720 cDNA 124 AAGATCATTGATGGAGATGGAGACAGTTTTGCTAA 158
AAGATTATAGATGAAGATGGAGACACTTTTGCTAA
predict_At1g58210 1751 AAGAGCTGAGATGTATTACCGTAAAAGACCAGAGA 1785
54-CY cDNA 159 AAGAGCTGAGATGTATTACCGTAAAAGACCAGAGA 193
at1g09720 cDNA 159 AAGAGCGGAGATGTATTACCGTAAAAGGCCAGAGA 193
AAGAGCTGAGATGTATTACCGTAAAAGACCAGAGA
predict_At1g58210 1786 TTGTTAATTTCGTGGAGGAAGCTTTTCGTTCATAC 1820
54-CY cDNA 194 TTGTTAATTTCGTGGAGGAAGCTTTTCGTTCATAC 228
at1g09720 cDNA 194 TTGTCAATTTCGTGGAAGAAGCATTTCGTTCATAC 228
TTGTTAATTTCGTGGAGGAAGCTTTTCGTTCATAC
predict_At1g58210 1821 CGCGCTTTAGCTGAACGTTATGATCATTTATCTAG 1855
54-CY cDNA 229 CGCGCTTTAGCTGAACGTTATGATCATTTATCTAG 263
at1g09720 cDNA 229 CGTGCTTTAGCAGAACGCTATGATCATTTGTCCAC 263
CGCGCTTTAGCTGAACGTTATGATCATTTATCTAG

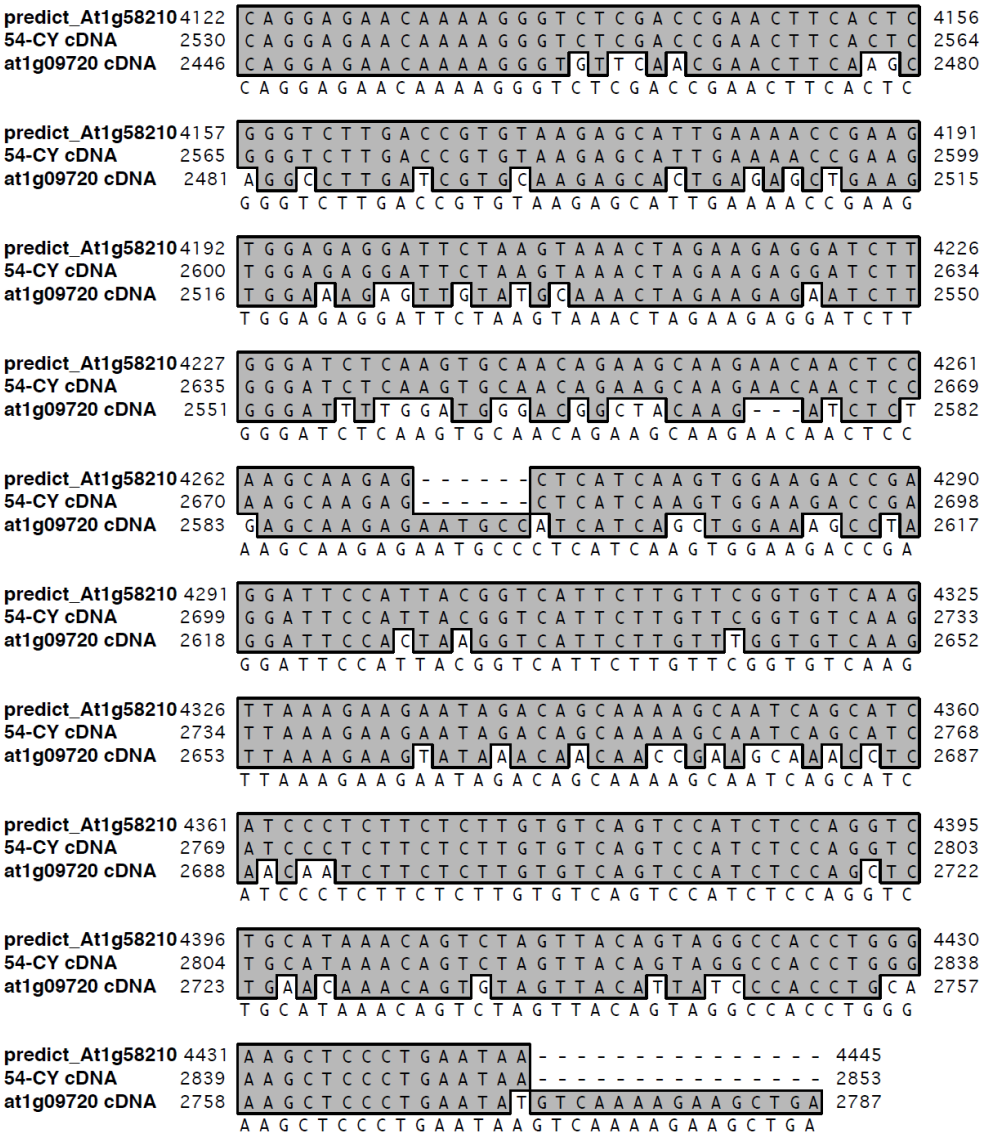

predict_At1g58210 1856 AGAGCTTCAAAGCGCAAACCGTACAATAGCCACAG 1890
54-CY cDNA 264 AGAGCTTCAAAGCGCAAACCGTACAATAGCCACAG 298
at1g09720 cDNA 264 AGAACTTCAAAGTGCCAATCATATGATCGCTACCG 298
AGAGCTTCAAAGCGCAAACCGTACAATAGCCACAG
predict_At1g58210 1891 CTTTCCCTGAACATGTTCAGTTTCCTTTAGAGGAT 1925
54-CY cDNA 299 CTTTCCCTGAACATGTTCAGTTTCCTTTAGAGGAT 333
at1g09720 cDNA 299 CTTTCCCTGAACATGTTCCATTTCCTCTTGTTGAT 333
CTTTCCCTGAACATGTTCAGTTTCCTTTAGAGGAT
predict_At1g58210 1926 GATAGTGATGAAAATGAAGATTACGACGGAAGGCC 1960
54-CY cDNA 334 GATAGTGATGAAAATGAAGATTACGACGGAAGGCC 368
at1g09720 cDNA 334 GATGATGACGATGATGATGAT------GATAATCC 362
GATAGTGATGAAAATGAAGATTACGACGGAAGGCC
predict_At1g58210 1961 ACGAAAACCGCCTAAGCATCTTCATCTTATTCCCA 1995
54-CY cDNA 369 ACGAAAACCGCCTAAGCATCTTCATCTTATTCCCA 403
at1g09720 cDNA 363 CAAGAAACCACCTAAGCATCTTCATCTCATTCCCA 397
ACGAAAACCGCCTAAGCATCTTCATCTTATTCCCA
predict_At1g58210 1996 AGGGGATCAACATACCTGAAGTCCCGGATATTCCT 2030
54-CY cDNA 404 AGGGGATCAACATACCTGAAGTCCCGGATATTCCT 438
at1g09720 cDNA 398 GTGGAACTAACATACCACAAGTCCCTGAGGTTCCC 432
AGGGGATCAACATACCTGAAGTCCCGGATATTCCT
predict_At1g58210 2031 AAGAAGAAAGATTTCAGGAGTCAGTCTATGATG-- 2063
54-CY cDNA 439 AAGAAGAAAGATTTCAGGAGTCAGTCTATGATG-- 471
at1g09720 cDNA 433 AAGAA---AGAGTTCAAGAGCCAATCTCTGATGGT 464
AAGAAGAAAGATTTCAGGAGTCAGTCTATGATGGT
predict_At1g58210 2064 -TTGTCTAGAAAAGGACCAGCTGATTTGAAAAGGA 2097
54-CY cDNA 472 -TTGTCTAGAAAAGGACCAGCTGATTTGAAAAGGA 505
at1g09720 cDNA 465 GTTGTCGAGAAAGGACCTGGTGTTTTGCAAAGTT 499
GTTGTCTAGAAAAGGACCAGCTGATTTGAAAAGGA
predict_At1g58210 2098 ATGTCTCTTCTGCACAAGCGAAGAGGGAAGCGGCG 2132
54-CY cDNA 506 ATGTCTCTTCTGCACAAGCGAAGAGGGAAGCGGCG 540
at1g09720 cDNA 500 CGGAAACATCT----------TCTGCA 516
ATGTCTCTTCTGCACAAGCGAAGAGGGAAGCGGCG
predict_At1g58210 2133 ATTGTGCGTTCGGGGTTGAGTAAAGAAGAGGGTTT 2167
54-CY cDNA 541 ATTGTGCGTTCGGGGTTGAGTAAAGAAGAGGGTTT 575
at1g09720 cDNA 517 TTGGTGAGTTCCGGATTGAGTAGAAGAGGCATT 551
ATTGTGCGTTCGGGGTTGAGTAAAGAAGAGGGTTT
predict_At1g58210 2168 GGAAGAGATTGATAAGCTTCAGAAAGGGATCTTGG 2202
54-CY cDNA 576 GGAAGAGATTGATAAGCTTCAGAAAGGGATCTTGG 610
at1g09720 cDNA 552 GGAGGAGATCGATAAGATCCATAAGGGGATCTTGG 586
GGAAGAGATTGATAAGCTTCAGAAAGGGATCTTGG
predict_At1g58210 2203 CGTTGCAGACAGAGAAGGAGTTTGTGAGGAGCTCA 2237
54-CY cDNA 611 CGTTGCAGACAGAGAAGGAGTTTGTGAGGAGCTCA 645
at1g09720 cDNA 587 TGCTGCAGACTGAGAAGGAGTTTGTGAGAAGTCCT 621
CGTTGCAGACAGAGAAGGAGTTTGTGAGGAGCTCA
predict_At1g58210 2238 TATGAAGAGTCTTATGAGAGATATTGGGATTTGGA 2272
54-CY cDNA 646 TATGAAGAGTCTTATGAGAGATATTGGGATTTGGA 680
at1g09720 cDNA 622 TATGAGCAGTCCTATGATAGGTATTGGAATTTGGA 656
TATGAAGAGTCTTATGAGAGATATTGGGATTTGGA
predict_At1g58210 2273 GAATGAAGTGACTGAGATGCAGAAGAGTGTTTGTA 2307
54-CY cDNA 681 GAATGAAGTGACTGAGATGCAGAAGAGTGTTTGTA 715
at1g09720 cDNA 657 GAATGAAGTTGAAGAAATGCAGAAGAGATGTTTGCA 691
GAATGAAGTGACTGAGATGCAGAAGAGTGTTTGTA


predict_At1g58210 2308 ACTTACAAGATGAGTTTGGTCTTGGTGCTTCTATT 2342
54-CY cDNA 716 ACTTACAAGATGAGTTTGGTCTTGGTGCTTCTATT 750
at1g09720 cDNA 692 GTTTACAAGATGAGTTTGGGGTTGGTGGAGAGATC 726
ACTTACAAGATGAGTTTGGTCTTGGTGCTTCTATT
predict_At1g58210 2343 GATGATAGCGATGCTAGGACGTTGATGGCGAGTAC 2377
54-CY cDNA 751 GATGATAGCGATGCTAGGACGTTGATGGCGAGTAC 785
at1g09720 cDNA 727 GAAGATGGTGAAGCTAGGACGCTGGTGGCAACCGC 761
GATGATAGCGATGCTAGGACGTTGATGGCGAGTAC
predict_At1g58210 2378 CGCTTTGAGTTCTTGTAAGGACACACTTGCTAAGC 2412
54-CY cDNA 786 CGCTTTGAGTTCTTGTAAGGACACACTTGCTAAGC 820
at1g09720 cDNA 762 TGCTCTAAGCTCTTGTAAAGAGACATTGCTAAGC 796
CGCTTTGAGTTCTTGTAAGGACACACTTGCTAAGC
predict_At1g58210 2413 TTGAGGAGAAACAGAAGATATCTGTTGAAGAAGCC 2447
54-CY cDNA 821 TTGAGGAGAAACAGAAGATATCTGTTGAAGAAGCC 855
at1g09720 cDNA 797 TTGAGGAGACGCAGAAGCGATTTCTGAAGACGCC 831
TTGAGGAGAAACAGAAGATATCTGTTGAAGAAGCC
predict_At1g58210 2448 GAGATTGAGAAAGGAAGGATTACTACTGCGAAAGA 2482
54-CY cDNA 856 GAGATTGAGAAAGGAAGGATTACTACTGCGAAAGA 890
at1g09720 cDNA 832 GGAATTGAGAAGGAAAGGATTGATACTGCAACAGA 866
GAGATTGAGAAAGGAAGGATTACTACTGCGAAAGA
predict_At1g58210 2483 AAGGTTTTATGCATTGAGGAACAAGTTTGAGAAAC 2517
54-CY cDNA 891 AAGGTTTTATGCATTGAGGAACAAGTTTGAGAAAC 925
at1g09720 cDNA 867 AAGGTGTGAAGCACTGAAGAAAAAGTTTGAGATTA 901
AAGGTTTTATGCATTGAGGAACAAGTTTGAGAAAC
predict_At1g58210 2518 CGGAGAGTGATGTTCTTGATGATGTTATTAGGACA 2552
54-CY cDNA 926 CGGAGAGTGATGTTCTTGATGATGTTATTAGGACA 960
at1g09720 cDNA 902 AGG----------TAGAAGAACA 914
CGGAGAGTGATGTTCTTGATGATGTTATTAGGACA
predict_At1g58210 2553 GACGAAGAAGAAGAGAAAGAGGCTGATGTGGT-TC 2586
54-CY cDNA 961 GACGAAGAAGAAGAGAAAGAGGCTGATGTGGT-TC 994
at1g09720 cDNA 915 AGCCAAGAAAGCTTTTCATGGTCAAGAATCGAGTT 949
GACGAAGAAGAAGAGAAAGAGGCTGATGTGGTGTC
predict_At1g58210 2587 AAGAATCGAGCTATGAATCTGAGAGGGAAGATTCA 2621
54-CY cDNA 995 AAGAATCGAGCTATGAATCTGAGAGGGAAGATTCA 1029
at1g09720 cDNA 950 ATGAATCTGTGAAAGAATCGAGACAGATTGATTTG 984
AAGAATCGAGCTATGAATCTGAGAGGGAAGATTCA
predict_At1g58210 2622 AATGAGAATCTTACTGTTGTGAAGCTTGCGGAGAA 2656
54-CY cDNA 1030 AATGAGAATCTTACTGTTGTGAAGCTTGCGGAGAA 1064
at1g09720 cDNA 985 AATGAGAACTTGAGTAATGTTGACTTTGCAGAGAA 1019
AATGAGAATCTTACTGTTGTGAAGCTTGCGGAGAA
predict_At1g58210 2657 GATTGATGATCTTGTGCATAGGGTTGTTTCATTGG 2691
54-CY cDNA 1065 GATTGATGATCTTGTGCATAGGGTTGTTTCATTGG 1099
at1g09720 cDNA 1020 GATAGATGAGCTTGTAGAGAAAGTGGTTTCTCTGG 1054
GATTGATGATCTTGTGCATAGGGTTGTTTCATTGG
predict_At1g58210 2692 AGACTAATGCTTCGTCTCATACTGCATTGGTGAAA 2726
54-CY cDNA 1100 AGACTAATGCTTCGTCTCATACTGCATTGGTGAAA 1134
at1g09720 cDNA 1055 AAACAACAGCTTTGTCTCATACGGCTTTGTTAAAG 1089
AGACTAATGCTTCGTCTCATACTGCATTGGTGAAA
predict_At1g58210 2727 ACATTAAGATCAGAGACAGACGAGTTACATGAACA 2761
54-CY cDNA 1135 ACATTAAGATCAGAGACAGACGAGTTACATGAACA 1169
at1g09720 cDNA 1090 ACATTAAGATCAGAGACAAATGAGTTACAGGACCA 1124
ACATTAAGATCAGAGACAGACGAGTTACATGAACA


predict_At1g58210 2762 TATTCGTGGTCTAGAGGAGGACAAAGCAGCTCTTG 2796
54-CY cDNA 1170 TATTCGTGGTCTAGAGGAGGACAAAGCAGCTCTTG 1204
at1g09720 cDNA 1125 TATCGTGATGTAGAGAAAGACAAGGCATGTCTTG 1159
TATTCGTGGTCTAGAGGAGGACAAAGCAGCTCTTG
predict_At1g58210 2797 TTTCGGATGCCACGGTTATGAAACAGAGGATAACG 2831
54-CY cDNA 1205 TTTCGGATGCCACGGTTATGAAACAGAGGATAACG 1239
at1g09720 cDNA 1160 TATCAGACTCCATGGACATGAAGAAAGGATAACG 1194
TTTCGGATGCCACGGTTATGAAACAGAGGATAACG
predict_At1g58210 2832 GTTCTTGAAGACGAGCTGAGAAATGTTAGAAAACT 2866
54-CY cDNA 1240 GTTCTTGAAGACGAGCTGAGAAATGTTAGAAAACT 1274
at1g09720 cDNA 1195 GTCCTTGAAGATGAGCTAAGAAAAGTTAAAACCT 1229
GTTCTTGAAGACGAGCTGAGAAATGTTAGAAAACT
predict_At1g58210 2867 ATTTCAGAAAGTGGAAGATCAAAACAAGAATCTAC 2901
54-CY cDNA 1275 ATTTCAGAAAGTGGAAGATCAAAACAAGAATCTAC 1309
at1g09720 cDNA 1230 GTTCAAAGAGTGGAAGACCAGAACAAGAATCTTC 1264
ATTTCAGAAAGTGGAAGATCAAAACAAGAATCTAC
predict_At1g58210 2902 AGAACCAATTCAAGGTAGCTAATAGGACTGTTGAT 2936
54-CY cDNA 1310 AGAACCAATTCAAGGTAGCTAATAGGACTGTTGAT 1344
at1g09720 cDNA 1265 ACAAGCACTTGACTGAAGCTAATAGTACAGCTAAG 1299
AGAACCAATTCAAGGTAGCTAATAGGACTGTTGAT
predict_At1g58210 2937 GATTTATCTGGCAAGATACAAGATGTGAAGATGGA 2971
54-CY cDNA 1345 GATTTATCTGGCAAGATACAAGATGTGAAGATGGA 1379
at1g09720 cDNA 1300 GACTTATCAGGGAAGTTACAGGAAGTGAAAATGGA 1334
GATTTATCTGGCAAGATACAAGATGTGAAGATGGA
predict_At1g58210 2972 TGAAGATGTTGAAGGAGCTGGAATTTTCCAGGAGT 3006
54-CY cDNA 1380 TGAAGATGTTGAAGGAGCTGGAATTTTCCAGGAGT 1414
at1g09720 cDNA 1335 TGAAGATGTCGAAGGTGATGGACTTAACCCAGAG- 1368
TGAAGATGTTGAAGGAGCTGGAATTTTCCAGGAGT
predict_At1g58210 3007 TGCCTGTTGTCTCAGGATCAGAAGATTCTCGAGAT 3041
54-CY cDNA 1415 TGCCTGTTGTCTCAGGATCAGAAGATTCTCGAGAT 1449
at1g09720 cDNA 1369 ----GACATCCAGG---AGGAGGATACTGTTGAG 1395
TGCCTGTTGTCTCAGGATCAGAAGATTCTCGAGAT
predict_At1g58210 3042 GATTTGAAGTCAGTCTCAACGGAGAAAACGAAGAA 3076
54-CY cDNA 1450 GATTTGAAGTCAGTCTCAACGGAGAAAACGAAGAA 1484
at1g09720 cDNA 1396 GATTCGGATTCAATATCAACGAGAGAGAGATAAA 1430
GATTTGAAGTCAGTCTCAACGGAGAAAACGAAGAA
predict_At1g58210 3077 AGATGTTATTGCAGTTAAGGAGAGTGAAGATGGTG 3111
54-CY cDNA 1485 AGATGTTATTGCAGTTAAGGAGAGTGAAGATGGTG 1519
at1g09720 cDNA 1431 GAATGCAGAGGAAATTAAAGAGG-CCATGGTAATC 1464
AGATGTTATTGCAGTTAAGGAGAGTGAAGATGGTG
predict_At1g58210 3112 AAAGAGCTCA-AGAAGAGAAACCAGAGATAAAAGA 3145
54-CY cDNA 1520 AAAGAGCTCA-AGAAGAGAAACCAGAGATAAAAGA 1553
at1g09720 cDNA 1465 AAACAGAGCAGAGATCAAGAGTCGATGCAGGAAGA 1499
AAAGAGCTCAGAGAAGAGAAACCAGAGATAAAAGA
predict_At1g58210 3146 TTCTTTTGCGTTATCAGAAACCGCAAGCACTTGTT 3180
54-CY cDNA 1554 TTCTTTTGCGTTATCAGAAACCGCAAGCACTTGTT 1588
at1g09720 cDNA 1500 GAAATCAGAGACAAGAGACTCTTGTGGTGGTTTAT 1534
TTCTTTTGCGTTATCAGAAACCGCAAGCACTTGTT
predict_At1g58210 3181 TCGGAACAGAAG-CTGAAGATCTGGTGACGGAAGA 3214
54-CY cDNA 1589 TCGGAACAGAAG-CTGAAGATCTGGTGACGGAAGA 1622
at1g09720 cDNA 1535 CAGAAACAGAAAGCACTTGCTTTGGT-ACAGAGGC 1568
TCGGAACAGAAGGCTGAAGATCTGGTGACGGAAGA


predict_At1g58210 3215 CGAGGATGAAGAGACACCAAACTGGAGACATTTGT 3249
54-CY cDNA 1623 CGAGGATGAAGAGACACCAAACTGGAGACATTTGT 1657
at1g09720 cDNA 1569 GGAAGATGAGGAGAGGAGAAACTGGAGGCAGTTGT 1603
CGAGGATGAAGAGACACCAAACTGGAGACATTTGT
predict_At1g58210 3250 TACCAG---ATGGCATGGAGGATAGAGAGAAAGTT 3281
54-CY cDNA 1658 TACCAG---ATGGCATGGAGGATAGAGAGAAAGTT 1689
at1g09720 cDNA 1604 TACCCGCAGATGGCATGGAGGATAGAGAGAAGGTT 1638
TACCAGCAGATGGCATGGAGGATAGAGAGAAAGTT
predict_At1g58210 3282 CTGTTAGATGAATACACATCAGTACTAAGGGATTA 3316
54-CY cDNA 1690 CTGTTAGATGAATACACATCAGTACTAAGGGATTA 1724
at1g09720 cDNA 1639 CTATTGGATGAATACTCATCAGTACTAAGAGACTA 1673
CTGTTAGATGAATACACATCAGTACTAAGGGATTA
predict_At1g58210 3317 TAGAGAAGTAAAGAGAAAGTTAGGTGATGTTGAGA 3351
54-CY cDNA 1725 TAGAGAAGTAAAGAGAAAGTTAGGTGATGTTGAGA 1759
at1g09720 cDNA 1674 TAGAGAAGTGAAGAGAAAGTTGAGTGAAGTTGAGA 1708
TAGAGAAGTAAAGAGAAAGTTAGGTGATGTTGAGA
predict_At1g58210 3352 AAAAGAACCGGGAAGGATTTTTCGAGCTAGCATTA 3386
54-CY cDNA 1760 AAAAGAACCGGGAAGGATTTTTCGAGCTAGCATTA 1794
at1g09720 cDNA 1709 AGAAGAACCGTGATGGCTTCTTTGAGCTGGCATTG 1743
AAAAGAACCGGGAAGGATTTTTCGAGCTAGCATTA
predict_At1g58210 3387 CAGTTGAGAGAACTCAAGAATGCTGTTGCTTACAA 3421
54-CY cDNA 1795 CAGTTGAGAGAACTCAAGAATGCTGTTGCTTACAA 1829
at1g09720 cDNA 1744 CAGTTGCGAGAGCTCAAGAATGCTGTTTCTTGCGA 1778
CAGTTGAGAGAACTCAAGAATGCTGTTGCTTACAA
predict_At1g58210 3422 AGACGTGGAGATTCAGTCGTTACGCCAAAAACTTG 3456
54-CY cDNA 1830 AGACGTGGAGATTCAGTCGTTACGCCAAAAACTTG 1864
at1g09720 cDNA 1779 AGACGTGGATTTCATTTCTTACACCAGAAACCAG 1813
AGACGTGGAGATTCAGTCGTTACGCCAAAAACTTG
predict_At1g58210 3457 ACACCACTGGGAAAGATTCACCACACCAAGGAGAA 3491
54-CY cDNA 1865 ACACCACTGGGAAAGATTCACCACACCAAGGAGAA 1899
at1g09720 cDNA 1814 AATTGCCCGGCCAGGGTTTCCTCATCCAG---TA 1845
ACACCACTGGGAAAGATTCACCACACCAAGGAGAA
predict_At1g58210 3492 GGAAATAACCAGTTGGAACATGAACAAGGACATCA 3526
54-CY cDNA 1900 GGAAATAACCAGTTGGAACATGAACAAGGACATCA 1934
at1g09720 cDNA 1846 GAAAGAACCGGG----------C 1859
GGAAATAACCAGTTGGAACATGAACAAGGACATCA
predict_At1g58210 3527 TGAAACTGTGAGCATTTCGCCAACTTCTAACTTTT 3561
54-CY cDNA 1935 TGAAACTGTGAGCATTTCGCCAACTTCTAACTTTT 1969
at1g09720 cDNA 1860 TGAAAGTGTGAGCATCTCTCATAGCTCTAACTCTT 1894
TGAAACTGTGAGCATTTCGCCAACTTCTAACTTTT
predict_At1g58210 3562 CGGTTGCCACCACGCCGCATCATCAAGTAGGAGAT 3596
54-CY cDNA 1970 CGGTTGCCACCACGCCGCATCATCAAGTAGGAGAT 2004
at1g09720 cDNA 1895 CATTCTCCATGCCACCACTCCTCAAAGAGGAGAC 1929
CGGTTGCCACCACGCCGCATCATCAAGTAGGAGAT
predict_At1g58210 3597 GTGAAGAGAACACCTGGAAGAACAAAGTCAACTGA 3631
54-CY cDNA 2005 GTGAAGAGAACACCTGGAAGAACAAAGTCAACTGA 2039
at1g09720 cDNA 1930 TTGAAGAGAGCATCTGAACAAGAAAAGGAAGACGG 1964
GTGAAGAGAACACCTGGAAGAACAAAGTCAACTGA
predict_At1g58210 3632 GGTTAGGGTGAAATTTGCAGACGTTGATGACAGCC 3666
54-CY cDNA 2040 GGTTAGGGTGAAATTTGCAGACGTTGATGACAGCC 2074
at1g09720 cDNA 1965 GTTTAAGGTGAAATTTGCAGGCATTAGTGACAGCC 1999
GGTTAGGGTGAAATTTGCAGACGTTGATGACAGCC


predict_At1g58210 3667 CGAGAACAAAGATTCCAACTGTGGAAGACAAAGTG 3701
54-CY cDNA 2075 CGAGAACAAAGATTCCAACTGTGGAAGACAAAGTG 2109
at1g09720 cDNA 2000 TAAGAAAGAAATTCCAACTGTGGAAGAAAAAGTG 2034
CGAGAACAAAGATTCCAACTGTGGAAGACAAAGTG
predict_At1g58210 3702 CGTGCAGACATTGACGCGGTGTTGGAAGAGAATCT 3736
54-CY cDNA 2110 CGTGCAGACATTGACGCGGTGTTGGAAGAGAATCT 2144
at1g09720 cDNA 2035 CGTGGAGATATTGATGCGGTGCTGGAAGAGAATAT 2069
CGTGCAGACATTGACGCGGTGTTGGAAGAGAATCT
predict_At1g58210 3737 CGAGTTTTGGTTAAGATTTAGTACATCGGTACATC 3771
54-CY cDNA 2145 CGAGTTTTGGTTAAGATTTAGTACATCGGTACATC 2179
at1g09720 cDNA 2070 TGAATTCTGGTTAAGATTTAGCACATCTGTTCATC 2104
CGAGTTTTGGTTAAGATTTAGTACATCGGTACATC
predict_At1g58210 3772 AGATACAGAAGTACCAGACAACAGTTCAAGATCTG 3806
54-CY cDNA 2180 AGATACAGAAGTACCAGACAACAGTTCAAGATCTG 2214
at1g09720 cDNA 2105 AGATACAGAAGTACCACACATCAGTTCAAGACCTG 2139
AGATACAGAAGTACCAGACAACAGTTCAAGATCTG
predict_At1g58210 3807 AAATCAGAGCTATCGAAACTGAGAATTGAAAGCAA 3841
54-CY cDNA 2215 AAATCAGAGCTATCGAAACTGAGAATTGAAAGCAA 2249
at1g09720 cDNA 2140 AAAGCAGAGCTATCGAAAAT------TGAAAGCAA 2168
AAATCAGAGCTATCGAAACTGAGAATTGAAAGCAA
predict_At1g58210 3842 GCAACAACAGGAATCTCCAAGAAGTAGTAGTAATA 3876
54-CY cDNA 2250 GCAACAACAGGAATCTCCAAGAAGTAGTAGTAATA 2284
at1g09720 cDNA 2169 GCAACAAGG---TAATGCTGGGAGCAGTAGTAACA 2200
GCAACAACAGGAATCTCCAAGAAGTAGTAGTAATA
predict_At1g58210 3877 CTGCAGTTGCATCAGAGGCAAAGCCTATCTATAGA 3911
54-CY cDNA 2285 CTGCAGTTGCATCAGAGGCAAAGCCTATCTATAGA 2319
at1g09720 cDNA 2201 CTGCACTTGCATCAGAGGCAAAGCCTATATATAGA 2235
CTGCAGTTGCATCAGAGGCAAAGCCTATCTATAGA
predict_At1g58210 3912 CACCTCAGAGAGATTCGAACAGAACTACAGCTATG 3946
54-CY cDNA 2320 CACCTCAGAGAGATTCGAACAGAACTACAGCTATG 2354
at1g09720 cDNA 2236 CACCTTAGAGAGATTCGGACAGAGCTGCAGTTATG 2270
CACCTCAGAGAGATTCGAACAGAACTACAGCTATG
predict_At1g58210 3947 GCTAGAGAACAGTGCGGTTCTTAAAGACGAACTCC 3981
54-CY cDNA 2355 GCTAGAGAACAGTGCGGTTCTTAAAGACGAACTCC 2389
at1g09720 cDNA 2271 GCTAGAGAACAGTGCAATTCTGAGAGACGAACTTG 2305
GCTAGAGAACAGTGCGGTTCTTAAAGACGAACTCC
predict_At1g58210 3982 AGGGAAGGTATGCGTCATTAGCTAATATCCAAGAA 4016
54-CY cDNA 2390 AGGGAAGGTATGCGTCATTAGCTAATATCCAAGAA 2424
at1g09720 cDNA 2306 AGGGTAGGTATGCAACACTTTGCAATATCAAAGAT 2340
AGGGAAGGTATGCGTCATTAGCTAATATCCAAGAA
predict_At1g58210 4017 GAAATCGCAAGAGTCACGGCTCAATCAGGTGGCAA 4051
54-CY cDNA 2425 GAAATCGCAAGAGTCACGGCTCAATCAGGTGGCAA 2459
at1g09720 cDNA 2341 GAAGTTTCAAGGGTCACGTCTCAATCGGGTGCAAC 2375
GAAATCGCAAGAGTCACGGCTCAATCAGGTGGCAA
predict_At1g58210 4052 TAAAGTAAGTGATTCAGAGATTAGTGGTTACCAAG 4086
54-CY cDNA 2460 TAAAGTAAGTGATTCAGAGATTAGTGGTTACCAAG 2494
at1g09720 cDNA 2376 TGAAGTAAGCAATACAGAGATAAGAGGTTACCAGG 2410
TAAAGTAAGTGATTCAGAGATTAGTGGTTACCAAG
predict_At1g58210 4087 CTGCAAAGTTCCATGGAGAGATACTTAACATGAAA 4121
54-CY cDNA 2495 CTGCAAAGTTCCATGGAGAGATACTTAACATGAAA 2529
at1g09720 cDNA 2411 CTGCAAAGTTCCATGGTGAGATACTTAACATGAAA 2445
CTGCAAAGTTCCATGGAGAGATACTTAACATGAAA


predict_At1g58210 4122 CAGGAGAACAAAAGGGTCTCGACCGAACTTCACTC 4156
54-CY cDNA 2530 CAGGAGAACAAAAGGGTCTCGACCGAACTTCACTC 2564
at1g09720 cDNA 2446 CAGGAGAACAAAAGGGTGTTCAACGAACTTCAAGC 2480
CAGGAGAACAAAAGGGTCTCGACCGAACTTCACTC
predict_At1g58210 4157 GGGTCTTGACCGTGTAAGAGCATTGAAAACCGAAG 4191
54-CY cDNA 2565 GGGTCTTGACCGTGTAAGAGCATTGAAAACCGAAG 2599
at1g09720 cDNA 2481 AGGCCTTGATCGTGCAAGAGCACTGAGAGCTGAAG 2515
GGGTCTTGACCGTGTAAGAGCATTGAAAACCGAAG
predict_At1g58210 4192 TGGAGAGGATTCTAAGTAAACTAGAAGAGGATCTT 4226
54-CY cDNA 2600 TGGAGAGGATTCTAAGTAAACTAGAAGAGGATCTT 2634
at1g09720 cDNA 2516 TGGAAAGAGTTGTATGCAAACTAGAAGAGAATCTT 2550
TGGAGAGGATTCTAAGTAAACTAGAAGAGGATCTT
predict_At1g58210 4227 GGGATCTCAAGTGCAACAGAAGCAAGAACAACTCC 4261
54-CY cDNA 2635 GGGATCTCAAGTGCAACAGAAGCAAGAACAACTCC 2669
at1g09720 cDNA 2551 GGGATTTTGGATGGAGGCTACAAG---ATCTCT 2582
GGGATCTCAAGTGCAACAGAAGCAAGAACAACTCC
predict_At1g58210 4262 AAGCAAGAG------CTCATCAAGTGGAAGACCGA 4290
54-CY cDNA 2670 AAGCAAGAG------CTCATCAAGTGGAAGACCGA 2698
at1g09720 cDNA 2583 GAGCAAGAGAATGCCATCATCAGCTGGAAAGCCTA 2617
AAGCAAGAGAATGCCCTCATCAAGTGGAAGACCGA
predict_At1g58210 4291 GGATTCCATTACGGTCATTCTTGTTCGGTGTCAAG 4325
54-CY cDNA 2699 GGATTCCATTACGGTCATTCTTGTTCGGTGTCAAG 2733
at1g09720 cDNA 2618 GGATTCCACTAAGGTCATTCTTGTTTGGTGTCAAG 2652
GGATTCCATTACGGTCATTCTTGTTCGGTGTCAAG
predict_At1g58210 4326 TTAAAGAAGAATAGACAGCAAAAGCAATCAGCATC 4360
54-CY cDNA 2734 TTAAAGAAGAATAGACAGCAAAAGCAATCAGCATC 2768
at1g09720 cDNA 2653 TTAAAGAAGTATAACACAACGAAGCAACACTC 2687
TTAAAGAAGAATAGACAGCAAAAGCAATCAGCATC
predict_At1g58210 4361 ATCCCTCTTCTCTTGTGTCAGTCCATCTCCAGGTC 4395
54-CY cDNA 2769 ATCCCTCTTCTCTTGTGTCAGTCCATCTCCAGGTC 2803
at1g09720 cDNA 2688 AACAATCTTCTCTTGTGTCAGTCCATCTCCAGCTC 2722
ATCCCTCTTCTCTTGTGTCAGTCCATCTCCAGGTC
predict_At1g58210 4396 TGCATAAACAGTCTAGTTACAGTAGGCCACCTGGG 4430
54-CY cDNA 2804 TGCATAAACAGTCTAGTTACAGTAGGCCACCTGGG 2838
at1g09720 cDNA 2723 TGAACAAACAGTGTAGTTACATTATCCCACCTGCA 2757
TGCATAAACAGTCTAGTTACAGTAGGCCACCTGGG
predict_At1g58210 4431 AAGCTCCCTGAATAA---------- 4445
54-CY cDNA 2839 AAGCTCCCTGAATAA---------- 2853
at1g09720 cDNA 2758 AAGCTCCCTGAATATGTCAAAAGAAGCTGA 2787
AAGCTCCCTGAATAAGTCAAAAGAAGCTGA

APPENDIX VIX AFFYMETRIX MICROARRAY CHIPS USED TO GENERATE THE 'FLOWERNET' CORRELATION NETWORK

3 chips: Regulation of the *Arabidopsis* anther transcriptome by DYT1 for pollen development (Feng, 2012 #83)

6 chips: MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis (Yang, 2007 #28)

7 chips: Transcriptome analysis of haploid male gametophyte development in *Arabidopsis* (Honys, 2004 #169)

13 chips: Transcriptional regulators of stamen development in *Arabidopsis* identified by transcriptional profiling (Mandaokar, 2006 #170).

21 chips: A gene expression map of *Arabidopsis thaliana* development (Schmid, 2005 #171).

Additionally, three experiments from the Wilson lab are included:

2 chips: Microarray analysis of buds from Y2H1 (At3g58110) mutant compared with wt L*er*

6 chips: Microarray analysis of L*er* buds through PMI/BC/PMII Developmental Stages

8 chips: Timecourse microarray analysis of buds from MYB26-inducible lines compared with L*er*

APPENDIX X GENE LISTS OF THE CLUSTERS IDENTIFIED AS MS1 DOWNSTREAM TARGETS BY 'FLOWERNET'

AGI code	Name	FlowerNet Cluster	Fold change in *ms1* young anthers	Fold change in *ms1* old anthers
AT1G01310		7		-3.146
AT1G03050		7		-3.72
AT1G04470		7		-10.527
AT1G13970	DUF1336	7		-10.585
AT1G19500		7		-11.666
AT1G24520	BCP1	7		-20.59
AT1G35490		7		-11.084
AT1G68110		7		-11.874
AT1G74000	SS3	7		-5
AT2G02140	LCR72	7		-9.538
AT2G07040	PRK2A	7		-4.888
AT2G07180		7		-2.641
AT2G13350	CaLB domain	7		-7.032
AT2G16730		7		-4.665
AT2G18470	PERK4	7		-7.457
AT2G20700	LLG2	7		-2.33
AT2G22180		7		-4.276
AT2G24450	FLA3	7		-21.494
AT2G27180		7		-9.99
AT2G36020	HVA22J	7		-10.803
AT2G43230		7		-8.627
AT2G45800		7		-25.583
AT3G04700		7		-24.911
AT3G04710		7		-5.595
AT3G05610		7		-16.085
AT3G05930	GLP8	7		-5.967
AT3G16040		7		-3.194
AT3G20220		7		-20.844
AT3G21700	SGP2	7		-9.314
AT3G25165	RALFL25	7		-12.837
AT3G26110		7		-26.396
AT3G26860		7		-14.563
AT3G42640	HA8	7		-18.575
AT3G43120		7		-10.937
AT3G51300	ROP1	7		-6.711
AT3G52600	CWINV2	7		-5.837

AGI code	Name	FlowerNet Cluster	Fold change in *ms1* young anthers	Fold change in *ms1* old anthers
AT3G54800		7		-11.734
AT3G61160		7		-16.28
AT3G61230		7		-4.553
AT3G62640	DUF3511	7		-12.222
AT3G62710		7		-20.055
AT4G03290		7		-2.578
AT4G04930	DES-1-LIKE	7		-9.042
AT4G05330	AGD13	7		-3.001
AT4G07960	CSLC12	7		-4.337
AT4G18395		7		-18.512
AT4G24640	APPB1	7		-10.654
AT4G25780		7		-9.577
AT4G27580		7		-13.386
AT4G28280	LLG3	7		-7.27
AT4G30140	CDEF1	7		-7.31
AT4G35180	LHT7	7		-4.916
AT4G35700	C2H2 type	7		-6.185
AT4G36490	SFH12	7		-16.436
AT4G38190	CSLD4	7		-3.244
AT4G39670	GLTP	7		-4.68
AT5G03690		7		-5.243
AT5G04180	ACA3	7		-17.45
AT5G15110		7		-9.181
AT5G20410	MGD2	7		-8.112
AT5G26700		7		-38.989
AT5G27980		7		-7.857
AT5G39880		7		-27.035
AT5G47000		7		-19.509
AT5G48140		7		-24.433
AT5G50830		7		-21.587
AT5G54095		7		-12.141
AT5G55930	OPT1	7		-3.196
AT5G58050	SVL4	7		-3.886
AT5G62750		7		-10.68
AT5G64790		7		-5.274
AT1G08730	XIC	14		-2.531
AT1G12070		14		-5.294
AT1G16360		14		-3.497
AT1G19890	MGH3	14		-8.183
AT1G22730		14		-2.34
AT1G26320		14		-2.351

AGI code	Name	FlowerNet Cluster	Fold change in *ms1* young anthers	Fold change in *ms1* old anthers
AT1G26480	GRF12	14		-8.759
AT1G62450		14		-3.181
AT2G01330		14		-2.063
AT2G13570	NF-YB7	14		-4.613
AT2G19330	PIRL6	14		-4.046
AT2G22950		14		-3.226
AT2G33670	MLO5	14		-2.839
AT2G46140		14		-3.488
AT3G18570		14		-2.111
AT3G46750		14		-4.39
AT3G51070		14		-4.782
AT3G52620		14		-5.022
AT4G20780	CML42	14		-3.087
AT5G02570		14		-2.683
AT5G13990	EXO70C2	14		-4.083
AT5G14670	ARFA1B	14		-2.131
AT5G19610	GNL2	14		-3.792
AT5G46200	DUF239	14		-3.103
AT5G51030		14		-5.948
AT5G59370	ACT4	14		-8.442
AT1G07850	DUF604	21		-2.817
AT1G52570	PLDALPHA2	21		-2.728
AT1G54560	XIE	21		-4.042
AT2G25890		21		-5.337
AT3G08560	VHA-E2	21		-4.866
AT3G17630	CHX19	21		-8.143
AT3G19090		21		-3.471
AT3G51490	TMT3	21		-3.557
AT4G02140		21		-3.443
AT4G10440		21		-2.298
AT4G11030		21		-2.584
AT5G66020	ATSAC1B	21		-2.571
AT1G10770		23		-26.539
AT1G28270	RALFL4	23		-25.121
AT1G29140		23		-41.886
AT1G55570	sks12	23		-60.775
AT1G73630		23		-3.572
AT2G46860	PPa3	23		-26.51
AT2G47050		23		-43.197
AT3G01240		23		-35.883
AT3G01250		23		-14.81

AGI code	Name	FlowerNet Cluster	Fold change in *ms1* young anthers	Fold change in *ms1* old anthers
AT3G01270		23		-90.422
AT3G01700	AGP11	23		-58.924
AT3G03430		23		-32.116
AT3G07820		23		-99.446
AT3G13390	sks11	23		-27.028
AT3G20865	AGP40	23		-35.24
AT3G21180	ACA9	23		-9.306
AT3G28750		23		-101.467
AT3G62170	VGDH2	23		-62.233
AT3G62180		23		-18.444
AT4G02250		23		-45.558
AT4G35010	BGAL11	23		-61.347
AT5G07420		23		-12.641
AT5G07430		23		-39.81
AT5G14380	AGP6	23		-53.233
AT5G19580		23		-28.103
AT5G20230	BCB	23		-4.726
AT5G26060		23		-4.248
AT5G61720	DUF1216	23		-13.033
AT1G04670		33		-11.438
AT1G07540	TRFL2	33		-2.116
AT1G50310	STP9	33		-5.381
AT1G52680		33		-22.351
AT1G54870		33		-3.528
AT1G71680		33		-9.53
AT1G80660	HA9	33		-52.04
AT2G07560	HA6	33		-38.229
AT2G28355	LCR5	33		-30.622
AT2G29790	MEG	33		-22.444
AT2G32890	RALFL17	33		-6.281
AT2G33690		33		-4.359
AT3G02480	LEA	33		-27.157
AT3G05960	STP6	33		-4.063
AT3G21970	DUF26	33		-10.838
AT3G53080		33		-4.162
AT4G01470	TIP1;3	33		-5.956
AT4G13230	LEA	33		-11.983
AT5G44300		33		-14.66
AT5G46940		33		-6.016
AT5G62850	AtVEX1	33		-17.328
AT1G01280	CYP703A2	37	4.229	

AGI code	Name	FlowerNet Cluster	Fold change in *ms1* young anthers	Fold change in *ms1* old anthers
AT1G02813		37	-24.842	
AT1G62940	ACOS5	37	2.929	
AT1G69500	CYP704B1	37	2.65	
AT1G76470		37	-11.783	
AT3G23770		37	-8.689	
AT4G14080	MEE48	37	3.584	
AT4G20420		37	3.572	
AT4G34850	LAP5	37	3.529	
AT4G35420	DRL1	37	2.924	
AT1G01780		58		-3.623
AT1G24620		58		-6.258
AT1G66360	CaLB domain	58		-3.05
AT1G79910		58		-4.439
AT2G19000		58		-20.948
AT2G22860	PSK2	58		-3.745
AT3G07830		58		-4.214
AT3G10460		58		-6.033
AT3G13065	SRF4	58		-3.27
AT5G14890		58		-4.221
AT1G23520	DUF220	73	-6.266	
AT1G23670	DUF220	73	-7.094	-2.011
AT1G28430	CYP705A24	73	-13.916	-8.281
AT1G30350		73	-12.508	-4.847
AT1G75930	EXL6	73	-42.587	-10.724
AT3G03910	GDH3	73	-6.313	
AT3G52810	PAP21	73	-7.682	
AT5G07510	GRP14	73	-21.803	-4.844
AT5G07520	GRP18	73	-26.946	
AT5G07540	GRP16	73	-31.518	
AT5G07560	GRP20	73	-118.807	-14.064
AT5G61605		73	-17.062	-8.101
AT1G13140	CYP86C3	81	-21.373	
AT1G67990	TSM1	81	-82.629	
AT1G71160	KCS7	81	-7.945	
AT2G19070	SHT	81	-14.678	
AT3G52160	KCS15	81	-11.169	
AT4G28395		81	-12.854	
AT5G13380		81	-14.599	
AT5G16960		81	-20.379	
AT5G48210		81	-28.474	

AGI code	Name	FlowerNet Cluster	Fold change in *ms1* young anthers	Fold change in *ms1* old anthers
AT5G49070	KCS21	81	-3.329	
AT5G65205		81	-12.919	
AT1G63060		110		-4.926
AT2G19770	PRF5	110		-22.542
AT3G28790	DUF1216	110		-26.322
AT3G28980	DUF1216	110		-21.476
AT4G13560	UNE15	110		-39.722
AT4G29340	PRF4	110		-29.606
AT5G09550		110		-7.955
AT5G17480	PC1	110		-6.338
AT5G38760	LEA	110		-17.118
AT5G53820	LEA	110		-11.052
AT1G14420	AT59	116		-16.581
AT1G49290		116		-9.227
AT2G17500		116		-2.847
AT2G31500	CPK24	116		-9.983
AT3G25170	RALFL26	116		-10.52
AT4G16480	INT4	135		-3.082
AT4G19960	KUP9	135		-2.663
AT5G19600	SULTR3;5	135		-2.844
AT5G25550	LRR	135		-2.385
AT1G76240	DUF241	156		-3.595
AT2G38240	2OG	156		-3.135
AT3G06490	MYB108	156		-3.229
AT3G11430	GPAT5	156		-10.053
AT3G22910		156		-8.045
AT4G25010		156		-4.098
AT5G13580		156		-3.088
AT5G50800		156		-2.767
AT1G06260		206		-50.313
AT1G06990		206		-7.668
AT1G23570	DUF220	206		-26.43
AT1G59740		206		-4.616
AT1G75910	EXL4	206		-81.15
AT1G75920		206		-32.481
AT3G25050	XTH3	206		-56.39
AT1G71880	SUC1	242		-3.777
AT3G28780	DUF1216	242		-11.399
AT3G28840	DUF1216	242		-25.806
AT5G42170		242		-5.432
AT1G18280		254		-6.497

AGI code	Name	FlowerNet Cluster	Fold change in *ms1* young anthers	Fold change in *ms1* old anthers
AT1G24400	LHT2	254		-4.666
AT4G08670		254		-5.579
AT5G20710	BGAL7	254		-6.669
AT1G68875		263		-133.448
AT1G74540	CYP98A8	263		-19.068
AT2G03740		263		-9.01
AT3G51590	LTP12	263		-93.889
AT4G14815		263		-7.852
AT4G37900		263		-33.951
AT3G47440	TIP5;1	277		-9.677
AT4G25590	ADF7	277		-12.796
AT5G52360	ADF10	277		-8.897
AT1G02790	PGA4	308		-36.039
AT3G13400	sks13	308		-29.541
AT3G28830	DUF1216	308		-17.827
AT3G57690	AGP23	308		-14.2
AT5G45880		308		-17.305
AT1G03390		347		8.133
AT2G42940		347		6.699
AT3G07450		347		4.447
AT3G52130		347		3.879
AT4G29980		347		2.951
AT2G31980	CYS2	406		-25.166
AT5G07530	GRP17	406		-48.241
AT5G07550	GRP19	406		-167.216
AT5G59845		406		-7.105
AT1G78460		435		-16.949
AT3G62230		435		-47.189
AT5G50030		435		-38.072
AT1G13150	CYP86C4	448	-13.843	
AT1G74550	CYP98A9	448	-9.911	
AT4G23660	PPT1	448	-3.923	
AT5G44400		448	-27.952	
AT1G20120		560	-49.165	
AT2G03850	LEA	560	-20.268	
AT2G46370	JAR1	560	-5.38	
AT3G51000		560	-5.608	