Cloning and Functional Investigation of a Thermo-sensitive Female Sterility Gene *TFS1* in Rice

LI, Haoxuan

A Thesis Submitted in Partial Fulfilment

of the Requirements for the Degree of

Doctor of Philosophy

in

Biology

The Chinese University of Hong Kong

September 2017

Thesis Assessment Committee

Professor LAM Hon Ming (Chair)

Professor ZHANG Jianhua (Thesis Supervisor)

Professor ZHONG Silin (Committee Member)

Professor XIA Yiji (External Examiner)

Acknowledgement

First of all, I am grateful to my supervisor, Prof. ZHANG Jianhua, sincerely for offering me the most precious opportunity to join the warm family. During the past five years, he gave me great and continuous support and trust on my PhD study and related project. His patience and motivation inspired me to perform my research. I was encouraged by his character and morals in the area of both scientific research and real life. His guidance helped me in all the time of research and writing of this thesis. And a very special gratitude goes out to Prof. LI Chuanguo, who offered us the female sterility mutant *tfs1*. And GU Haiyong, for his kind assistance of planting the rice in Baiyun base, Guangzhou.

Besides, I also give my thanks to my thesis committee: Prof. LAM Hon Ming and Prof. ZHONG Silin, for their attendance of my graduate seminars in the past four years, for their insightful comments and encouragement on my research, for their concerns about my experimental process. I am also thankful for Prof. XIA Yiji's instruction and comments on my thesis.

I would like to thank The Chinese University of Hong Kong for the studentship and financial support for my PhD study. I also would like to acknowledge our laboratory technicians Ken LAU and Rita Fu, for their support of ordering reagents, booking gene garden for planting rice; Gene Garden worker Sam, Thomas, Andrew, and Ms. PING, for their kind works of taking care of rice in gene garden. With a special mention to university staff Freddie, for his kind help in preparing the material of scanning microscope, CLSM, TEM, paraffin section and his patient guidance in using the instruments.

I want to express my heartfelt gratitude to my labmates JIN Yu, WANG Guanqun, SONG Tao, LIU Tieyuan, CHEN Moxian, ZHU Fuyuan and XIU Zhihui, to two visiting professors, LIN Sheng and CHENG Dongliang, for their assistance in the field work in Guangzhou and Shenzhen, and to other labmates for the constructive discussions and suggestions on my project and the fun we have had in the last five years. In particular, I am grateful to Dr. YE Nenghui, for enlightening me the first glance of research and continuous guidance of experiment, also the modification of my thesis. In

the past five years, we worked together and overcame hardship together. It will be the most precious time in my life.

I want to give my thanks to the most important people in my life, my dear parents and my little sister, for their continuous trust, encouragement and support in the past five years, and for their selfless love and accompany.

Last but not least, I want to thank my friends XIAO Zhixia and YU Sheng for their endless help on analyzing sequencing data, and to thank Dr. LV Peitao for his help of protoplast transformation. Acknowledgement should also be given to everyone in my life who made me feel cared and loved.

Declaration

I hereby declare that this thesis represents my own work which has been done after registration for the degree of PhD at The Chinese University of Hong Kong, and has not been previously included in a thesis, dissertation submitted to this or other institution for degree, diploma or other qualification.

Signature: Li Haoxuan

Date: September 2017

Abstract

Male sterility lines play a fundamental role in hybrid rice breeding, whereas female sterility line seems useless due to the extreme difficulty in maintaining a homozygous line. Luckily, we have identified a rice germplasm showing female sterility separated from variety 4266 population. Further experiments have shown that it is thermo-sensitive sterile, which can be partially recovered when grown at temperature of 25-27℃. Genetic analysis shown that it is controlled by a single recessive gene that we named as temperature-sensitive female sterility 1 (*TFS1*). Pollen viability testing demonstrated that there is no significant difference between *tfs1* and 4266 plants on pollen morphology. Pollen from both 4266 and *tfs1* plants were able to germinate and grow into the ovary of both phenotypes. Embryo sac development was studied by CLSM techniques and results showed that embryo sac in *tfs1* is similar with that in 4266 before pollination, whereas the egg cell and central cell were still observed 3 DAPs. Embryo development analysis, together with pollen tube elongation experiment, indicating the double fertilization is blocked in the ovary of *tfs1* plant. Our cytological studies on *tfs1* have tentatively illustrated sterility mechanism and provide very good information for the following research on biological function of the mutant gene.

Map-based cloning is a fundamental technique and has been widely used in genetic research for gene cloning. In this study, a F2 population with 20000 individuals was generated by crossing the mutant with *nipponbare*. SSR markers were then used to locate the mutated position on chromosome3 and a 2.17Mb region was demonstrated linked closely with the mutant position by fine mapping. Unfortunately, the region contains a centromere which made it difficult to further locate another SSR marker that is closer to mutant position. Then we carried out MutMap method and find a candidate gene which belongs to the Argonaute family. To further identify the mutant gene of *tfs1*, we have a complement vector with whole genome DNA gene of this candidate gene and transferred into *tfs1*. Results shown that the complementary line displayed a recovered phenotype, indicating that we have successfully clone the mutant gene. In addition, a TFS1-GFP fusion protein has been constructed to study the subcellular localization of TFS1, results shown that TFS1 locates in cytoplasm where pre-RNAs are further processed.

To figure out which step is interrupted during double fertilization and the regulating mechanism of temperature sensitive, the transcriptome and small RNA sequencing were applied to find the candidate small RNAs which are differentially expressed between 4266 and *tfs1* plants. By analyzing the data of small RNA sequencing and RNA sequencing, we found that the target of DE miRNAs is consistent with the DE genes found from transcriptome analysis, which means *TFS1* regulated the expression of both miRNAs and genes. Our sequencing results have alsp provided some primary insights into the regulation mechanism of TFS1 protein. To further explore the molecular function of *TFS1,* more experiments using reverse genetic techniques will be carried out in the future.

摘要

近年來水稻雄性不育的研究已經取得了突破性的進展，但是對於雌性不育的研究卻非常的不足。在大田生產中，我們在 4266 群體中發現一株雌性不育的水稻，在較低的溫度下其育性可以被部分恢復，統計後代分離比後確認一個隱性單基因控制的這個性狀。細胞學觀察表明突變體的花藥和花粉發育正常，進一步觀察表明 *tfs1* 的胚囊發育與 4266 並無差別。我們隨後檢測了 *tfs1* 上花粉管伸長以及授粉後的胚胎發育情況，結果表明即使在授粉後第三天，*tfs1* 胚囊中的卵細胞區域和中央細胞一直沒有消失，表明 *tfs1* 中雙受精過程受到了阻遏。小結來說，我們的細胞學機理研究初步揭示了 *tfs1* 雌性不育的機理，並為後續進一步研究 *tfs1* 不育的分子生物學機理研究工作提供了重要的資訊。

圖位克隆是一個被廣泛應用到遺傳學研究上開展基因克隆的技術。在本研究中，我們通過將 *tfs1* 突變體和日本晴雜交構建了一個具有 20000 個單株的 F2 分離群體，然後使用 SSR 分子標記將突變基因的範圍縮小到 3 號染色體上一段 2.17M 的片段上。然而由於該區段包含了著絲粒而導致的極低交換律，我們無法再將該區段做更進一步的縮小。為解決這一問題，我們通過 MutMap 和圖位克隆相輔助的方法獲得了一個候選基因。隨後我們利用這個候選基因構建了互補植株，部分互補株系發生了互補表型，說明該候選基因正是我們所要克隆的基因。此外我們還對該基因做了亞細胞定位實驗，結果表明 *TFS1* 基因定位於細胞質中，而該處正是各種 RNA 分子加工成熟的場所。

為了進一步研究 *TFS1* 基因調控水稻雌性不育的機理，我們採用了轉錄組測序和小 RNA 測序技術來研究 *tfs1* 突變體和 4266 之間的基因表達差異。通過對比兩種組學數據的關聯分析，我們發現有差異表達的 miRNA 的目標基因的表達變化與差異表達 mRNA 之間有著顯著一致性，這一結果表明 *TFS1* 基因不僅調控了下游靶基因的表達，還影響 miRNA 的豐度變化。我們的測序結果為揭示 *TFS1* 基因調控雌性不育的生物學機理提供了一些思路，在未來的研究中，我們仍需要開展一些新的實驗來解釋 *tfs1* 突變基因與溫敏型雌性不育之間的關係。

List of Figures

List of Tables

List of Abbreviations

AGO	Argonaute
ARF	Auxin response factor
ATAF	Arabidopsis transcription activation factor
BAC	Bacterial artificial chromosome
BSA	Bulked Segregant Analysis
CLSM	Confocal laser scanning microscope
cmiRNA	Canonical miRNA
CRCK	Calmodulin-binding receptor-like cytoplasmic kinase
CUC	Cup-shaped cotyledon
DAP	Day after pollination
DCL4	Defective chloroplast and leaves 4
DIC	Differential interference contrast
DNA	Deoxyribonucleic acid
DOF	DNA-binding One Zinc Finger
dsRNA	Double-strand RNA
EGMS	Environment-sensitive genic male sterility
FPKM	Fragments per Kilobase Million
GFP	Green fluorescence protein
GO	Gene Ontology
HITS-CLIP	High throughput sequencing of RNAs isolated by crosslinking immunoprecipitation
IRGSP	International Rice Genome Sequencing Project
Kan	Kanamycin
kb	kilobase
kD	kilo Dalton
lncRNA	Long noncoding RNA
LRR	Leucine-rich repeat protein kinase family protein
Mbp	mega base pairs
MFS	Major facilitator superfamily protein
miRNA	micro RNAs
mRNA	Messenger RNA

NAC domain	NAM ATAF CUC domain
NAM	No apical meristem
nat-siRNAs	natural cis-acting siRNAs
NB-ARC	Nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4
NGS	Next Generation Sequencing
NK58S	Nongken 58S
nt	nucleotide
ORF	Open reading frame
PA64	Peiai 64
PARE	Parallel analysis of RNA end
PBS	Phosphate-buffered saline
PCD	Programmed cell death
PCR	Polymerase chain reaction
PGMS	Photoperiod-sensitive genic male sterility
Pol I	Polymerase I
PSMS	Photoperiod-sensitive male sterility
PTGS	Post-transcriptional gene silencing
QRT	Quantitative RT-PCR
RdDM	RNA-directed DNA methylation
RDR6	RNA-dependent RNA polymerase 6
RISCs	RNA- induced silencing complexes
RNA	Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease
RT-PCR	Reverse transcription PCR
SAM	Shoot apical meristem
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEM	scanning electron microscopy
siRNA	Small interfering RNA

SPL	Squamosa promoter binding protein-like
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
ssRNA	Single-strand RNA
ta-siRNA	Trans-acting small interfering RNA
TFS1	Thermo-sensitive female sterility 1
TGMS	Thermo-sensitive genic male sterility
TGS	Transcriptional gene silencing
TPM	Transcripts Per Kilobase Million
TPR	Tetratricopeptide repeat
UTR	Untranslated region
UV	Ultraviolet
WT	Wild type
WGS	whole-genome sequencing

Table of Contents

Chapter 1 General Introduction

1.1 Flower development in double fertilized plant

1.1.1 The structure of rice flower

Rice (*Oryza sativa* L.) is the most fascinating monocotyledon plant, which is widely used for research as model plant. The structural units of rice inflorescence include rachis, ciliate ring and peduncle, the primary branch grows on the rachis, and secondary branch grows on the first branches. Florets insert grow on both the primary and secondary branches which is called spikelets. Flowering in rice is a complicated process, initiates from the inflorescence and ends at the floret opening which is also called anthesis.

Rice is monecious plant, with one pistil and six stamens, and a pair of lodicule at the bottom. Additionally, the grass specific organ husk comprises of lemma and palea. Gametophyte development is an important process in flowering plant, microspores and megaspores generated from meiocyte will develop into male and female organs, respectively. The establishment of ABC model is a fruitful achievement that help us better understand flower development, which was raised by Haughn and Somerville in 1988 (Haughn and Somerville, 1988), as a description of genetic mechanism of flower organ establishment. A, B and C represent three kind of genes that control the generation of different components in a flower. A gene was deemed to control the development of sepal, B together with A or C to decide the development of petal and stamen respectively, while the development of pistil was controlled by C (Bowman et al., 2012; Litt and Kramer, 2010). Studies in the past decades have shown that most of the genes controlling rice reproductive organ development belong to MADS-box family (Lopez-Dee et al., 1999; Yoshida and Nagato, 2011).

1.1.2 Female organ development in rice

Rice female organ comprise of three parts, stigma, style and ovary. Rice has only one ovule inside the ovary without funiculus, which harbors the female gametophyte. The ovule primordium is a bulge structure derived from the placenta of the ovary wall, then

elongates and forms the nucellus. An archesporial cell starts to differentiate and develops into megasporocyte, which will form the megaspore tetrad after two meiosis. Three micropylar megaspores will programmed degenerate, leaving the chalazal-most megaspore. Concurrently, the outer integument and inner integument generates from the bottom and enclose the functional megaspore and develops to the embryo sac. The terminates of the integument is defined as the site where pollen tube enter the ovule, named micropyle (Reiser and Fischer, 1993; Yoshida and Nagato, 2011).

The functional megaspore increases in size and undergoes three consecutive rounds of nucleate division to form an eight-nucleate embryo sac. Along with karyokinesis, the vacuole increases in size and pushes the nuclei to the poles from the first karyokinesis. A mature embryo sac is a *Polygonum*-type seven-cell eight-nucleate polarized structure with four cell types, including one egg cell flanking by two synergids located at the micropylar end, two synergids come from a pair of nuclei while another pair of nuclei became an egg cell and a polar nucleus. This polar will fuse with another polar nucleus which differentiate by a chalazal end and migate to the micropylar cytoplasm to form the central cell. Three left nuclei at the chalaza end will became three antipodal cell(Bowman et al., 2012; Yang and Sundaresan, 2000).

1.1.3 Male organ development in rice

Rice male organ consist of anther and filament. When young spikelets develop to stamen and pistil differentiation stage, the stamen primordium differentiates into anther and filament. And the former is comprised by undifferentiated cell. After differentiation, the young anther becomes a four-lobed structure with an archesporial cell in each lobe, separated by septum initials. Which will undergo the first periclinal division and generates primary parietal cell and primary sporogenous cell. The former one will generate the four layers of non-productive tissue surrounding the microsporocytes, from the surface to the interior, including epidermis, endothecium, middle layer and tapetum by two periclinal and anticlinal divisions, while the latter will develop to microsporocyte. The microsporocyte undergoes two successive meiosis and give rise to tetrads of microspores (Wilson and Zhang, 2009).

The mature pollen grain of grass is 3-celled type, generated by two asymmetric mitotic division. The first mitotic forms a sexual cell and a vegetative cell, while only the sexual one proceeds to the second mitosis, which form a tricellular structure that harbors two sperms and a large vegetative cell. Pollen tube is generated by the vegetative cell and delivers the sperms to the embryo sac (Raghavan, 1988).

It has been well documented that the process of anther development involves a complex coordinate interaction between sporophytic and gametophytic tissues. All these somatic cell layers play key roles in pollen development and sperm release, especially the tapetum, researches showed programmed cell death (PCD) happens during the late stage of pollen development. The failure development of any part of somatic cell layers may lead to sterility (Wilson and Zhang, 2009).

1.1.4 The process of double fertilization in rice

Pollination is a synchronous process with flowering, which usually takes place after pollination in different plants. Double fertilization is a highly orchestrated process and controlled by a complicated mechanism in flowering plants, including two dimorphic female gametophytes, the central cell and egg cell, fuse with two non-motile sperm cells and subsequently develop into endosperm and embryo respectively. Double fertilization is a key process to integrate paternal genes into the progenies and the bridge between generations (Berger et al., 2008; Russell, 1992).

Pollen grains landing and adhere to the stigma is the starting of fertilization. Since the embryo sac are well wrapped by the maternal tissues, the sperm need a channel to access to them. The delivery system comprised by two sperm cells and a large vegetative tube cell (Chaudhury et al., 1997; Raghavan, 1988). Along with the germination of pollen grains and pollen tube penetrating through the papillar cells, two immotile sperm reside at the front of the pollen tube followed by a vegetative nucleus. Pollen tube polarized extension through the transmitting tract subsequently emerge on the surface of the placenta and entry into the embryo sac from the micropyle(Feng et al., 2000; Wilson and Zhang, 2009).

There are two main phases in pollen tube guidance, pollen tube extension relies on the communication with the maternal tissue before reaching to the micropyle. However,

the guidance of pollen tube to micropyle and enter the embryo sac to release sperms is believed to be controlled by mature female gametophyte. Especially the synergids, many researches revealed that the filiform produced from synergids extent towards the micropyle and one pollen tube which first target the filiform will pass through this apparatus and enters one of the two synergids. Two sperms and a vegetative nucleus in the pollen tube will be released in the receptive synergid. Other researches also reveals that synergids have function in sperm migration to target zone and gametophyte karyogamy (Higashiyama, 2002; Yang and Sundaresan, 2000).

1.1.5 Embryo and endosperm development

Seed developments starts form the fusion of sperm with central cell and egg cell. The triploid product of the second fertilization develops to endosperm that store carbohydrates and proteins during seed maturation. There are mainly three types of endosperm in angiosperm: nuclear type, cellular type and helobial type, rice endosperm is nuclear type and its development is usually earlier than embryo, including three phases: syncytial, cellularization and differentiation (Berger, 1999; Kranz and Kumlehn, 1999). After fuse with a sperm, the egg cell will develop to the pro-embryo, meanwhile the central cell together with the sperm nucleus will develop to primary endosperm nucleus. The primary endosperm nucleus will move to the antipodal cell. The initiation of endosperm development is the successive divisions of the triploid zygotic nucleus without cytokinesis. The endosperm is a syncytium at the initial stage and isolated by cell wall at the cellularization stage (Grini et al., 2002; Luo et al., 2000). During the differentiation stage, the endosperm develops to a complex structure with three components. The starchy endosperm is the main part of which stores starch, while the aleurone layer locates between seed coat and the starchy endosperm which has functions in starch degradation during seed germination(Curtis and Grossniklaus, 2008). The embryo surrounding region is believed to function in nutrients support and physical isolation for young embryo development. Researchers believed that free nuclei division is more efficient without cytokinesis, which makes the free nuclei endosperm distribute on the surface of embryo sac to rapidly to create a suitable situation for embryo development(Chen et al., 2014).

The fertilized egg will undergo two phases to form a mature embryo, which is diploid. The development of pro-embryo starts from the first asymmetrical division of zygote at 18 hours after flowering, usually occurs horizontally (slightly oblique) to form apical and basal cells which will form a globular embryo. The followed division of both apical and basal cell happen 24 hours after flowering and generate a 4-cell pro-embryo in T shape. After 30 hours, the number of cells in globular embryo exceeds 20. The 200-cell stage seems to be a turning point for the organs differentiation in embryo. Scutellum primordium appears at the top while the coleoptile primordium arises at the ventral of the embryo. At day 4 after pollination, formation of SAM and coleoptile primordium start and the radicle can be recognized almost simultaneously at fixed positions. Different with other dicots in which the SAM usually fix during embryo development, grass has three leaf primordium incorporating into embryo before embryo maturation, indicating that SAM starts to generate young leaf primordium in embryo differentiation stage. 5 days after pollination, the structure of radicle, including root cap, epidermis, cortex, stele and coleorhiza arise and develop at day 6. The first leaf primordium can be recognized and the procambium differentiates to form the connection between scutellum and plumule by hypocotyls (Chen et al., 1985; Kranz and Kumlehn, 1999). From 7 days to 8 days, the second and the third leaf primordium appear consecutively, organs keep enlarging and the SAM becomes dome-shaped. There is no obvious morphologically change of rice embryo after 10 days, except keep enlarging until 20 days after pollination when the embryo will stop growing and become dormant (Sentoku et al., 1999).

1.2 Hybrid rice breeding and two-line system

1.2.1 Advantages of hybrid rice

Rice (*Oryza sativa* L.) is one of the most important crops in the world. In the last half-century, two great breakthroughs in rice yield have been recorded. The first one was referred to the First Green Revolution with the usage of the semi-dwarf gene during 1960s. The second increase started in 1970s in China with the utilization of heterosis by producing hybrid cultivars. Statistics reveal that China occupies less than 10% cultivated land and 5% of water resources, whereas feed 20% of the population in the earth. With such insufficient resources, however, from 1950s to early 1990s, although

nearly half of China's arable lands were reduced, the grain production achieved a 4.5-fold increase, which is much higher than the national populations increase (2.5-fold). Hybrid rice is a kind of rice bred from two different parental varieties, the two parental varieties should have different good agronomic trait and combining ability. Hybrid rice will not only inherit those good agronomic traits from the parent material, but also more advantages over them which is well-known as hybrid vigor, or called heterosis, which is a general phenomenon in biological kingdom. Hybrid rice usually outperform the parents such as has idle morphology, developed root system, large panicles with vigorous growth. Hybrid rice also has satisfied performance in response to stress condition. Research showed that the yield can be raised by increasing the degree of heterosis. Compared to inbred varieties, hybrid rice has significant increase the rice yield by more than 30%. The utilization of heterosis in rice breeding and worldwide crop production to improve both the quality and quantity is the main achievement of modern agriculture.

1.2.2 Male and female sterility in rice

Male sterility which was first reported in 1763 is wide spread in plant, usually refers to the failure in producing functional pollen. To date, male sterility has been reported in over 600 plants and provides important resources to inveatigate the development of anther. Any abnormal development of male gametophyte may cause infertility. Many researches revealed that abnormal sporophytic development mutant may have aberrant tapeta while gametophytic mutant usually has defective microspores, which thus lead to the abortion of pollen grain. Male sterility can be caused by hereditary factor from both cytoplasm and nucleus which can be divided to three kinds, cytoplasmic male sterility (CMS), genic male sterility (GMS) and cytoplasmic-genetic male sterility (CGMS). In CMS and GMS, the mitochondrial genes and a single or several nuclear genes control the infertility respectively, while in the CGMS, the system controlling male sterility includes interactions between mitochondrial and nucleonic genes (Wang et al., 2006).

Mitochondria are semiautonomous organelle in plant which contain fraction of the hereditary factor that is essential for plant development. Sequencing results show that there are only approximately 60 known genes in mitochondria while proteomics

analysis showed that plant mitochondria contain more than 1,000 proteins. Plant mitochondrial genomes sequencing reveals that there are many recombination-activated sequences which may recombine with the genome to create new molecular structures and novel ORFs, including CMS genes (Luo et al., 2013). Many researches demonstrated that some nuclear gene can target to mitochondria and regulate the function of mitochondrial genes and vice versa. The nuclear restoration of fertility genes (Lark et al.) is a kind of well know nuclear genes that can rescue fertility of CMS genes which established the foundation of three-line system (Chen and Liu, 2014; Fan et al., 2016).

With the discovery of environment-sensitive genic male sterility (EGMS) mutants in rice, researchers have a better understanding on effect of environment factors on rice fertility. Researches revealed that some male infertility mutants are sensitive to day length (PGMS), and some mutants are sensitive to temperature (TGMS). Male sterility in EGMS is usually reversible, which starts a new opening of two-line system. It is believed that epigenetic control by noncoding RNAs is involved in EGMS, which can restore its fertility under different growth conditions. Besides, drought was also reported can induce male sterility (Sheoran and Saini, 1996). Although there are many researches have been done on male sterility, only a few genes have been reported, and their functions are still obscure (Zhang et al., 1994).

Male sterility rice plays a key role in three-line system hybrid breeding, whereas female sterility rice seems useless due to its difficulty in generating the progeny. Female sterility is also widespread in plant kingdom, which has the similar classification with male sterility by hereditary factor in cytoplasm or nucleus. Female sterility including any abnormal development during the whole reproduction period may lead to the failure of seed formation, from the female gametophyte differentiation to seed development.

The Asian cultivated rice (*Oryza sativa* L.) can be classified into two main subspecies, *indica* and *japonica*, which domesticated independently from various subpopulations of *O.rufipogon.* Although the hybrid vigor between different varieties within the same subspecies is widely applied in hybrid rice breeding in decades, the heterosis is not always significant in F1. Hybrid sterility between subspecies is the main obstacle to appply heterosis in production. Embryo sac abortion occurs frequently in the spikelet

of *indica/japonica* hybrids which was first reported by Oka in 1957(Kubo and Yoshimura, 2005). To date, many mutants have been reported in both Arabidopsis and rice and studied at cytological level. Central cell mutants have wrong number of polar nuclei in central cell, which may lead to the failure of divisions of the triploid zygotic nucleus failed, and thus lead to form a non-endosperm seed. Egg apparatus mutants have embryo sac with nonfunctional egg cell or without egg apparatus which may generate a seed with embryo deletion. While the embryo sac degenerate mutant has no or abnormal small embryo sac inside where double fertilization cannot process(Li et al., 2007; Robinson-Beers et al., 1992; Zeng et al., 2009; Zhao et al., 2007). The reproductive barrier between subspecies also affect the postzygotic development which may lead to double fertilization interruption (Kubo, 2013; Song et al., 2005; Xiao et al., 2004).

It has been shown that there is a complicated signal network in plants by mechanism studies. However, only 40 genes in rice have been found to involved in hybrid sterility (Long et al., 2008; Rick, 1966). Investigations on hybrid unfertility have proposed two models. The first one is interlocus epistasis model which is also known as the Bateson-Dobzhansky-Muller (BDM) model (Cutter, 2012; Fierst and Hansen, 2010; Kaplan, 2008). The second one named the one-locus allelic interaction model is also called sporogametophytic interaction(Lark et al., 1995; Long et al., 2008). Recently, many studies focus on one-locus model have revealed the unique properties of its molecular structures.

1.2.3 Three-line and two-line system

Conventional rice breeding mainly relies on rice varieties, and usually focuses on selecting idle plant morphology and high tolerance properties. Since rice is a kind of self-pollinating plant, breeders have to remove the functional pollen grains to prevent self-pollination contamination in hybrid rice breeding. Emasculation in rice relies on manual labor, not only the efficiency is low, but also difficulty in avoiding contamination, not to mention realize large-scale mechanized production. With the discovery of CGMS and GMS lines, which cannot generate functional pollen to produce the homozygous progeny, male lines became the idle material for hybrid seed production. The first application of CGMS is in maize in 1950s, which has significant

increased the production of hybrid maize. The commercial hybrid rice was first released in China in 1976 and occupied over half of the planting land since 1980s. The most famous breeding system based on CGMS is three-line system, invented by Yuan Longping in 1970, which requires three different breeding lines: the CGMS line, the maintainer line, and the restorer line. The CGMS line is male sterility which controlled by mitochondria gene and used as the female parent, while the maintainer line which has the same nuclear genome with the CGMS line but has normal cytoplasm genome can serve as the male parent to generate CGMS progeny. The restorer line that has a functional nuclear restorer of fertility (Lark et al.) gene or genes and good combining ability with the CGMS line will serve as male parent to produce hybrid seeds with strong hybrid vigor.

In contrast to CGMS, most GMS mutants are not suitable for hybrid seed production, because their male-sterility traits cannot be efficiently maintained. However, the discovery of EGMS mutants has enabled some GMS traits to be used for hybrid crop breeding. the start of two-line system breeding is the discovery of Nongken 58S (NK58S) in 1973, which is a natural recessive photoperiod-sensitive male sterile line belongs *japonica*. The original NK58S is completely male sterile when grown under long-day conditions but male fertile when grown under short-day conditions. Further study shows that the male sterile is induced by day length is between secondary branch differentiation and microsporogeneisis during spikelet development, which indicates that photoperiod induced sterility is independent with flowering. The reversibility of male fertility in NK58 enables the two-line system hybrid breeding by serving as both the male sterile line and maintainer line. For further application in two-line system, the photoperiod-sensitive trait was introgressed into an *indica* background variety Peiai 64 (PA64), which became a thermos-sensitive line instead of photoperiod-sensitive line. The new mutant shows predominant male sterile in normal rice grown temperature (25-28℃) while fertility rescued under lower temperature (21-23℃) (Ding et al., 2012). The long-day and higher temperature which can maintain the male stertility of EGMS is defined as the restrictive conditions while the short-day or low-temperature conditions which can restore fertility of EGMS is permissive conditions. Since nucleic gene controls the male infertility in EGMS, which means all normal varieties possess the wild-type fertility gene alleles can restore male fertility, and thus they can be used

as the restorer line for hybridization. Therefore, a two-line system simplifies hybrid seed production and reduces costs(Chen and Liu, 2014; Zhou et al., 2012).

The genetic and molecular basis of the EGMS in NK58 and PA64 were both investigated by Dr Qifa Zhang's lab and Dr Yaoguang Liu's lab. Mapping results locate the PGMS locus in NK58 and TGMS locus in PA64 in the same position 22 258 571 of chromosome 12, a single base C-to-G nucleotide substitution. Further research reveals that the mutation is within a 21-nt small RNA, which is preferentially expressed in young panicles, mediates the PGMS and the TGMS trait (Zhu and Deng, 2012). Their results suggest that this small RNA may target downstream gene or genes to response to length and temperature, respectively, in NK58 and PA64. A small interfering RNA will target the promoter region of the lncRNA gene and affect its DNA methylation efficiency in NK58s and WT rice, which will thus regulate the expression of this lncRNA gene. Although it has not yet been identified any target gene or genes for the lncRNA and osa-smR6864 genes, a mechanism revealed by these studies demonstrate that male fertility can be regulated by the interaction between environmental factors and the genetic or epigenetic elements.

1.3 Map-based cloning and MutMap analysis

1.3.1 The genome of *Nipponbare* and 93-11

As one of the most important staple food crops worldwide, rice is among the first sequenced plant species. Rice has the smallest genome of the major cereals, dense genetic map and relative ease of genetic transformation, which make rice as an important model plant for researches. The complete reference genome for a single inbred cultivated subspecies, *nipponbare*, which belongs to *japonica*, was pooled by the IRGSP (International Rice Genome Sequencing Project) from international sequence groups. To date, more than a hundred of plant genomes have been published, the genome of *nipponbare* remains as the only monocot genome that has been sequenced and annotated at high-quality level since its release in December 2002. IRGSP used a bacterial artificial chromosome (Reflinur et al.)-based cloning strategy to sequenced ten-fold coverage of the genome. The genome size of *nipponbare* was calculated to be 388.82Mb, including estimated length of gaps (International Rice

Genome Sequencing, 2005). The centromere is the most complex region for both sequencing and assembling, which contains highly repetitive satellite DNA of 155-165bp within the functional domain at the center and centromere specific retrotransposons in the flanking regions.

Although the high-quality map-based sequence of *Nipponbare* genome has been published by IRGSP, researchers are still making continuous effort to refine the genome assembly and improve the annotation, especially on the centromeres and telomeres. Sequencing comparison of telomere-ends suggesting clearly suggests that the chromosomal ends are highly transcriptional expression region enriched gene. Moreover, researchers have constructed a plasmid library of 14 telomere-ends distributed on 12 chromosomes.

As the representative variety of *indica*, 9311was sequenced by Beijing Institute of Genome using a whole-genome shotgun approach with four-fold coverage of the genome in April 2002. The genome size of 9311 is calculated to be 430Mb. Two major issues for dogging sequenced genomes today are assembly errors and DNA structural defects, especially for *indica* 9311 (Yu et al. 2006). Short-read sequencing technology is not able to resolve the problems at sequence structural level, such as insertions, deletions, translocations and copy number variations. In both rice genomes, the unfilled gaps were found to contain complicated sequences and referred as “dark matter”.

In 2016, the parents of an elite Chinese hybrid, Minghui63 (MH63) and Zhenshan97 (ZS97) which also represent the two major cultivated varieties of *indica* in China and southeast Asia, were sequenced with higher accuracy and coverage using BAC-by-BAC approach, supplemented with Illumina whole genome shotgun. The gradual perfection of rice genome makes it a powerful tool in gene location and functional analysis of some desirable agronomic traits.

1.3.2 Molecular basis of Map-based Cloning and MutMap

To date, there are mainly two approaches to connect the sequence and the function of a gene: forward and reverse genetics. Forward genetics is an approach used to identify the genetic mechanisms of specific phenotype, such as genetic mapping to locate the target gene on chromosomal region. It is the classic hereditary method to investigate

gene function, which is relies on particular mutated phenotypes and genome database. While reverse genetics rely on disruption of genome sequence information to investigate genes function. Chemical and radiation mutations have been widely used in gene function researches, which typically cause single base-pair mutation and are different with the T-DNA insertion.

Positional cloning (Map-based cloning) is an indirectly approach (forward genetics), raised by Alan Coulson at Cambridge university in 1986. The strategy of map-based cloning is to use markers distributed in the whole genome to walk down on one chromosome and eventually find the marker which tightly links to the locus controlling the phenotype, thus locate the gene in a small region (Lukowitz et al., 2000). Sequencing of this small region will then be used to identify the mutation locus. The genetic background information of the research material plays a key role in map-based cloning by providing high quality markers. Indicating that greater difference of genetic background between two crossing parents will provide higher density of molecular markers.

MutMap is a different cloning approach by cross the mutant line with a remotely related wild-type plants, which makes the number of segregating loci responsible for the phenotypic change is minimal, the derived F2 plants of the mutant phenotypes will pooled and sequenced. SNPs called from two parents will be used for analysis(Chen et al., 2014; Patel et al., 2012; Takagi et al., 2015).

1.4 Functions of Argonaute proteins in small RNA mediated gene expression regulation

Since its discovery in *nematodes*, microRNA (miRNA) was firstly treated as useless genes, However, it becomes the focus of research and initiates a new insight in gene expression regulation. Researchers found that miRNA involved in many important physiological and pathological pathways genes expression such as plant growth, formation of tumor and immune response. miRNA-mediate gene regulation pathways include a series of protein, including Argonaute proteins which are component of RISC (RNA-induced silencing complex). RISC is responsible for the RNA interference (RNAi) gene silencing phenomenon. Except miRNA, small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs) are also involved in mRNA cleavage or

translation inhibition by binding AGO protein and guide AGO proteins to target genes through sequence complementary (Mallory and Vaucheret, 2010).

AGO proteins are highly conserved and can be divided into two subfamilies in eukaryotic: the AGO and the PIWI. AGO proteins are typically composed of 700~1200 amino acid residues, having a molecular weight of ~100kDa. AGO proteins are conserved and expressed in a wide range of species and tissues: yeast has only one AGO protein, Drosophila possesses 5 AGOs and worm has 27 AGOs. Plant Argonaute proteins are all in the AGO subfamily. The AGO genes have been identified in several model plants, their results show that the number of AGO genes also varies greatly among different plants: Arabidopsis has 10 AGOs, maize has 17 AGOs whereas rice has 19AGOs. As the core component of RISC, Argonaute proteins can affect the growth and development as well as the responses to abiotic and biotic stresses. Further studies reveal that AGO proteins not only regulate gene expression, but also participate in miRNA biogenesis (Mi et al., 2008).

1.4.1 Functional domains of Argonaute proteins

The structure of Argonaute was first determined by x-ray in the archaebacterium *P.furiosus*. In the past decades, high-resolution structures of human AGO2 with guide RNA has been reported and made as the typical structure of AGOs. Although there is no crystal structure of plant AGOs, the reposted structure of AGOs in plant are highly similar. Previously study about AtAGOs shows that the MID (Middleton et al.) domain crystal structure is similar with human AGO2. AGO proteins consist of six domains: a variable N-terminal domain, a conserved PAZ (PIWI-ARGONAUTE-ZWILLE), MID, PIWI and two linker domains flanking PAZ domain.

PIWI and MID domain are at the C terminal and has a conserved triad composed motif, DDX (Asp-Asp-Asp/Glu/His/Lys), in the catalytic center, which is similar with RNase H family of endonuclease. This similarity enables parts of AGOs to have cleavage activity, but different with RNase H family of endonuclease that using DNA as template, PIWI prefers ssRNA (single strand RNAs) as template when targeting mRNA. MID have a rigid loop that can recognize the 5'-phosphate of siRNA or miRNA. The siRNA or miRNA are anchored at the conserved pocket at the interface between PIWI and MID domain of AGOs. The MID domain contains a MC domain, which is

homologous to cap structure binding motif, and is necessary for efficient regulation of translation (Hutvagner and Simard, 2008) .

N-terminal domain is less investigated, in vitro studies on human AGOs revealed that N-terminal domain was involved in unwinding the siRNA or miRNA: miRNA* duplex during RISC assembling and required for small RNA loading. The changes of conformation of AGO protein will drive the wedging of the N domain between the two duplex strands, which will thus result in opening and unwinding of the miRNA: miRNA duplex. It is also revealed that the unstructured loop in N-terminal is necessary for cleavage but the mechanism is not clear.

PAZ domain is found in both Argonaute and dicer proteins, both of which play key roles in RNAi. Researches revealed that the PAZ domain interacts directly with small RNA in RISC by forming a OB fold (oligonucleotide/ oligosaccharide-binding) which contains a central cleft line that can recognize the 3'-overhangs thus bind and anchors the ssRNAs(Liu et al., 2011). Unwinding of the duplex which is slicer-independent also needs the contribution from PAZ domain (Hutvagner and Simard, 2008).

1.4.2 Loading of sRNAs onto Argonaute proteins

Previous studies on Drosophila and human revealed that AGOs not only regulate the expression of genes by forming RISC with sRNA, but also involved in miRNA and siRNA formation. Small RNAs are a short double-stranded RNA of about 20-24nt in length and widely existed in plant and animals which are produced form long hairpin RNA or long double strand RNA by DICER proteins thus generates (Fang and Qi, 2016) . There are two different RNA-binding sites in dicer protein, the long dsRNA binding site and siRNA rebinding site, which are responsible for the cleavage of long dsRNA and rebinding of the cleaved of siRNA, respectively. Asymmetry sensing takes place during the rebinding of siRNA.

In plants, different small RNAs are generated by different DCLER proteins, miRNAs processed by DCL1, ta-siRNAs produced by DCL4 whereas DCL3 is responsible for the formation of hc-siRNAs (heterochromatic siRNAs). However, DICER and its partner proteins are not always necessary for RISC formation in Drosophila Ago1-RISC and human Ago2-RISC. Further researches revealed that DICER directly interact

with AGOs in determining which strand in the duplex will be the guide RNA and finally form RISC leading to the strand with less stable paired 5′ end is preferentially loaded into AGO proteins.

Mature miRNA-loaded AGOs form RISC together with Dicer and its partner protein, which needs chaperone HSP90 (heat shock protein 90) and its cochaperones that can modulate the ATP-hydrolyzing rate to keep AGO proteins in an open state, allowing them to accommodate the RNA duplex. It has been found in different species. In Drosophila, HSP90 and HSC70 can form the duplex HSC70/HSP90 which is proved functional in siRNA duplex binding, thus effect the formations Ago1-RISC and Ago2-RISC. Similar mechanism is also found in Arabidopsis, whereas the cochaperone of HSP90 is needed to accelerate the assembly of RISC with AGO1. After loading onto AGOs, only one strand of the sRNA duplex will be remained as guide RNA while the other strand will be removed by the endonuclease. In summary, a dsRNA-binding protein, an AGO protein and the HSC70-HSP90 system are the minimal components of RISC complex. However it is notable that some species-specific or sequence-specific factors will affect the RISC loading (Fang and Qi, 2016).

1.4.3 Different sRNA sort into specific AGO complex

AGOs can load different sRNAs to perform specific functions. The mechanisms of how AGOs choose sRNAs are varies in different species. Studies reveal that Ago1 prefers to load miRNAs while Ago2 choose siRNAs by determining the structure of the duplex in Drosophila. Studies in Arabidopsis show that the 5’ nucleotide is involved in determining AGOs: miRNAs with a 5’U is preferentially selected by AGO1 and siRNAs with a 5’A by AGO2, AGO4 and AGO5 mainly associate with hc-siRNAs bearing a 5’A and sRNAs with a 5’C, respectively. In Arabidopsis, miR390 which has a 5’U is recruited by AGO7 but not by AGO1, miR165/166 with 5’A is associated by AGO10 but not by AGO4, suggesting that the identity of 5’ is necessary but not sufficient for sRNA loading to AGOs, additional mechanisms exist to regulate sRNA sorting (Montgomery et al., 2008; Zhu et al., 2011). Previous research revealed that loading of specific miRNA into AGO1 or AGO2 are partially dependent on the structure of miRNA duplex which is recognized by PIWI domain. The recognition relies on the QF-V motif and DDDE catalytic tetrad (Zhang et al., 2014). AGO1 branch

proteins incorporate the cmiRNAs initiating with 5'U, following the loading rule directed by 5'nucleotide (Fang and Qi, 2016). However, AGO4 protein recruit the DCL3-dependent lmiRNAs despite the base of the their 5' end. Indicating that the of 5' terminal nucleotide is less important than the enzymatic machinery for the process of micro RNAs. (Inui et al., 2010). Besides, AGO1 and AGO4 can recruit both 21- and 24-nucleotide sRNAs, which indicates the length of the sRNAs is unlikely a determinant factor (Chitwood and Timmermans, 2010).

Collectively, these researches reveal AGO recruit the certain sRNA is controlled by three factors, the first one is nuclear base of the 5' end, the second one is the sRNA duplex's structure that recognized by PIWI, the last one is signals from the upstream biogenesis machinery. For the last two determined factors, mechanisms are remained unclear (Moore et al., 2014).

1.4.4 Functions of Argonaute protein and RISC

Previous studies reveal that AGOs are essential to RNAi in eukaryotes by forming RISC. Since the first identified AGO has cleavage activity, AGOs were referred as slicer proteins for a long time. Although the structure of eukaryotic AGOs are conserved, only part of the AGO proteins have cleavage activity, the other AGO proteins regulate gene expression by a different mechanism. In animals, miRNA usually control gene expression by preventing RNA translation in cytoplasmic. without cleavage activity in this approach and the target RNA can duty-cycle operation in this regulation. The function of AGOs in plant are quite diverse, studies in Arabidopsis demonstrated that there are two specific pathways in regulation of gene expression by AGO protein and sRNA, including gene regulation at the transcriptional level through directing DNA methylation, which is well known as TGS (transcriptional gene silencing) and PTGS (posttranscriptional gene silencing), which functions in gene regulation at the posttranscriptional level through endonucleolytic cleavage (Mi et al., 2008).

In plants, AGO1, AGO2, AGO4, AGO7 and AGO10 are illustrated to have slicer activity that catalyzes the cleavage of target RNAs under the guidance of sRNA. Endogenous classes of sRNA trigger TGS and PTGS depends on their length, 21- and 22-nucleotide in length including miRNAs, ta-siRNAs and nat-siRNAs can direct

PTGS. While the 24-nucleotide sRNA including p4- and p5-siRNAs can direct DNA methylation and TGS (Rubio-Somoza et al., 2009).

In plant, it has been reported that high extent of sequence complementary to their target mRNAs for miRNAs is required, which mediate the cleavage of target mRNA by Argonaute proteins. Whereas the degree of sequence complementary in animal is relatively low, and the way miRNAs regulate gene expressions is different, usually through translational inhibition without cleavage. These differences between plant and animal suggest a switch function between mRNA cleavage and translational inhibition for the degree of complementary between a sRNA and its target. On the contrary, other studies have proved that many miRNA from plants can suppress translation. Recently, a well-designed study showed that translation of target genes is prevented by miNRAs on the endoplasmic reticulum. An integral membrane protein, Altered Meristem Program1 (AMP1), associates with AGO1 and plays a key role in translational repression mediated by miRNA (Kidner and Martienssen, 2005; Mallory and Vaucheret, 2010). Interestingly, both AMP1 and AGO1, which are compenents for translational repression, are not always necessary for target mRNA cleavage directed by miRNAs, suggesting that these two processes take place in different cellular compartments. The cleavage of at least some target transcripts by miRNA is reduced by the deficiency of isoprenoids synthesis, a chemical which is necessary for many processes such as sterols generation, indicating that activity AGO1 involved target cleavage is concerned with membrane. Nevertheless, it is not clear yet about the identity of different organelle for miRNA directed target mRNA cleavage (Zhai et al., 2016).

1.4.5 Clades of Argonaute proteins

The *AGO* gene family have been identified in several model plants and the number of *AGO* genes also varies greatly among different plants: There are 10 AGOs in Arabidopsis and 19 AGOs in rice. These AGOs have diverse functions which involved in various molecular function and biological processes. There are mainly three clades in Arabidopsis and an additional clade in rice with an evolved AGO18 (Wu et al., 2015).

Clade AGO1/AGO5/AGO10 and Clade AGO18: Arabidopsis has only one AGO1 while rice contains four AGO1 homologs (AGO1a, AGO1b, AGO1c, and AGO1d) that

are all ubiquitous expressed. Loss of function of these AGO1s show multiple effect on the development of mutants. AGO1s can mediate genes expression in almost the whole growth period. Further researches reveal that AGO1s are also involved in defense against virus by recruiting some virus derived siRNAs. Both rice and Arabidopsis has only one AGO10 which is mainly expresses at meristem and leaf primordial, which plays important roles in the maintenance of stem cells of shoot apical meristem (SAM) and the establishment of leaf polarity (Montgomery et al., 2008). Arabidopsis has only one AGO5 while rice contains five AGO5 homologs (AGO5a, AGO5b, AGO5c, AGO5d and AGO5e). AGO5 in Arabidopsis is associate with megagenesis, which has function the somatic cells cover the megaspore mother cell. Whereas in rice, AGO5c is specifically expressed in germ cells and controls the meiosis during sporogenesis. AGO18 is an evolved clade in monocots which is expressed in seedling. There is only one AGO18 in rice while maize has two AGO18 homologs (AGO18a and AGO18b).

Studies in Arabidopsis and rice reveal that both AtAGO10 and OsAGO18 can compete with AGO1 for specific miRNAs and abrogate their activities through sequestration. For example, AtAGO10 predominantly binds miR165/166, which prevents their recruiting by AGO1, thus derepress the HD-ZIP III family genes that are mediate by miR165/166 to control meristem growth (Juarez et al., 2004). In inflorescences, AGO10 will combine with vsiRNA and together with AGO1 to play a modest antiviral function (Ji et al., 2011).

The expression level of *AGO*s can also be regulated through miRNA pathway, AGO 1 encodes mRNA that is targeted by miR168. Virus infection induced OsAGO18 preferentially associates with miRNA168, results in decrease of AGO1 cleavage directed by miR168, which thus leads to the accumulation of AGO1 protein and higher antiviral defense. Although AGO10 from Arabidopsis and AGO18 from rice have structure which is necessary for cleavage, both AGOs bound by miR165/166 and miR168, respectively, are not shown active in degrading mRNAs (Yang et al., 2013). According to these findings, the catalytic activities of these AGOs should be suppressed when they are decoying miRNAs. However, the inhibition mechanism need to be further investigated. To date, there are three envisions about this mechanism. The first one is that Arabidopsis AGO10 and rice AGO18 have weaker intrinsic catalytic activities than AGO1. The second propose is not be demonstrated interacting protein

which may inhibit the catalytic activities. The third one suggests that AGO10/AGO18 and AGO1 recognize different miRNAs in subcellular compartmentalization, which will prevent these two AGOs getting close to their target mRNA (Axtell et al., 2006; Baumberger and Baulcombe, 2005; Kidner and Martienssen, 2005).

Clade AGO2/AGO3/AGO7: Both rice and Arabidopsis have only one AGO2 which is important in plant virus resistant. In vivo study revealed that AGO2 can recruit exogenetic vsiRNAs and catalyze target mRNA cleavage, thus repress the copy of viral. Through translational inhibition, AGO2 inhibit a Golgi-localized gene MEMB12 directed by miR393b, which lead to the extracellular secretion of PR1 protein and higher antimicrobial propery. AGO3 is able to bind siRNAs and cleave viral RNAs in vitro but little is known about its working mechanism. AGO7 specifically recruits miR390 which is uniquely adapted to initiate ta-siRNAs biogenesis. Studies in Arabidopsis, rice and maize reveal that *TAS3* ta-siRNAs are dependent on AGO7. *TAS3* ta-siRNAs target several ARF (AUXIN RESPONSE FACTOR) genes and associate with HD-ZIPIII genes expression, both of which are important in developmental timing regulation.

Clade AGO4/AGO6/AGO8/AGO9: Rice encodes four AGO4 homologs (AGO4a, AGO4b, AGO15 and AGO16) which can bind lmiRNAs (long miRNAs) and 24-nucleotide siRNAs. Function analysis by using *AGO4a* and *AGO4b* knockdown mutant revealed that AGO4 may affect DNA methylation level by binding 24-nucleotide siRNAs. AGO4 is also involved in plant defense against DNA virus by mediating DNA methylation level. AGO6 has similar function with AGO4 which are all required for DNA methylation at most of their functional site. AGO4 is widely expressed throughout plant while AGO6 is mainly express in meristem or immature tissues and researchers revealed that they might have sequential function (Fang and Qi, 2016; Wang et al., 2011).

AGO8 and AGO9 genes is next to each other in genome and their sequences of have a highly degree similarity, except AGO8 has a deletion in the middle of PIWI domain, which probably excludes the correct folding of AGO8. However, it has been thought to be nonfunctional because the expression of AGO8 has not been discovered in any tissues for now. while AGO9 is detected peculiar expressed in cytoplasmic of somatic

cells in both rice and maize, which has function in cell fate controlling by preventing the somatic fate in germ cells. Studies in rice reveals that AGO9 prefers 24-nt siRNAs that is generate from TEs and thus repress its function in female gamete and their accessory cells.

1.5 Objectives of this study

The goal of this project is to clone a thermo-sensitive female sterility gene (*TFS1*) and to reveal the possible mechanism by which *TFS1* controls the double fertilization and thus the female fertility in rice. Double fertilization is a highly orchestrated process and controlled by a complicated mechanism in flowering plants. Understanding its mechanism has direct implications in agriculture, such as hybrid rice production. However, unlike the male sterility rice, which has been focused in hybrid rice study for a long time, little is still known about the female sterility and its application in hybrid rice breeding.

Male sterility line plays a fundamental role in hybrid rice breeding, whereas female sterility line seems useless due to its extreme difficulty in maintaining a homozygote line. We have identified a rice germplasm showing female sterility from progenies of a crossing between *japonica* and *indica* varieties. Interestingly, further experiments have shown that it is thermo-sensitive female sterility line while day length has no effect on its fertility. Genetically analysis result shown that it is controlled by a single recessive gene that we named as temperature-sensitive female sterility 1 (*TFS1*). Pollen viability testing demonstrated that there was no significant difference between *tfs1* and 4266 plants on pollen morphology and activity. Cytological studies revealed that development of embryo sac and pollen tube germination in *tfs1* was normal compared with 4266. However, egg cell and central cell were still observed after pollination for more than 3 days, indicating the double fertilization is blocked in *tfs1* mutant. In addition, we have located the mutant position to a 2.17-Mb DNA region by fine mapping. And find a candidate gene by whole genome sequencing. In this proposed study, the cytological mechanism of double fertilization in *tfs1* will be investigated. At the same time SNP and Indel markers inside this region will be developed for the fine mapping using a F2 population with more than 10,000 individuals of *tfs1*-nipponbare cross. Besides, whole genome sequencing and MutMap analysis will also be used to

accelerate the cloning of *TFS1* gene. Finally, complementary experiments and other molecular and reverse genetic approaches (including RNA-seq, His-CLIP, transgenic method, mutant method and so on) will be used to study the possible mechanism of *TFS1* in controlling double fertilization.

Hybrid rice is a critical but still labor-costing technology in modern agriculture. Although different kinds of hybrid technologies have been well developed, mechanized production of F1 hybrid seed is still far from application due to the selfing of restorer line. Farmers need to inter-plant male sterility line and restorer line and cut off the restorer line after pollination, which will reduce hybrid efficiency and, importantly, is labor intensive.

Thus, how to decrease the cost of making F1 hybrid seed has been the focus of hybrid rice study for decades and with little progress. Different techniques for mechanized production of F1 seeds have been proposed. But none of them has been widely applied because of the higher cost than the traditional way. Female sterility line has great potential in solving this high-cost problem. Female sterility restorer line which could have been a perfect way to realize the mechanized production of F1 seed is not successful too due to seedless of this kind of restorer line.

Thus, the outcome of this research shall help us better understand the genetic control of rice reproductive processes and the mechanism of female sterility with an application value in hybrid rice breeding.

Chapter 2 Cytological mechanism investigation of a thermos-sensitive female sterility rice mutant *tfs1*

2.1 Introduction

As the most important monocot model plant, the developmental studies of rice focus on the morphological changes and staging systems have been made great achievement, which reveals the complex and well-organized system in plant life cycle. Studies on molecular function and genetic mechanism that based on stage-specific defective plants have supply new insights into rice development. The reproductive phase includes a complicated regulating process. Fertility transformation in response to environmental cues is widely found in different species. As mentioned above, temperature and day length are two main elements involved in plant fertility control.

Similar with male sterility, female sterility that due to the abnormal development of ovule, embryo sac and embryo is also widely found in higher plant, such as rice, soybean, Arabidopsis and so on, which is due to the abnormal development of ovule, embryo sac and embryo (Huang and Sheridan, 1996; Reiser and Fischer, 1993; Rotman et al., 2003). Accord to the mutation position, female sterility in plants is divided into three groups. The first one has incomplete pistil or without any pistil (Fladung et al., 1991; Kerstetter et al., 1997). The second one displays an abnormal embryo sac due to the unsuccessful development of ovule or magegametophyte (Huang and Sheridan, 1996; Zeng et al., 2007). The last but not least one has complete pistil, however, the embryo development is defective after fertilization (Nowack et al., 2006; Tsukamoto et al., 2010). Analysis of two rice T-DNA insertion mutant *osapc6* and *osmads13* shows that *OsAPC6* is involved in the continuity of the cell cycle in the process of female gametophyte formation (Kumar et al., 2010) and *OsMADS13* is involved in the process of ovule formation (Dreni et al., 2007; Lopez-Dee et al., 1999), both of them mutate may lead to female unfterility. Female sterility in rice usually occurs in the progeny from *indica* and *japoniaca* hybrid (Li et al., 1997; Wan et al., 1996). The formation of abnormal embryo sac is mainly due to the development of ovule or spore mother cell is arrested, Song et al (2005) carried out QTL analysis of the progeny of *indica* and *japonica*, find an embryo sac development related QTL. By gene cloning for *indica* IR24 and *japonica* Asominori hybrid rice, they find three genes came from three different parents which controlled the fertility of progeny, and epistatic effect

determines the absence of female sterility (Song et al., 2005). Since synergids play a key role in pollen tube attraction and sperm release, many synergid related mutants can also cause female sterility (Kasahara et al., 2005). As reported previously, an *Arabidopsis* mutant *feronia* disrupted the female gametophytic control of pollen tube reception. When the pollen tube reach the receptive synergid in embryo sac of a *feronia* mutant, it is keep extension without release the sperm cells, and invaded the embryo sac, subsequently double fertilization didn't occur (Pagnussat et al., 2005). In addition, there are also many genes related to development of embryo and endosperm thus affect the seed formation, such as *DEMETER* in *Arabidopsis*, which has function for endosperm gene imprinting (Raghavan, 2005). Since the female gametophytes of rice are deeply embedded into the maternal tissues, which made the observations of embryo sac development of double fertilization process are not as easy as that observed in Arabidopsis. However, we hope that as the development of micro observation technologies, the detailed embryo sac development and double fertilization processes are able to be investigated in rice like that in Arabidopsis.

In the present study, we have identified a female sterility line from the F9 generation of *indica/japonica* crossing variety named 4266. Crossing and segregation experiments indicated that female sterility of this mutant was controlled by a recessive nuclear gene. Our cytological studies revealed that the pollen availability, as well as the development of embryo sac in *tfs1* is normal, similar with 4266. Embryo development analysis, together with pollen guidance experiment, reveal that the double fertilization was stopped at the step of pollen tube interact with synergid cell. Together, our cytological studies on *tfs1* have tentatively illustrated sterility mechanism.

2.2 Materials and Methods

2.2.1 Plant materials and field growth condition

4266 are hybrid variety with 70% of *indica* background and 30% *japonica* background. *tfs1* is a female sterility mutant separated from 4266 populations. The rice plants examined under natural field conditions were grown in the Baiyun Base of Guangdong Academy of Agricultural Sciences, Guangzhou, China. For lower temperature condition (25-27°C) and long day condition (>14h) at inflorescence stage, both 4266 and *tfs1* were grown in greenhouse which can control the day length and temperature in The Chinese University of Hong kong, Hong kong, China.

2.2.2 Pollen sterility detection by I-KI staining

Pollen sterility was determined from at least 1000 pollen grains from both *tfs1* and 4266. The florets were taken in the morning, about two hours before flowering, in which the anthers were mature and ready for anthesis. The pollen grains were stained by 1% iodine potassium iodide (I_2-KI) solution and viewed under light microscope (Palmer et al., 1978).

2.2.3 Embryo sac observation by paraffin section

Fixation: Pistils were taken in the morning, about two hours before flowering, in which the embryo sacs were mature and ready for fertilization. Pistils were fixed in FAA (formaldehyde: acetic acid: 50% ethanol = 3.7: 5: 91.3) for at least 24 hours at 4°C, then washed by 70% ethanol three times. Dehydrated the sample sequentially in 80% ethanol, 90% ethanol and 100% ethanol, each for 30 minutes at room temperature, followed by immerse the tissue in 100% xylene three times for 20 minutes each. After that, the samples were embedded with paraffin at 58 °C.

Staining: The samples were cut at 5-15 μm thick by rotary microtome. Float the sections in 45°C water and mount onto slides. The slides with paraffin-embedded section need to be dried overnight at 50°C oven. Before staining, the samples are dewaxed by 100% xylene for two times, each for 5 minutes each. Then hydrated the tissue in a 5% decreased graded ethanol series from 100% to 70% ethanol, two minutes for each. The tissues were first stained by 1% Safranin O solution (dissolved in 50% ethanol) for 1hour, and then washed by tap water. Dehydration by 20% increased graded ethanol series from 30% ethanol to 90% ethanol and stained by 0.1% Fast Green (dissolved in 95% ethanol) for 5 seconds, then washed by 100% ethanol twice and clear in 50% xylene and followed by 100% xylene. The sample were sealed by cover slides using Permount TM mounting medium after the slides were complete dried and observed under light microscope (Nikon E80i) (Duckett et al., 1994).

2.2.4 Embryo sac observation by CLSM

Fixation: Pistils were taken in the morning, about two hours before flowering, in which the embryo sacs were mature and ready for fertilization. Pistils were fixed in 2.5% glutaraldehyde in PBS (sodium phosphate buffer, pH=7.2) for at least 24 hours at 4°C,

then washed by PBS three times. Dehydrate the sample sequentially in 30% ethanol, 50% ethanol and 70% ethanol at room temperature, each for 30 minutes, and dissect the pistils in 70% ethanol under stereoscope to remove the stigma.

Staining: Samples were hydrated sequentially in 50% ethanol, 30% ethanol and distilled water followed by 2% aluminium potassium sulphate (which can allow the dye to enter the embryo sac more readily), each for 20 minutes. The dye used for staining is 10mg/L Eosin B ($C_{20}H_6N_2O_9Br_2Na_2$, FW 624.1) dissolved in 4% sucrose solution and then stained overnight at room temperature. The samples were washed by 2% aluminium potassium sulphate and twice by distilled water and then dehydrated by 20% increased graded ethanol series from 30% ethanol to 90% ethanol. After that, the samples were transferred to 50% methyl salicylate diluted by ethanol for 1 hour and placed the sample in pure methyl salicylate for transparentize the sample.

Embryo sac scanning: The sample were separately placed to a cultural dish with a cover slide at the bottom separately for scanning under Confocal Microscope (Olympus FV1000 IX81-SIM) with mission light wavelength between 550~630 nm. For the orientation of the pistil, try to keep the ovary parallel with the cover slide and immerse in methyl salicylate. The thickness for scanning and number of figures depends on the sample, which need to be observe for the whole structure including antipodal cell, central cell and egg apparatus (Zeng et al., 2007).

2.2.5 Pollen germination and pollen tube extension observation

Pollen germination in vitro: Over 1000 fresh pollens were used for germination detection. The florets were taken in the morning, about two hours before flowering, in which the anthers were mature and ready for anthesis. Shake the floret above the solid medium (15% sucrose, 0.1% agar, 0.2 mM $Ca(NO_3)_2$, 0.4mM H_3BO_3) and take photo by using the Stereo-microscope (Olympus SZX16) at room temperature (Fan et al., 2001; Hong-Qi and Croes, 1982).

Pollen germination and pollen tube extension in vivo: Florets with stem (cut at the bottom) were taken in the morning, about two hours before flowering and placed in water, in which the anthers were mature and ready for anthesis. Removed the anther gently and avoid damaging the stigma, while some florets were remained the anther.

When the florets with anther were flowering, shake the florets above the emasculative floret and checked the stigma by stereoscope quickly. The pollinated pistils were sampled at 5 minutes and 30 minutes and fixed in ethanol and acetic acid mixed (3:1) solution for at least 30 minutes. After which, immersed the pistils in 1M KOH solution overnight at room temperature for softening. The washed the samples by distilled water for 3 times and stained by 0.1% aniline blue dissolved in K_3PO_4 buffer (PH=8.5) over two hours at room temperature. Washed the sample by distilled water for 3 times and immersed them in 50% glycerol diluted by K_3PO_4 buffer (PH=8.5). Place a single pistil on the slide and sealed by cover slide, gently press the cover slide to make the pistil mounted in glycerol, which was ready for observation under the fluorescence microscope (Nikon E80i) (Chhun et al., 2007; Fujii and Toriyama, 2005; Martin, 1959).

2.2.6 Seed formation observation after pollination by CLSM

Pistils were taken at one day after pollination (1DAP), two days after pollination (2DAP), three days after pollination (3DAP), four days after pollination (4DAP), and five days after pollination (5DAP). The fixation, staining and embryo sac scanning are the same with embryo sac observation mentioned above.

2.2.7 Floral primordial development observation by SEM

For scanning electron microscopy (SEM), young panicles at about 1mm to 1cm length were taken and were fixed in 2.5% glutaraldehyde in PBS (sodium phosphate buffer, pH=7.2) for at least 24 hours at 4°C, then washed by PBS three times. Dehydrate the sample sequentially in increased concertation of ethanol for 30 minutes each at room temperature. The samples were dried in tetramethylsilane, then decant the solution and dry the sample at 28°C for more than 1 hour, which was then sputter-coated with platinum, and observed by the scanning electron microscope (Hitachi S-4000, Tokyo) at an accelerating voltage of 15 kV (Li et al., 2010; Li et al., 2009; Prasad et al., 2001).

2.3 Results

2.3.1 *tfs1* is thermos-sensitive female sterile mutant with defective hull

tfs1 mutant was identified in the population of F9 generation named 4266 from a multi-crossing of four different cultivars, which was a fully sterile germplasm with abnormal husk grown under normal condition during field production. We didn't get any seed of these sterility lines at the first time, but lucky we found another 20 sterility plants from the other three lines in this F9 generation. The total number of rice plants from these three lines is 63, indicating that female sterility is controlled by a single recessive nuclear gene accord to the genetic segregation. Tillers from these sterility lines were transferred to another field irrigated by cool water. Interestingly, the seed setting rate was increased to 20%, and the seed from which still displayed a fully sterility phenotype under normal condition.

As shown in figure 2-1, the florets form both 4266 and *tfs1* were compared at different stage. The size of florets and the length of hulls form *tfs1* were similar with 4266 plant, except a defective lemma and palea (Fig.2-1, A1-F2). Anthers form both 4266 and *tfs1* in the floret had the same height, as well as the similar filament (Fig.2-1, A1-F2). At the later stage of seed development, the pistil dried but no seeds were produced in *tfs1* under normal condition (Fig.2-1 G). While in lower temperature condition, *tfs1* can produce seed, and three kinds of seeds were discovered. The first one was the needle like seed with defective hulls, part of which was capable of germinating (Fig.2-1H). The second kinds of seeds had the same size of 4266 seed except a slightly incomplete hull (Fig.2-1 I-J). The third one was just the same with 4266 seed (Fig.2-1 K). The last two kinds of seed were found to germinate normally. Then the panicles from 4266 and *tfs1* mutant plants under normal conditions at mature stage were examined and compared with those from *tfs1* rice grown at lower temperature (Fig. 2-2 A-D). As shown in Figure 2-2, the size of panicle and floret number from a panicle were similar between 4266 plant and *tfs1* mutant plant under normal condition (Fig. 2-2 A, D), the panicle from *tfs1* grown under normal condition was 100% sterile (Fig. 2-2 D). However, when the plants were grown under lower temperature (25-27℃), the fertility of *tfs1* mutant was partially rescued, the seed setting rate of this condition ranged from 20 to 50% among different individual plants (Fig. 2-2 B, C).

To further explore the phenotype difference between 4266 and *tfs1* plants, which may provide more valuable information for mechanism studies, we have collected panicles at different stage from these two phenotypic plants. Our morphological studies have revealed that although the size and floret number of the panicle were highly similar, the

defective lemma and palea phenotype of *tfs1* plant was observed at the very beginning stage of panicle development (Fig. 2-3), indicating that the phenotype of defective hull, together with the female sterility phenotype, were determined at the spikelet primordium formation stage.

To further illustrate the seed formation process of *tfs1*, ovaries from different stage after pollination were studied. As shown in figure 2-4, for 4266 plant, ovaries swell gradually and filling normally from 1DAP to 14 DAP. On the contrary, ovaries of *tfs1* did not swell at all and became dry and dead at 14 DAP. These results indicate that the development of embryo and endosperm is abnormal arrested in *tfs1*. Temperature gradient experiments were performed to determine the optimal temperature for *tfs1*, which would maximally rescue the fertility of *tfs1*. Our results shown that when *tfs1* plants were grown under 25-27℃, the fertility recover efficiency of *tfs1* was greatly improved, ranged from 20% to 50%, both higher or lower temperature cannot rescue *tfs1* fertility or in very low ratio. Besides, further experiments revealed that day length had no effect on *tfs1*'s fertility.

It is well known that both male and female sterilities are frequently found in progeny populations of *indica-japonica* cross, the incompatibility between *indica* and *japonica* and so-called hybrid breakdown. We have crossed *tfs1* as male parent with both *nipponbare* and 9311, which belongs to *japonica* and *indica* respectively. F1 seeds were harvested with normal husk and sterility individuals with abnormal husk were found again in F2 populations due to segregation. However, we could not get any seed when crossed *tfs1* as female parent with both *nipponbare* and 9311. These results indicate that *tfs1* is thermos-sensitive female sterile mutant.

Figure 2-1. The florets and seed of *tfs1* and 4266 in normal and lower temperature.

(A1-E1) Florets of *tfs1* before flowering at different stage (time order). (F1) Floret of *tfs1* at flowering. (A2-E2) Florets from 4266 before flowering at different stage (time order, correspond to *tfs1*). (F2) Floret of 4266 at flowering. (G) Aborted seed from *tfs1*'s in normal condition. (H) Needle like seed grain from *tfs1*'s at lower temperature. (I-J) Seed grain from *tfs1* at lower temperature. (K) Caryopsis form 4266 in normal condition.

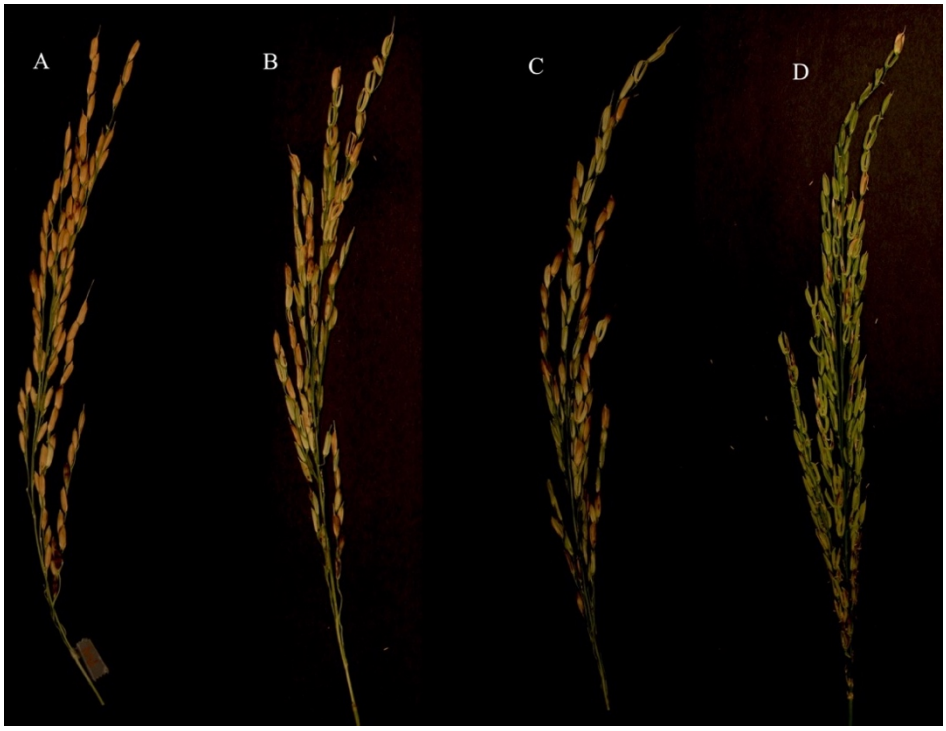

Figure 2-2. Phenotype of matural panicles of *tfs1* and 4266 grown in normal and lower temperature conditions.

Panicles of 14 days after pollination in normal and lower temperature were collected for photographing. (A) Panicle of 4266 in normal rice growth condition. (B-C) Panicles of *tfs1* which have partially rescued fertility grown at 25-27℃. (D) Total sterile panicle of *tfs1* in normal rice growth condition.

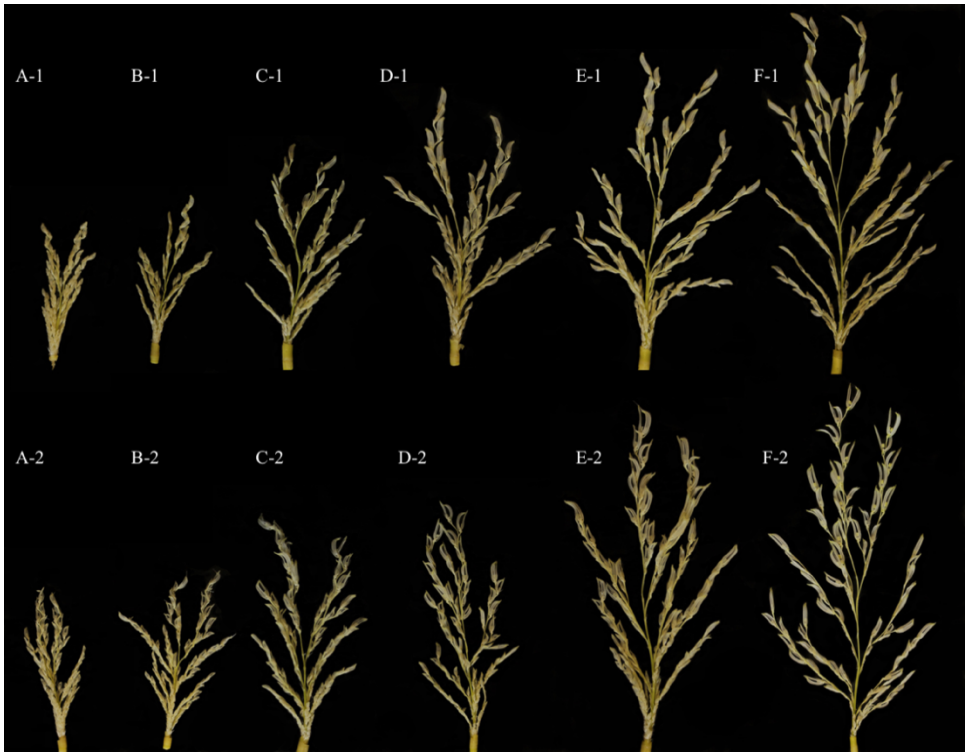

Figure 2-3. Yong panicles of *tfs1* and 4266 at different stages before heading under normal rice growth condition.

(A1-F1) Panicles of 4266 before heading at different stage show normal closed hull (time order). (A2-F2) Panicles from *tfs1* before heading at different stage show defective open hull (time order, correspond to 4266).

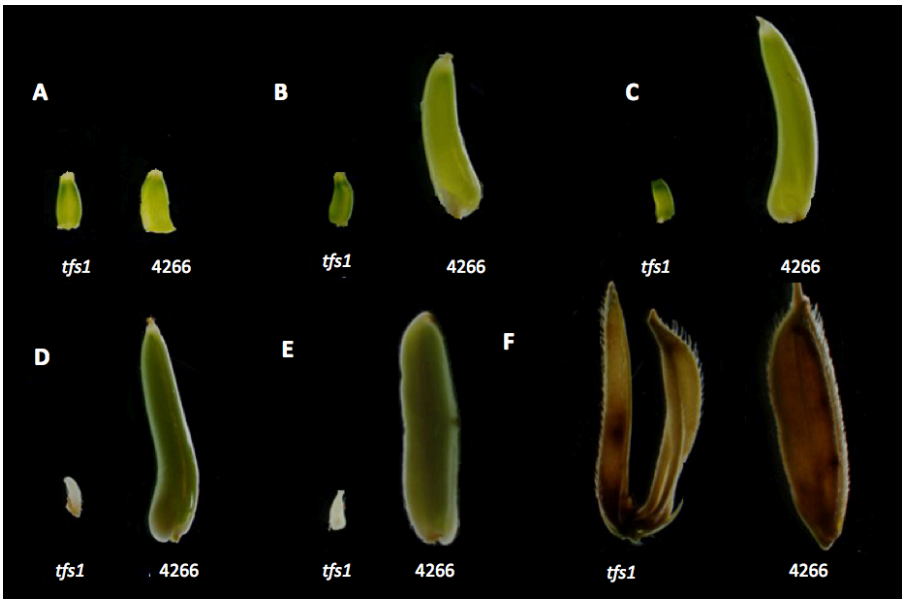

Figure 2-4, Ovaries of *tfs1* and 4266 at different stages after pollination.

(A-E)1 to 5 days after pollination respectively: On the left is *tfs1*, on the right is 4266, the grain filling in 4266 is normal. Compared to 4266, ovaries of *tfs1* did not swell at all. (F) Ovaries form *tfs1* became dry and dead 14 days after pollination (left) ovaries of 4266 generate a plump seed (right).

2.3.2 Pollen viability of *tfs1* was similar with 4266

To find out the reason why seed development of *tfs1* is arrested after pollination, we firstly examined the phenotype of anther and the vitality of pollen from *tfs1*. The morphology of *tfs1*'s anthers (Fig. 2-5 B, C) was quite similar with that from 4266 plant (Fig. 2-5 E, F). Including the size and color, as well as the pollen grain inside it. Then we detected the pollen viability by using I-KI, result shown that pollen from *tfs1* (Fig. 2-5 A) was normal when compared with 4266 plant (Fig. 2-5 D). Coincident with the results of hybridization experiments with *nipponbare* and 9311. Then we examined the pollen germination to detect the activities of pollen from *tfs1* in vitro. Our result further revealed that the pollen from *tfs1* mutant plants were quite similar with that from 4266 plant (Fig.2-6), the mutation did not affect the development of pollen and its activity in *tfs1* mutant.

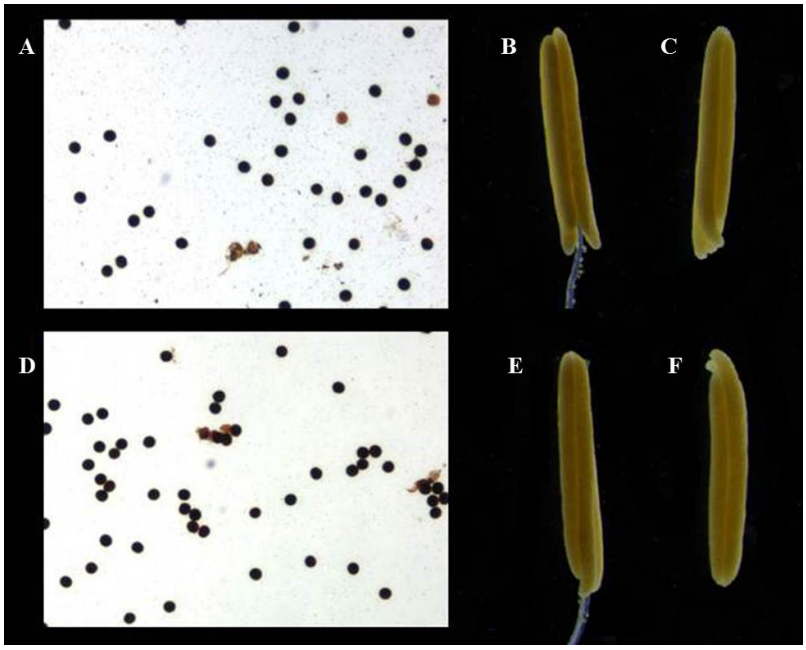

Figure 2-5. Mature pollen and anther from *tfs1* and 4266 plant.

(A and D) Mature pollen stained with I-KI of *tfs1* and 4266 respectively. The fertility of *tfs1's* pollen is similar with 4266. (B-C) Mature anther of *tfs1* before anthesis. (E-F) Mature anther of 4266 before anthesis. Both the size and morphology of anthers are similar between *tfs1* and 4266.

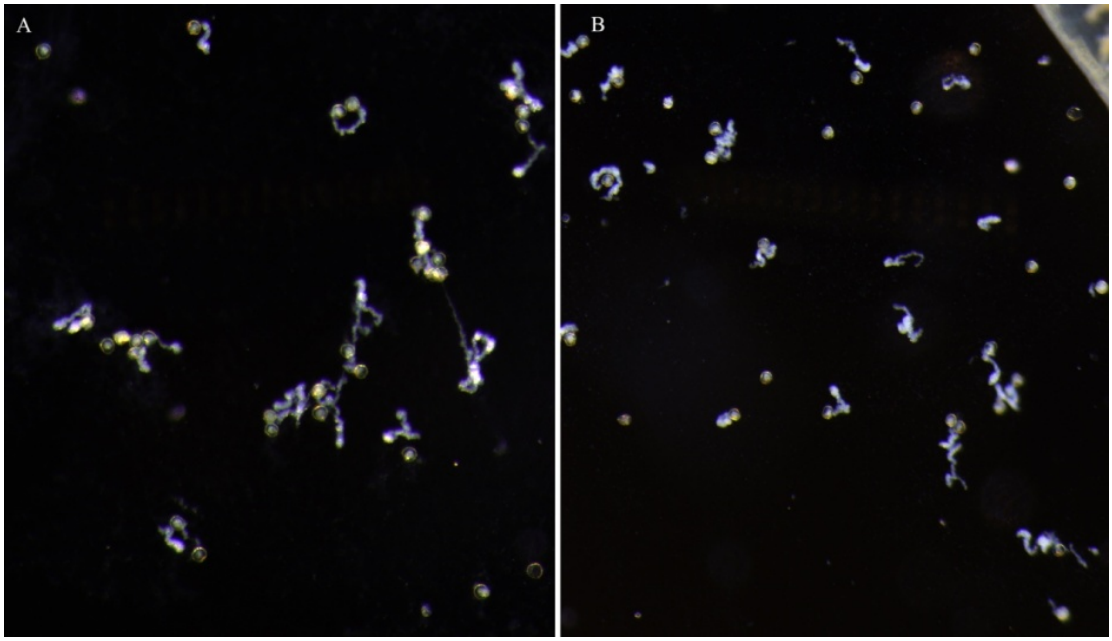

Figure 2-6. Pollen in vitro germination in vitro for 30 mins after anthesis from *tfs1* and 4266 plant.

(A) Pollen grain from 4266 plant. (B) Pollen grain from *tfs1* plant. The pollen from both *tfs1* and 4266 plant were collected and germinate on medium for 30min, and the pictures were taken by sterecscope.

2.3.3 *tfs1* has complete pistil and embryo sac structure

Accord to the morphological studies and hybridization studies on *tfs1*, as well as its pollen viability studies, we have concluded that this mutant is a female sterility mutant. To further controlling mechanism on female sterility, here we focused on the studies on pistil. As shown in figure 2-7, the pistils from *tfs1* mutant 4266 plant do not have any significant difference (Fig. 2-7). Both pistils have intact stigma, style and ovary, suggest that some problem within the ovary could be the reason of female sterility. The photo t was taken by stereoscope.

In previous study, rice embryo sac belongs to polygonum type. A mature embryo sac is a huge coenocytic cell with eight nuclei that adopt a 7cells-8nuclei configuration structure: 3 antipodal cells, one central cell (two polar nuclei), one egg flanking with two synergids. For rice, the number of antipodal cell is more than three, so we can observe a mass at the top of embryo sac (Fig.2-9). As the core of ovule, female gametophyte has always been the focus of plant biology study. To explain why *tfs1* mutant has normal male fertility but can't produce seed after pollination, we have examined the development of ovule, particularly the embryo sac by confocal laser scanning microscope (CLSM) which has been extensively used in pollen and ovule

development studies. However, we found that ovules of *tfs1* look normal and the embryo sac inside it was also similar with that in ovules of 4266 plants. A seven-cell structure was clearly observed in *tfs1* and 4266 plants grown under both normal temperature and lower temperature. To further confirm this result, pistils from *tfs1* mutant were observed again by paraffin section. Pistils were stained by safranin and fast green. Result shown that the 7cells-8nuclei configuration structure clearly exist in *tfs1* pistils (Fig. 2-9 B). indicating that cell division and differentiation and embryo sac development is not disturbed in *tfs1* mutant.

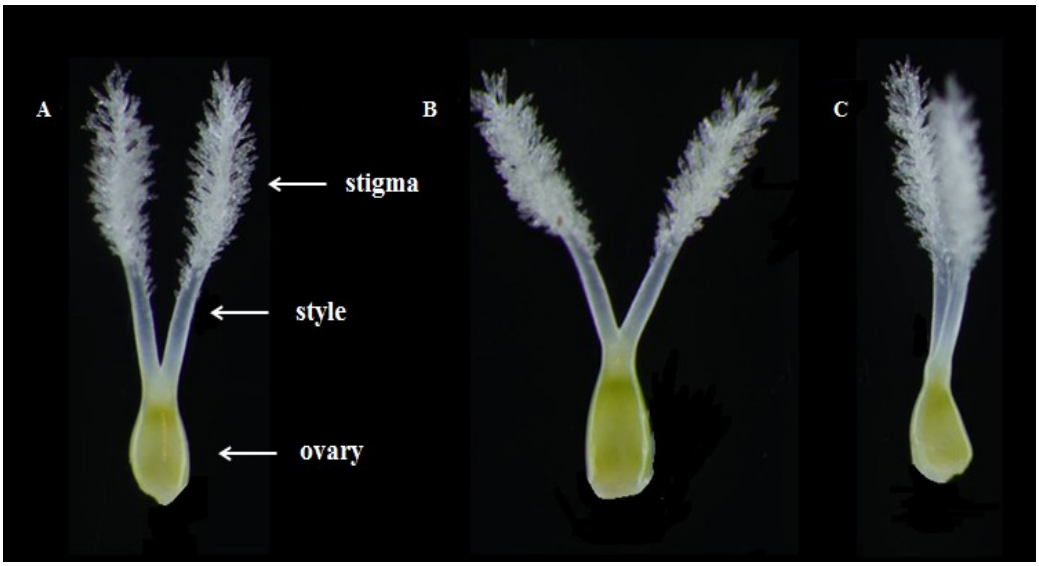

Figure 2-7. The pistils from 4266 and *tfs1*.

Pistils form 4266 and *tfs1* plants were collected before flowering, photo were taken by stereoscope (A) 4266's pistil with stigma, style and ovary. (B-C) *tfs1*'s pistils show similar morphology with 4266.

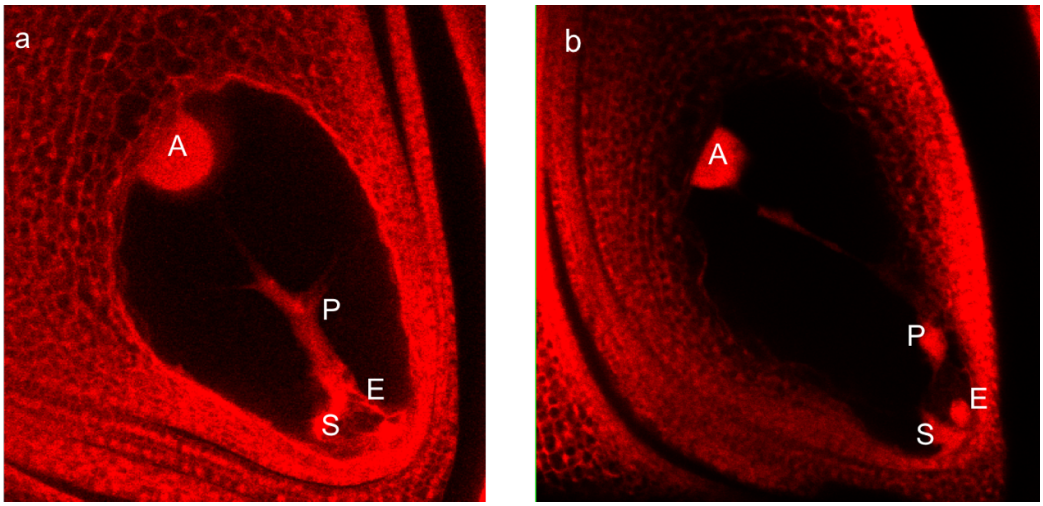

Figure 2-8. Embryo sac structure from 4266 and *tfs1* examined by CLSM.

(a) The embryo sac of 4266, shows complete 7celled and 8 nucleated structure. (b) The embryo sac of *tfs1*, also shows complete 7celled and 8 nucleated structure. A, antipodal cells; P, polar nuclei; E, egg; S, synergids.

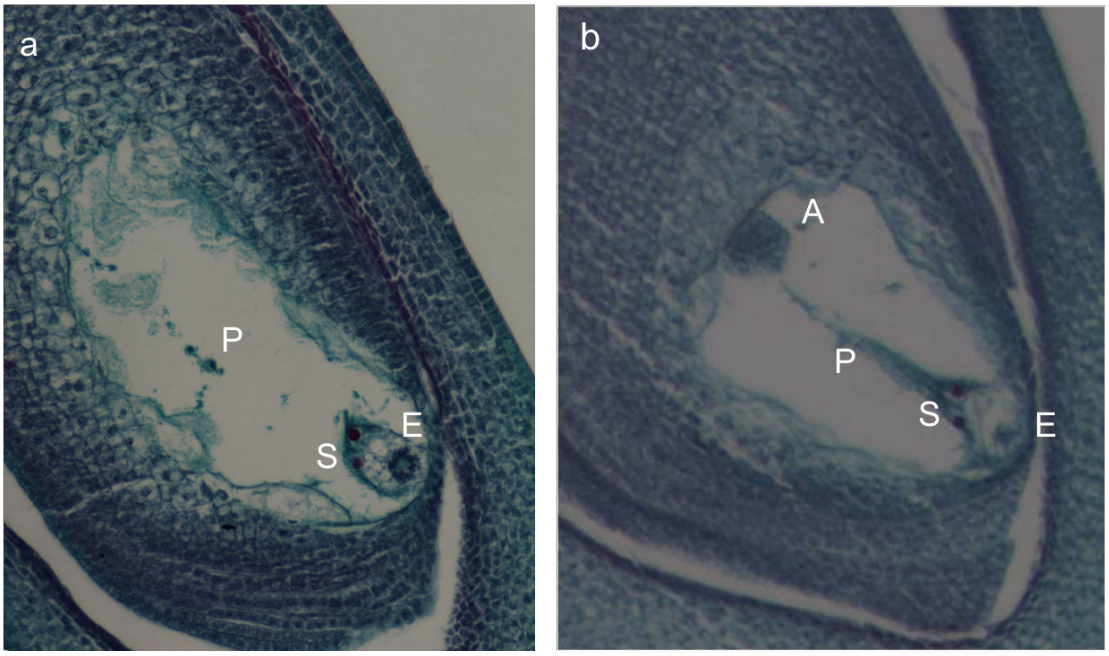

Figure 2-9. Embryo sac structure in 4266 and *tfs1* by paraffin section.

(a) The embryo sac of 4266, showing embryo sac structure, the antipodal cell is not shown in this figure. (b) The embryo sac of *tfs1*, shows complete 7celled and 8 nucleated structure, the nuclei of central cell is not clear in this figure. A, antipodal cells; P, polar nuclei; E, egg; S, synergids.

2.3.4 Pollen germination and pollen tube extension in vivo

Double fertilization comprises of pollen germination and fertilization(Berger et al., 2008; Higashiyama, 2002; Russell, 1992). Here in this study, the process of pollen grain germination on stigma and pollen tube extension in vivo was studied by using aniline blue. Pistils were taken from both 4266 and *tfs1* mutant and artificially anthesis by using anther from *tfs1* plant. The pistils were sampled and stained by aniline blue after the pollens were landed on stigma for 5 minutes (Fig. 2-10, A1 and A2) and 30 minutes (Fig. 2-10, B1-D1 and B2-D2), and then visualized by fluorescence microscopy. As shown in figure 2-10, the germination of pollen grain on stigma and pollen tube extension are similar between 4266 and *tfs1* plant. Almost all the pollen on the stigma can normally germinate, and the pollen tubes extent through style were normal in both 4266 and *tfs1* mutant. Pollen tubes extend and reach to the micropyle in both ovary, indicating that the pollen from *tfs1* mutant can germinate and extend normally on both phenotype of rice, the female sterility of *tfs1* is not induced by pollen germination and extension in pistil.

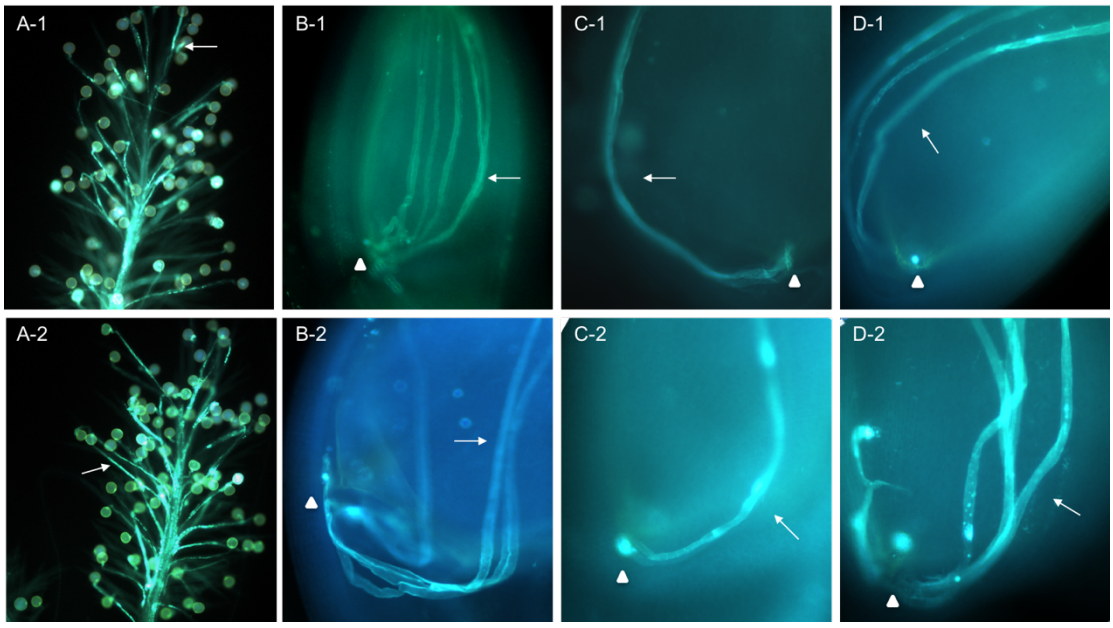

Figure 2-10. Pollen germination and pollen tube extension for 30mins after pollination.

(A1) Pollen germinated on 4266's stigma. (B1-D1) Pollen tube extent through the style and enter embryo sac from the micropyle in 4266. (A2) Pollen germinated on *tfs1* stigma. (B2-D2) Pollen tube extent is normal in *tfs1* and can also reach the micropyle; arrows show the pollen tube; arrow heads show the micropyle

2.3.5 Seed formation is failed in *tfs1*

Seed development is a fundamental process in flowering plant and thus a prime target in agricultural research because seed development is tightly linked to cereal crop production. To further find out the reason that lead to female sterility of *tfs1*, embryo and endosperm development were also observed by CLSM and DIC (differential interference contrast) (Canales et al.) after artificial pollination (Hu et al., 2009; Rotman et al., 2003; Zeng et al., 2007). Pistils of 1DAP to 5DAP were sampled to investigate the embryo and endosperm development stage, which takes place after double fertilization, in *tfs1* and 4266 plants. For unfertilized ovary, 7celled and 8 nucleated structure is integrative in both *tfs1* and 4266 plants (Fig. 2-11 a, b). In 4266 plant, two sperms have fused with egg and polar nuclei respectively at 1DAP, so that we can't see polar nuclei and egg, but a pro-embryo at the bottom of embryo sac (Fig. 2-11 c). Whereas in *tfs1*, the egg apparatus, polar nuclei and antipodal cells still exist one day after pollination (Fig. 2-11 d). In 4266 plants, embryo develops gradually from 1DAP to 5 DAP. Embryo of day 5 after pollination in 4266 plant differentiates the first leaf primordium (Fig. 2-11 c, e, g, i, k). Whereas in *tfs1*, the central cell and synergid cells could be found even 5 days after pollination, but the signal became weaker form 4DAP

(Fig. 2-11 k). This result indicate that double fertilization was failed in *tfs1*, which leads to the sterility of *tfs1* plant.

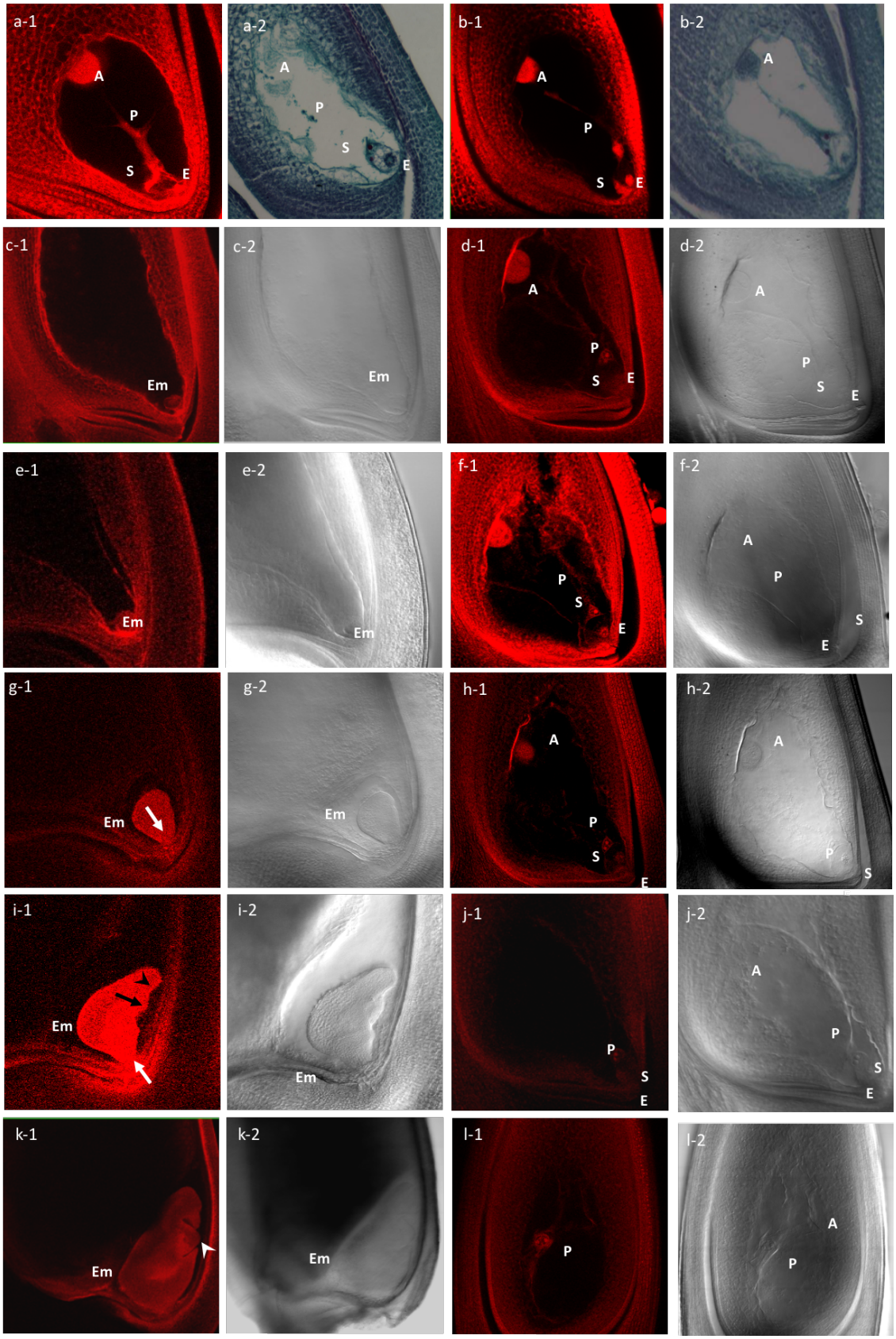

Figure 2-11. Examination of development of embryo after pollenation in4266 and *tfs1* by CLSM (left) and DIC (right).

(a) 7celled and 8 nucleated structure in unfertilized ovary of 4266 plant. (b) 7celled and 8 nucleated structure also exists in *tfs1* before fertilized. (c and d) One day after pollination, embryo sac structure of 4266 and *tfs1* respectively. In 4266, sperm fuse with egg and polar nuclei and a pro-embryo generate at the bottom of the embryo sac. But in *tfs1*, double fertilization didn't occur. (e) Two days after pollination, the development of embryo in 4266 is normal, but in *tfs1* we can still see polar nuclei exist (f). (g) Three days after pollination, embryo development is normal in 4266, while polar nuclei still exist in *tfs1* (h). (i) Four days after pollination, embryo of 4266 differentiates to coleoptile primordium (black arrowhead), SAM (arrow) and radicle primordium (white arrow). (j) Polar nuclei still exist in *tfs1*. (k) Five days after pollination embryo of 4266 differentiates to the first leaf primordium (white arrow head), (l) polar nuclei still exist in *tfs1*; A, antipodal cells; P, polar nuclei; E, egg; S, synergids; Em, embryo.

2.3.6 The development of floral primordial

The development of young panicle is in strict order, first is the differentiation of panicle primordium, then followed by the differentiation of primary branch primordium. The secondary branch primordium differentiation is just behind the primary branch primordium. The spikelet primordium and floret primordium are then differentiate on both primary and secondary branches. The structure of floret primordium is composed of, glume, lemma and palea in whorl 1, and differentiate and in order, two white lodicules at the adaxial (lemma) side in whorl 2, six stamens in whorl 3 and one gynoecium with two stigmas in whorl 4. The differentiation of pistil and stamen, including the development of male and female gametophyte, will initiates after the differentiation of floret primordium is finished.

In the present study, except the female sterility property, *tfs1* mutant also display a defective hull phenotype. To examine the defective hull formation mechanism, we use SEM technology to study the formation of floret. As shown in figure 2-12, four whorls in floret appear in strict order. The arrows show the lemma and palea overlapping part. At the very early stage, in figure A1 and A2 (Fig.2-12), the lemma and palea primordium appeared, and no significant difference between 4266 and *tfs1* was detected. In figure B1 and B2, with the elongation of lemma and palea, the anther primordium appears. The morphology of anther primordium of 4266 is similar with that of *tfs1,*

however, a bulge at the overlapping position of lemma and palea was also observed. The whorl 1 and whorl 3 keep elongating, the anther primordium is becoming more obvious (mark with white star) and the pistil primordium becomes easily to be observed (in the middle of white star). Both the number and morphology of the anther and pistil primordium of 4266 and *tfs1* are similar. At the same time the bulge at the lemma and palea overlapping position becomes more and more obvious ((Fig.2-12, C1-D2), indicating the defective hull has already been determined before this stage. At the later stage, the lemma and palea keep elongating and cover the anther and pistil gradually in 4266 plants, whereas in *tfs1* mutant, the abnormal development of lemma and palea has finally result in a defective hull (Fig.2-12, E1-G2). Eventually, the lemma and palea in 4266 is closed but lemma and palea in *tfs1* is opened (Fig.2-12, H2), which is consistent with the husk phenotype in spikelet.

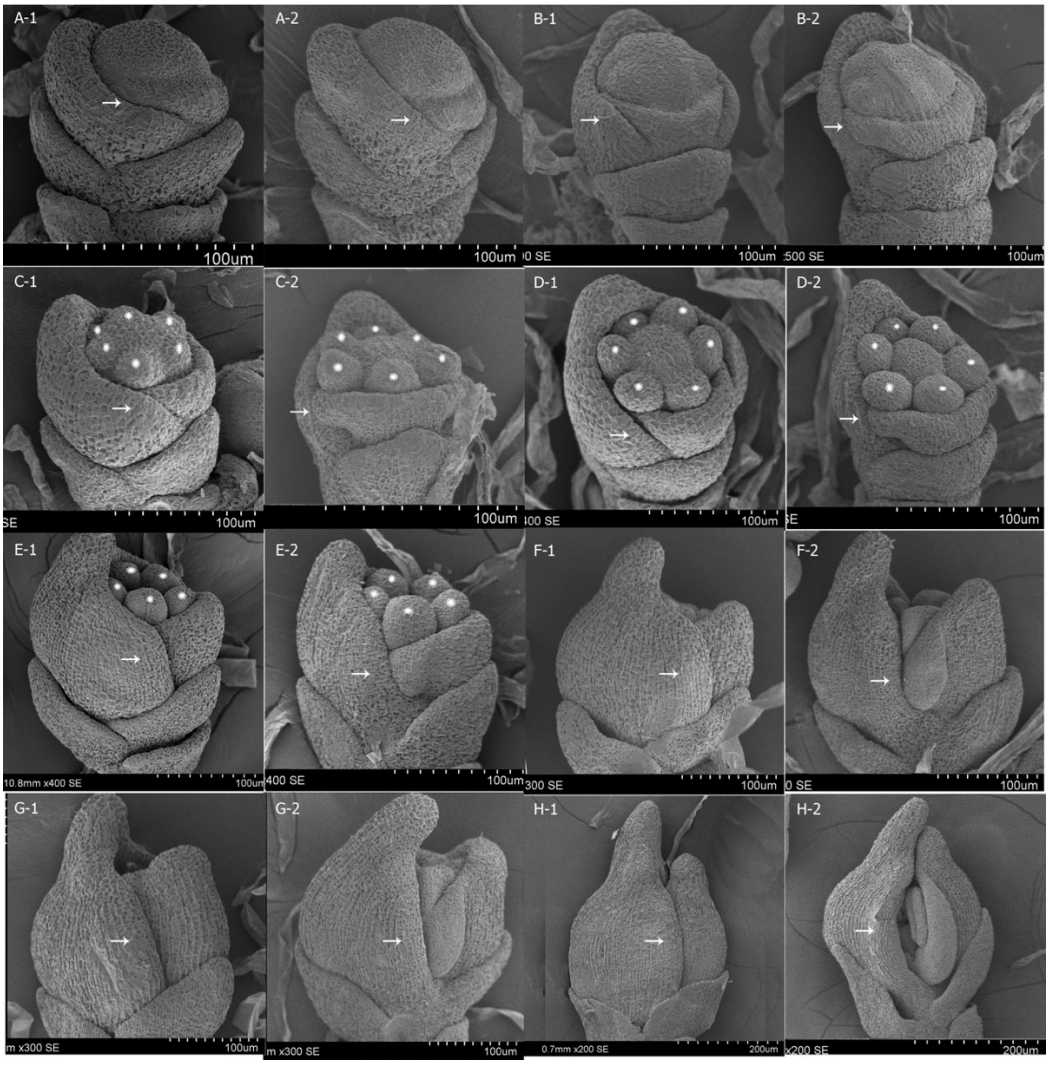

Figure 2-12. SEM study on floret primordium development of young panicle in 4266 and *tfs1*.

(A1-H1) Florets primordium of 4266 plants, in time order. (A2-H2) Floret primordium of *tfs1* plant, in time order; arrows show the lemma and palea overlapping area and stars show stamen primordium, the pistil primordium shows in the middle of stamen promordium.

2.4 Discussion

In the past decades, microscopic observation technologies have been widely used in plant researches, especially the development of germ cells, such as the male and female gametophyte development, the process of double fertilization(Kerim et al., 2003; Raghavan, 1986; Raghavan, 1988). Many progresses have been achieved in cytological mechanism study about plant reproductive development. The development of female and male gametophyte and double fertilization are well-organized and complicated processes with refined molecular mechanism. Although the pollen of angiosperm has simple structure and tiny bulk, which has only two or three cells, whereas expressed over half of total genes, and approximately 10% of these genes are specific expressed in pollen (Twell, 2002). As more and more genes related to pollen development have been investigated in both monocot and dicot, researchers have a clear knowledge of their expression and the control mechanism of pollen development. By analyzing mutant of specific phenotype, researchers found many genes control the differentiation of cells. During the formation of microspore mother cell, gene, *EXS (EXTRA SPOROGENOUS CELLs)*, can affect the division of archesporial cell which leads to the formation of an extra primary sporogenous cell in Arabidopsis. Its orthology gene in rice (Canales et al., 2002; Yang et al., 2005), *MSP1 (MULTIPLE SPOROCYTE1)*, has similar function (Dasheng et al., 2009; Zhang and Yang, 2014). As to the molecular mechanism during meiosis, more mutants were found in both rice and maize, such as *pair1* and *pair2* found in rice, the association and separation of Homologous chromosome in these mutants are abnormal, leading to the failure of meiosis (Kurata et al., 2005). The asymmetric division of microspore cell is the key point of pollen fate, a special promoter found in tomato, *LAT52*, which is expressed in vegetative cell but not in reproductive cell. Use the colchicine to induce the symmetry division of microspore cell, *LAT52* was found to express in two daughter cell, which means the asymmetric division of microspore cell will repress some genes expression (Twell et al., 1989). All these processes during cell differentiation can be observed by different microscopy methods, such as paraffin section or SEM, CLSM, and so on (Bedinger, 1992;

Blackmore et al., 2007; Knox, 1984; Stanley and Linskens, 2012; Ye et al., 2010).

In Arabidopsis, many genes involved in ovule development have been identified. In the past few years, investigators have gained significant insight into the mechanisms that control pistil development in this model plant (Yang et al., 2010). Whereas in rice, although female sterility has been identified for decades, little is known about the controlling mechanism of female sterility in rice due to the limitation of its application on agriculture. Two T-DNA insertion mutants, *apc6* and *mads13*, were found to be involved in pistil development, by which the cell cycle progression during megagametogenesis and ovule identity were disturbed, respectively (Kwee and Sundaresan, 2003; Sang et al., 2012).It is well known that both male and female sterilities are frequently found in progeny populations of *indica-japonica* cross, the incompatibility between *indica* and *japonica* and so-called hybrid breakdown. Firstly, a locus involved in embryo sac development has been identified in progeny of *indica-japonica* cross (Song et al., 2005). Then two genes, *S5* and *Sa*, leading to female and male sterility respectively *indica-japonica* hybrids were cloned and characterized (Long et al., 2008; Zhao et al., 2006). Our preliminary results from cytological observation have shown that *tfs1* is totally sterile after pollinated by pollens from different cultivars, including the 4266 plants, although pollens from *tfs1* can fertilize 4266 plants and produce normal seed. Pollen viability examinations have further support our crossing result that the male part of *tfs1* is normal. To find out the reason that results in female sterility, we have used CLSM and paraffin section technologies to study the development of embryo sac in *tfs1*. However, our result turns out that embryo sac in *tfs1* is complete too, which also has a typical 7 celled and 8 nucleated structure (Fig. 2-8 and Fig.2-9). Fertility of rice is not only controlled by male and female gametophyte developments, but also by the interactions between these two kinds of gametophytes. In *tfs1*, the normal embryo sac implies that the interaction between embryo sac and pollen or the afterward process, embryo development, are defective in *tfs1*.

To further explore the sterility mechanism, the CLSM and DIC microscope technologies, which have been frequently used in studies of embryo sac development and embryo development, were used again to observe the development of embryo and endosperm in *tfs1* after pollination. Obviously, the embryo development in *tfs1* is totally defective, the egg cell and synergid cells are still observed in embryo sac of *tfs1* 3 days after pollination (Fig. 2-11), suggesting the female sterility of *tfs1* is because of the failure of interaction

between embryo sac and pollen tube, the double fertilization process. Double fertilization is an unique process in angiosperms, which is start form pollen germination on stigma, and end with the fusion of two sperms with egg and polar nuclei, respectively (Steer and Steer, 1989; Taylor and Hepler, 1997). The first step of double fertilization is pollen guidance, which starts from pollen germination on stigma and pollen tube extending through the style then to the micropyle. Many mutants have been used to illustrate this step. It has been well-documented that after entering the embryo sac through the micropyle, pollen tube will pass through the filiform apparatus and reach to one of the two synergid cells, resulting in the degeneration of this synergid cell. In maize, a small molecular protein, *ZmEA1* which express in filiform apparatus, decrease the expression of this gene may lead to pollen tube guidance failure. In Arabidopsis, a gene *MYB98*, a member of R2R3-MYB family, is a key regulator of transcriptional events in synergids (Eckardt, 2007). It also has been reported that an Arabidopsis mutant, named *feronia*, has problem of synergid cell degeneration after interacting with pollen tube, which makes the pollen tube continue growing in the embryo sac, thus lead to the failure of double fertilization.

In the present study, we have pollinated the *tfs1* and 4266 pistils, aniline blue staining result demonstrated that both pollens from 4266 and *tfs1* could germinate and grow through the stylar tissue. It was also clear that pollen tube on both pistils can find the micropyle, indicating the growing of pollen tube is unblocked on *tfs1* compared with 4266 plants (Fig. 2-10). Our preliminary result seems to approve that pollen guidance in *tfs1* is normal, further experiments are still needed to explore cytological sterility mechanism of *tfs1*. However in rice, since the integument and somatic cell layers are much thicker than that in Arabidopsis, the transparent method is not suitable for pollen tube observation, which makes it more difficult in rice for double fertilization investigation. However, when we investigated the embryo development of *tfs1* and 4266 plants after pollination by CLSM and DIC micro technologies, we found that not only the egg cell was still found even at 5DAP, but also the synergid cells were observed in *tfs1*'s embryo sac at 3DAP. There was no embryo generate totally. On the contrary, embryo development in 4266 plants started from 1DAP to 5DAP when the first leaf primordium arises (Fig. 2-11). The failure of synergid cell degeneration in *tfs1* pistil suggested that pollen guidance in this step was abolished, which leads to the failure of

sperm release and thus the double fertilization. Our findings further supported that the degenerate of the synergid cell is important for sperm release and movement. Besides, further experiments have showed that only when two synergids were distorted, the pollen tube guidance to micropyle will failed, which means although the two synergids have different function, the remained synergid also participate in pollen tube guidance.

Development of female reproductive organ is a highly orchestrated and extremely complicated process. Genes involved in this process include those that function in polarity establishment, meristem maintenance, floral organ determination, ovule identity, and structure specification. Besides, auxin gradient, small RNA, as well as RNA processing are also proved to regulate the development process (Olmedo-Monfil et al., 2010; Pagnussat et al., 2009). Studies in the past few decades have presented a complex controlling network for this developmental process. Based on the studies on Arabidopsis, tobacco, petunia and rice, researchers have identified many genes that controlled the formation of placenta, ovule primordium and the differentiation of ovule apical-basal axis and dorsal-ventral axis, the initiation of integument. Their results revealed that these biological processed were controlled by a complicated signal network, and the MADS-box genes played an important role in this network (Nayar et al., 2014). Such as *OsMADS3*, antisense RNA expressing transgenic plants had phenotype that pistil changed to sterile floret, and stamen became pistil. Mutant of gene *OsMADS16*, also shows stamen changed to carpel. Except the female sterility property, *tfs1* also displayed a defective hull phenotype (Fig. 2-1, 2-4). Both phenotypes were strictly linked, which facilitated us to identify the mutant plants. Accord to our results, both phenotypes probably are controlled by a single gene which is the mutation gene. Thus, it will be of great help to find out the developmental stage when the defective hull phenotype is determined. We used SEM to study the development of floret, including the formation of lemma and palea. Our results revealed that the defective hull phenotype was determined at the best early stage of floret development, an abnormal growth of lemma and palea starts from the primordium stage (Fig. 2-11). These results suggest that the defective hull phenotypes, together with female sterility phenotype, are determined before the development of panicle primordium.

In conclusion, we have identified a female sterility line separated from 4266 population, which is the F9 generation of *indica/japonica* crossing. Crossing and segregation

experiments indicated that female sterility of this mutant was controlled by a recessive nuclear gene. Our cytological studies revealed this *tfs1* mutant also had an incomplete hull phenotype. Pollen availability of *tfs1* is similar with 4266 plant, including the in vitro germination and in vivo growth. Interestingly, the structure of embryo sac in *tfs1* is also complete, similar with 4266, suggesting that female sterility of *tfs1* is not due to the abnormal development of embryo sac or the other processes before this. Embryo development analysis, together with pollen guidance experiment, reveal that the double fertilization was stopped at the step of pollen tube interact with synergid cell, because the synergid cell was still observed 3DAP in *tfs1*'s embryo sac. Further SEM analysis result shown that the incomplete hull phenotype has been determined before the initiation of panicle primordium. Together, our cytological studies on *tfs1* have tentatively illustrated sterility mechanism and provide very good information for the following research on biological function of the mutant gene.

Chapter 3 Cloning of *TFS1* gene by Map-based cloning and MutMap

3.1 Introduction

Map-based cloning is an effective forward genetic technique to identify genes underlying phenotypic variations. It has been widely applied to different organisms and many functional genes has been reported in the past few decades. However, map-based cloning depends on genome-wide recombination events which resulted from crossing. Besides, map-based cloning usually requires large numbers of recombinants to narrow down a causal mutation to a single gene level, which makes it a time-consuming process. Especially for identifying genes in centromere and telomere regions where crossing over rate is very low. With the advance of Next Generation Sequencing (NGS) technology, MutMap, a method based on whole-genome sequencing (WGS) on pooled DNA from individuals with target phenotype has become an effective way to isolate candidate gene with important phenotypes. BSA (Bulked Segregant Analysis) are used both in map-based cloning and MutMap to identify markers that are genetically linked to the mutation position. BSA needs two groups of F2 individuals with opposite phenotype. In the two groups, alleles in all loci except the mutant locus will be randomly distribute and thus become the same between the two groups. By using molecular markers, we can locate the mutation on one chromosome and even narrow down to a small DNA region (Takagi et al., 2015).

Forward genetics has been a powerful tool in dissecting the genes in many biological processes, especially in crop breeding. Many genes controlling reproductive development have been cloned in rice and other plants. Female sterility has been widely found in higher plants. Studies on mechanism of female sterility have been intensively carried out in model plant Arabidopsis and many genes involved in ovule development have been identified. Whereas in rice, although female sterility has been identified for decades, little is still known about the controlling mechanism of female sterility in rice due to the limitation of its application on agriculture. To date, none of female sterility germplasm that could be restored by certain day length or temperature has been reported yet. In this study, we use traditional map-based combined MutMap analysis to clone the temperature sensitive female gene in *tfs1*. This research will not only help us

understand the plant reproductive biology and the mechanism of female sterility, but also brings the application value for rice hybrid production.

AGOs are usually the core component in small RNA directed silencing of target gene, including the siRNA and miRNA in plants. Argonaute proteins contain six domains, the PIWI, MID, PAZ, N-terminal domain and two linkers, and thus belong to the PIWI protein superfamily. These six domains, MID-PIWI and N-PAZ, together with two linkers form a bilobal scaffold, which binds to the guide and target RNAs. The PIWI domain is responsible for the slicing activity for the target RNAs, while the MID domain which consists of a Rossmann-like fold will recognize the phosphorylated first nucleotide and base features of target strand. The crystal structure of PAZ domain shows it contain a hydrophobic pocket which can binding a single oligonucleotide, is related to the 3'-end of sRNA. Besides, PAZ domain is also necessary for the assemble of RISC in non-cleavage AGOs by unwinding the small RNA duplex. Structure investigation of the Argonaute-sRNA complex in Thermus thermophilus (TtAgo) showed that the N-terminal domain, which is the least well characterized domain of the four domains, stop the extension of base pairing between the guide RNA and target mRNA at the 3'-region of the guide RNA. In rice, there are 19 AGO proteins, which have a variety of functions ranging from post transcriptional gene silencing in eukaryotes to host defence in prokaryotic systems. Thus mutation of one AGO genes may lead to various phenotype, from seedling growth to seed development.

3.2 Materials and methods

3.2.1 Mapping population for map-based cloning

Plant materials and growth condition were described as above. To determine the position of *TFS1* gene, map-based cloning was used in this study. F2 population was used for map-based cloning which was generated from the cross between *tfs1* and *nipponbare* under normal condition. BSA performed in two F2 pools F (fertile) and S (sterile) with 40 individuals of each pools for primary mapping and 20000 more individuals for fine mapping.

3.2.2 Mapping population construction for MutMap

Plant materials and growth condition were also same with described above. The derived F2 population used for MutMap was generated from the cross between *tfs1* mutant and

4266. DNA form 90 individuals show sterile with abnormal hull form F2 population were pooled. Each pool involves 30 individuals with Equal amounts of DNA from selected individuals according to Nanodrop™ 2000 (Thermo Scientific) spectrophotometer measurement. Then three pools together with 4266 and *tfs1* plants were sequenced by whole genome sequencing (Patel et al., 2012). MutMap sequencing were carried out by sequencing company (Annoroad, Beijing).

3.2.3 Genomic DNA extraction

For the F2 population of map-based cloning, total DNA was extracted using SDS method with modification. After disrupt rice leave (0.1g) in liquid nitrogen, 900ul 1.2% SDS extraction buffer (1.4 M NaCl; 100 mM Tris-Cl; 20 mM EDTA) was added. The homogenate was incubated in 65 ℃ for 30 min with periodic mixing. Then 270ul 5M KAc was added, mix and incubated on ice for more than 20 min. Centrifuge the lysate for 10 min at 13000 g and transfer the supernatant to a new tube, add 600ul isopropyl alcohol and incubate in -20℃ for 30 min. Centrifuge the mixture for 10 min at 13000 g and discard the supernatant. Wash the pellet by 70% ethanol twice, dry the pellet at RT and add 100ul milli-Q water to dissolve the pellet.

For the for MutMap sequencing, DNA of F2 population, 4266 and *tfs1*, was purified by adding pheno/chloroform/isoamyl alcohol (25:24:1, v/v/v) and mixed gently, centrifuged the mixture for 10 min at 13000 g, transfered the supernatant to a new tube added chloroform/isoamyl alcohol (24:1, v/v) and mixed gently, centrifuged the mixture for 10 min at 13000 g, transfered the supernatant to a new tube and added isopropyl alcohol, the further steps were the same as above.

3.2.4 SSR markers selection for map-based cloning

The SSR markers have polymorphism between *tfs1* and *nipponbare* were selected for map-based cloning. 81 SSR markers well-distributed on twelve chromosomes were used for the primary mapping and more SSR markers around the candidate mutation position were developed based on the reference DNA sequence of *nipponbare* and 9311 for fine mapping. All those selected SSR markers were tested for polymorphism between *tfs1* mutant and *nipponbare*, which were the two parents for mapping population construction.

3.2.5 Identification of the causal SNPs for MutMap

DNA from 30 individuals have sterility and abnormal hull phenotypes were equally mixed and applied to whole genome sequencing by using Illumina High-seq 4000 sequencer. Sequencing of each pool produced about 104 Million clean reads (150-bp), approximately 30x coverage of whole genome. Sequencing of *tfs1* and 4266 obtained 193 and 160 Million clean reads, respectively. Reads were aligned to the *Nipponbare* reference genome (IRGSP-1.0) (Kawaharaet al. 2013) using software BWA (version 0.7.0) (Li and Durbin 2009), 527 SNPs called between *tfs1* and 4266 were used for MutMap analysis after stringent filtering. We generated SNP index described by Akira Abe to show the extent SNP linked with target gene. SNPs unrelated to the target gene will segregate randomly and the ratio of SNPs segregation in 3 pools of F2 will be 1:1, resulting in the SNP-index of around 0.5. While the SNP closely linked to the target gene will be shared by all the F2 individuals thus the SNP-index will be around 1.

3.2.6 Generation of transgenic plants

Gene overexpression: cDNA of *TFS1* was amplified by PCR using mRNA isolated from 4266, the primers sequence was (*tfs1*-F:5'- CGGACAGGTTAAAGCGGTGT-3'; *tfs1*-R:5'- GCGGTGAAGCTAGCAGTAGA-3'). The sequence of the amplified fragment was verified and subcloned into vector pOX downstream of the ubiquitin promoter. Construced vectors were introduced into both *tfs1* plant and 4266 plants via Agrobacterium-mediated genetic transformation.

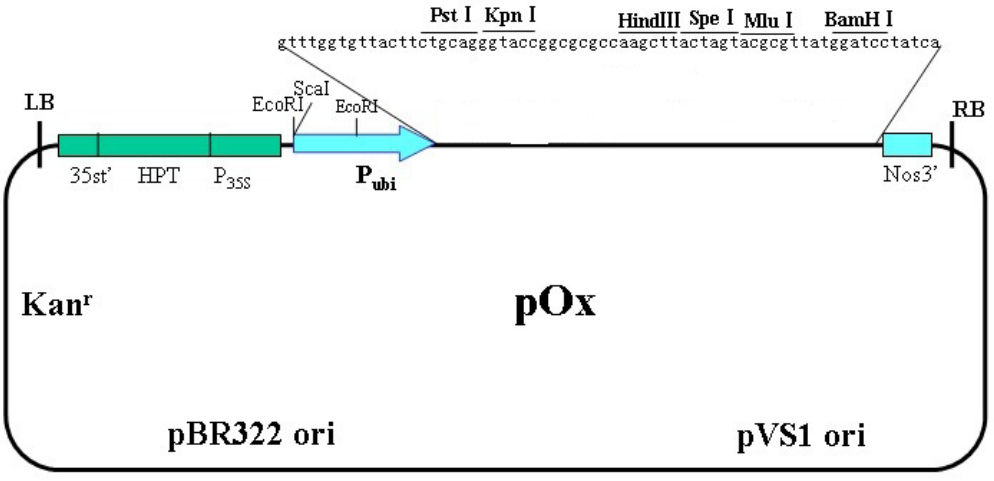

Figure 3-1. Map of vector pOX for gene overexpression.

The target gene was subcloned into the MCS by using restriction endonuclease site PstI (5') and BamHI (3'), under the control of the ubiquitin promoter.

Complementation vector: *TFS1* was amplified by using genomic DNA isolated form 4266 plant, a total of 10 kbp fragment include 5kbp before 5'UTR region and 1kbp after 3'UTR region were used for transgenic complementation. The whole gene was amplified into two 5kbp fragments which has 20bp overlap. Fragments were sequencing verified and subcloned to pCambia1300 which has no promoter to activate the gene, between BamHI (5') and SalI (3') by In-Fusion method (ClonExpress one step cloning kit, Vazyme). Primers used for fragments amplification were shown as follow (*tfs1*-gDNA-F1:5'- AGCTCGGTACCCGGGGATCCGGAACCAAAGTCCA-3'; *tfs1*-gDNA-R1: 5'-CTCCCGCTTCAGTCCACACT-3'; *tfs1*-gDNA-F2: 5'-AGTGTGGACTGAAGCGGGAG-3'; *tfs1*-gDNA-R2: 5'-TGCCTGCAGGTCGACGGTTGCAGGATCAAACTGA-3'). Both the complementation vector and control vector were introduced into *tfs1* via Agrobacterium-mediated genetic transformation.

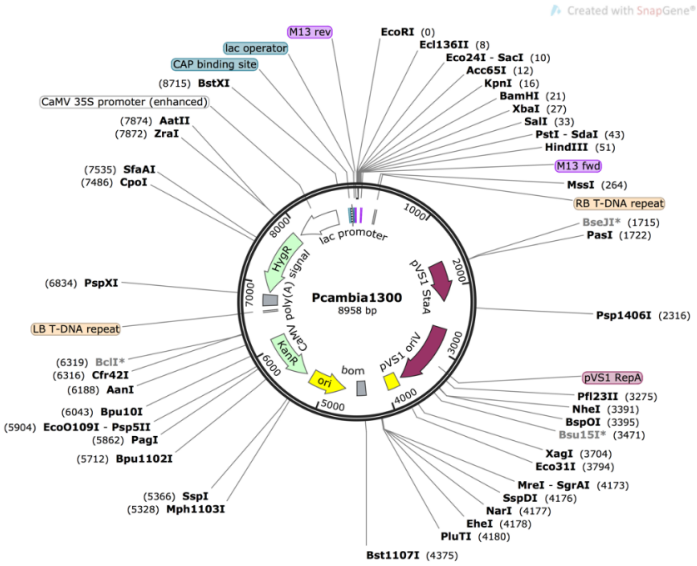

Figure 3-2. Map of vector pCambia 1300 for complementation experiments.

Vector pCambia 1300 has no promoter upstream the MCS. The whole gene of *TFS1,* including promoter, intron and Extron, was subcloned into the MCSbetween BamHI (5') and SalI (3').

Genetic editing by CRISPER/Cas9 system: The CRIPSR/*Cas9* multiplex genome targeting vector system carrying CAS9 coding gene and with sgRNA cassettes driven by *OsU6* was used for gene *TFS1* editing.

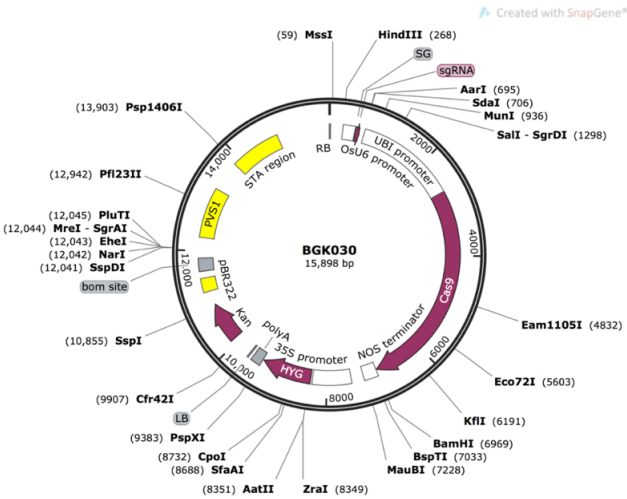

Figure 3-3. Map of vector BGK030 for CRISPER/Cas9 system.

Specific site of *TFS1* gene was designed and recombined into BGK30 then transformed via agrobacterium rhizogenes.

3.2.7 RNA isolation and quantification

RNA samples used for qRT-PCR and cDNA amplification were prepared form leaf by Plant RNA Extraction Kit (Qiagen) and DNA was removed by RNase-Free DNase Set (Qiagen). RT-PCR used RevertAid First Strand cDNA Synthesis Kit (Thermo) and Equal amount of RT products were used to perform PCR. qRT-PCR analysis was performed three repeats for each sample using iQ™ SYBR® Green Supermix (Bio-Rad) with CFX Real-Time PCR Detection system (Bio-Rad). The cDNA levels of target gene were normalized to the internal standard gene UBQ5 (UBQ5- qRT -F: 5'-AAACCCTAACGGGGAAGACCATAA -3'; UBQ5- qRT -R: 5'-CCACAGTAATGGCGATCAAAATGA -3'). For the expression level of *TFS1*, the primer was designed near the 3'UTR region and the sequences of primers were shown as follow (*tfs1*- qRT – F2: 5'-GCTACACATTTGCTCGGTGC -3'; *tfs1*- qRT -R2: 5'-TCTGAGCTTAGGTAGCGGCA -3').

3.2.8 Location of TFS1 in rice protoplast

Protoplast preparation: the isolation of protoplast was mainly based on the methods reported before with modifications.7 to 10-day-old rice seedlings cultured at 28~30°C were used for protoplast isolation. Stem and leaf sheath tissues from 40-60 rice seedlings were cut into approximately 0.5-1 mm strips then immediately transferred

into 0.6 M mannitol and incubate 30 minutes for a quick plasmolysis treatment, followed by enzymatic digestion (MES, PH=5.7; $CaCl_2$, 10Mm; Mannitol, 0.6M, Cellulase RS, 1.5%; Macerozyme R-10, 0.75%; BSA, 10%; Amp, 50ug/ml) in the dark with gentle shaking. The protoplasts were collected by filtration through 40 μm nylon meshes, and centrifuge 5 minutes at 200-300g at room temperature, and remove the supernatant. Quickly add W5 solution (MES, PH=5.7; KCl, 5Mm; NaCl, 5Mm; $CaCl_2$,125mM) to resuspend the protoplast, and centrifuge at less than 800g to remove the supernatant, resuspend by MMG solution (MES, PH=5.7; Mannitol, 0.4M; $MgCl_2$, 15mM) and prepare for transformation.

Vector construction and transformation: The open reading frame of *TFS1* was amplified by RT-PCR by using mRNA isolated from both 4266 and *tfs1* seedlings without stop codon, and fuse GFP to the N-terminal. The sequence of the amplified fragment was subcloned into vector pHW-avi-GFP (Fig. 3-4). Expression vectors carry genes from 4266 and *tfs1* were introduced into protoplasts, mediated by PEG (PEG 4000, 40%; Mannitol, 0.2M and 0.1 M $CaCl_2$), 5-10μg of plasmid DNA were mixed with 100μL protoplasts (about 2×10^5 cells), Then incubated at room temperature for 10-20 min in the dark. After that, add 440 μL W5 solution slowly followed by gently inverting the tube gently. Then centrifugation at 1,500 rpm for 3 min, and resuspended the protoplasts gently by using 1 mL WI solution (0.5 M mannitol, 20 mM KCl and 4 mM MES at pH 5.7). Finally, the protoplasts were transferred into multi-well plates and cultured under light or dark at room temperature for 6-16 h. And observed by confocal microscope under GFP exicitation length (488nm).

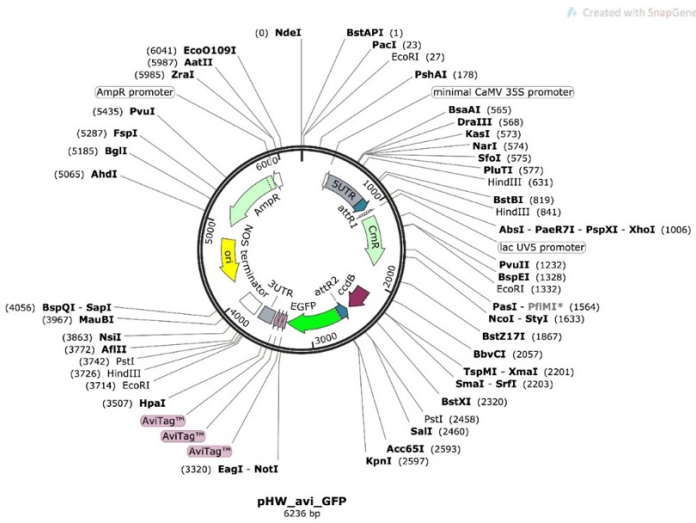

Figure 3-4. Map of vector pHW_avi_GFP for subcellular localization.

Two *TFS1* genes from 4266 and tfs1 mutant plant, without terminator codon, were fused into pHW_avi_GFP vector, respectively. The recombined vector will then produce TFS1-GFP fusion protein.

3.3 Results

3.3.1 Markers selection for map-based cloning

For primary mapping, we have designed SSR markers based on the DNA sequence of *nipponbare*. Then 81 polymorphic SSR markers between *nipponbare* and *tfs1* were selected, as shown in Figure 3-5, left lane stand for *nipponbare* and right lane is *tfs1*. All the selected markers were well-distributed in rice genome (Fig. 3-6), which were suitable for primary mapping. The primer of these selected SSR markers are shown in Table 3-1.

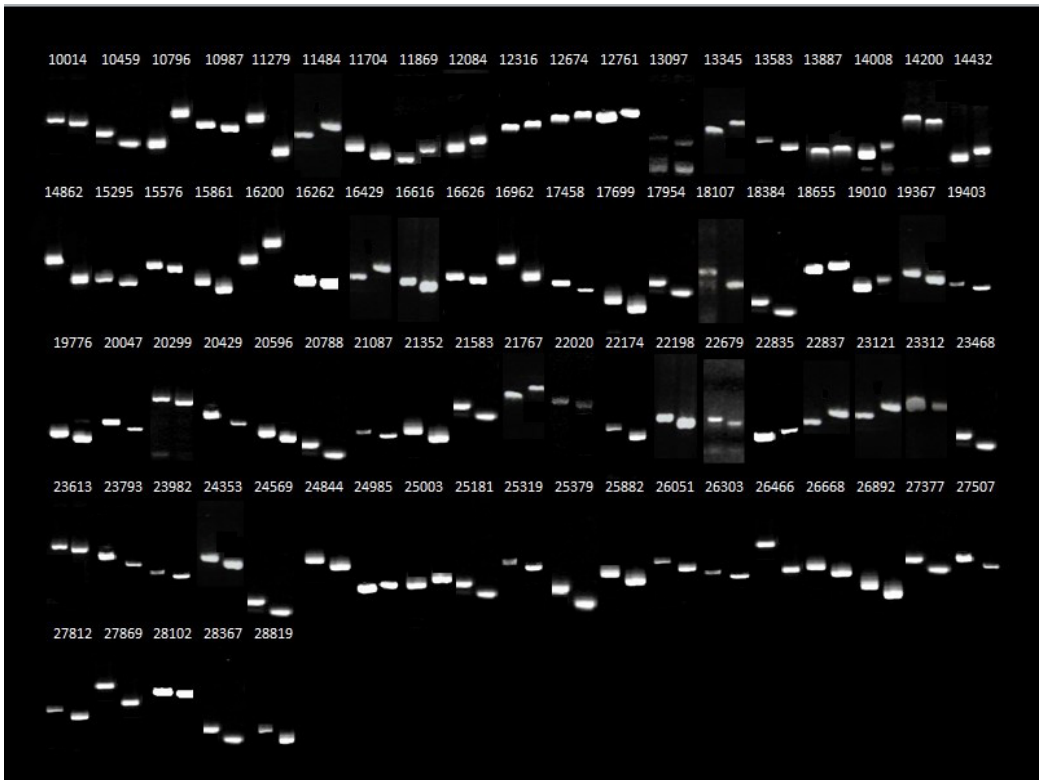

Figure 3-5. Identified polymorphic SSR markers for primary mapping.

Template DNA from *nipponbare* and *tfs1* was extracted by SDS method. SSR markers were examined by PCR and the left lane is *nipponbare*, right lane is *tfs1*.

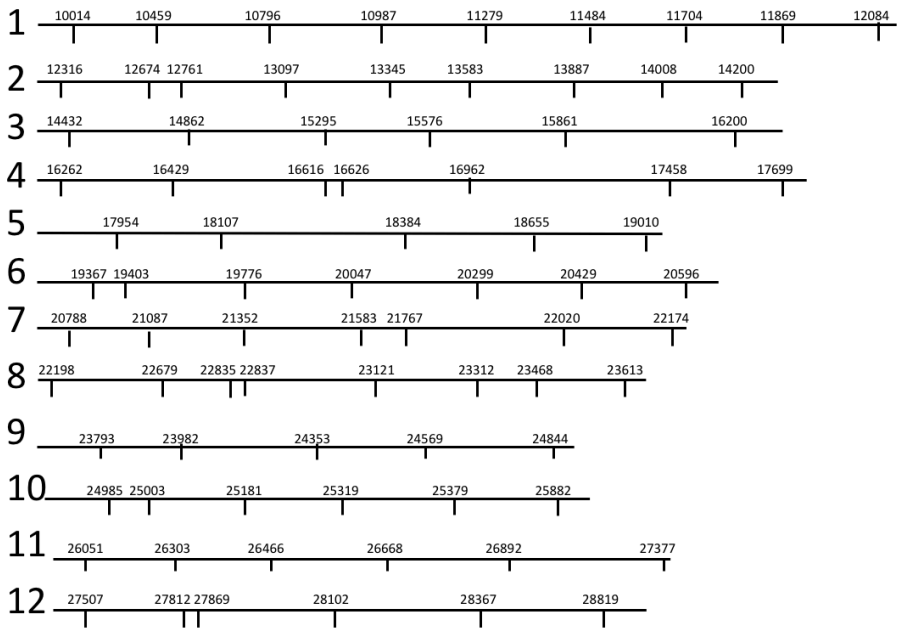

Figure 3-6. Positions of selected polymorphic SSR markers on 12 chromosomes for primary mapping

Table 3-1. The name and the sequence of selected SSR markers for primary mapping.

No.	Chromosome	SNP Marker	Forward Primer	Reverse Primer	SSR start	SSR end
1	1	10014	ATGATGATGGACGACGACAACG	TGAATCCAAGGTGCAGAGATGG	216000	216020
2	1	10459	TTTGAGCCCTCACTCTCTGTCC	GCTTGATCCTAATCACGCTCTGC	7361621	7361692
3	1	10796	ATCCTCTTCTCTGTTCTTCCACTTCC	CTTGTCCCTGCTCCATTCTCG	12672634	12672655
4	1	10987	CCTCACCATTGATCTCCATCACC	ACTTGCCCTCCTCTTCTCAACG	17274013	17274046
5	1	11279	GGCCCTCATCACCTTCGTAGC	CGTCCTCCCTCTCCCTATCTCC	23322208	23322235
6	1	11484	GCAAGTCGATAGGCTAATCAATCC	CATGTGTAGTGGGAAGATGTGTGG	27504750	27504815
7	1	11704	ACAGGGCTATGCAGACACAGTGC	AGCAAGCGAAGGGAAGTGACC	32093949	32094020
8	1	11869	CAAATCTCGCAACAACACAGAGC	CATACCAGCCCAACCGAAACC	36291780	36291799
9	1	12084	AATCGGCGAGGTTTGCTAATGG	ATACGTGGTACGTGACGCTTTGC	40100318	40100339
10	2	12316	AGCCAAGATCGTCTTCATCTCTCG	TCGTCTCCTTCTTCCTTCTTCTTCC	183362	183393
11	2	12674	TAAATGCCAACCAACTCCAAGC	AACTGCGTTTGGGAATATCTCG	5687665	5687686
12	2	12761	GAAAGCGAGGTGAAACGAAGC	CTAGGCCCAAGGTTTACCAAACG	7284919	7284940
13	2	13097	GGGCTTAAGGACTTCTGCGAACC	AGCGATCCACATCATCAAATCG	13190528	13190590
14	2	13345	CTTCTCCTGCTCCTCCCGATCC	CTTCCATGGCTGTCCGTCTTACC	19361461	19361490
15	2	13583	AAGCTGGGTGATCTTAGTTTGG	GAGAGTAGTGGCTATTTGCTTGC	23862405	23862434
16	2	13887	GGCTACGCTCCAGTTAGTGAAGG	TGGTAGGAAGAAGGGAGAAGTTTAGG	29655977	29655998
17	2	14008	AGAGAGGACATCCACCACTAGCC	AAATCTCTCCCAAGTCCCAATCC	32641163	32641184
18	2	14200	AATGACGTCTAAGGTGCTTTCAGC	CTGGGCTTAGTTCTTGACAGTGG	35416268	35416287
19	3	14432	CCTGCTATTAAGGAGTCAAGATCTGC	GCTATCCTATGAAATTGGCTGTGG	3417456	3417475
20	3	14862	GGGTATTTGTAAGGTGAGGTGTCG	GTCTGGATTTCGGGTTTGAGG	11635786	11635837
21	3	15295	GGAGGGCGTACATGTCAATAAGG	CCAGACCGGGACTTCTCATAGG	19324752	19324783
22	3	15576	CCAAACATGGCCTTGTAGTAGACG	CTGTGGCTATGCCTTTGGTTGG	24919550	24919605
23	3	15861	GGTGTGTGAGAAGAGGAGATTGG	GGCCCTTTGAAGACAACACTAGG	29784121	29784186
24	3	16200	GTGGTAGGGCGAAATGATCTGC	ATCACGCGCTCCTACCTCACC	35674153	35674176
25	4	16262	CTTACTCCATTGGGCTGGAACC	TGTAGGGTGGTAAGAGATCCACTCC	168711	168746
26	4	16429	TGTTGGGTGGGTAAAGTAGATGC	CCCAGTGATACAAAGATGAGTTGG	4418222	4418259
27	4	16616	TAGGCACTGGGCTCAAGTAGTGG	TTCTCATCCTTCGCCATCTAACC	12100013	12100066
28	4	16626	ACATGATTGCTGGCTTGCTTACC	GCCACGCAGTGTTGTTTCAGC	12778457	12778480
29	4	16962	GTTTGAACTCTGGTCTCAATGC	ACAGTATACGTGGCTCAATCAGG	20780840	20780884
30	4	17458	GAGATAACACCTTCCTACTGTCG	TGACATGAGCTACCATATCACC	30725662	30725709
31	4	17699	AGAGAGCCCTGCTGCATTTGTCG	TGGGTCTTTCGTTCGTCACTACCC	35393294	35393320
32	5	17954	ATTTCAGTACAAGGCACCCATGC	GTAGACGAGGGAGTACCAACTTGC	3591046	3591123
33	5	18107	CGTATGGACTTGCCTTGAGTCG	TCCAATCTGCCAAGCTTTACACC	6742128	6742193
34	5	18384	GCAGCAGAAAGGGAGAGAGTATGG	CAGCAACGTACGTACCAACAGG	14707368	14707433
35	5	18655	GGTAGGCACCAAAGAGTTTGACG	GGCATCACCTTATCCAATCACC	19830882	19830947
36	5	19010	CAGTTTGACACAAGCAGCAAGC	CTCCACAGCTTGAGGGAATGC	25888255	25888282
37	6	19367	CGTCATGTCGCGGAGGTAAGC	AGGCGTACGTGGAGCAGAGTGC	2335265	2335285
38	6	19403	GTTTGACGCGATAAACCGACAGC	ATGAGGACGACGAGCAGATTCC	2831543	2831572
39	6	19776	ACCTGCTCCATCCATCTCTACGG	AGCAACGTGGTACAGATTACAGAAGC	9232250	9232270
40	6	20047	AACGAGCGGCTGAGATTTATGG	ATGTACCAGCCCTTTGAGAGTGG	15875183	15875218
41	6	20299	TGGGCTTTGAAGTCAGGTAATCC	GGCAGTTTAGGAGCGATTGTAGG	22335308	22335363

42	6	20429	AGTTTCCTAGCGCTTCAGCATCC	TGTGCGTATGAGAACCATCATGC	25043532	25043624
43	6	20596	AACTTCCTTTCCAGGCTTTCAGC	TTCACTGAGCCTGAACACATTGC	27706647	27706672
44	7	20788	CCAAGAAGCAATCACGGATAAGG	AGCAGGTCGAGTACCAATTCAGC	156988	157049
45	7	21087	CAGCTTCAGCTCCTAATGGTTGC	GCCTCTCAGAGCCTTCTTCTTCC	4198722	4198741
46	7	21352	GGACGAAAGGGTATTTGATTGG	CCTCAAGGTGGTCTCCTTCTCC	10020417	10020466
47	7	21583	GGTACAAGCAACTTGCACTGTGG	CATCGCCTGGGAGATCATACG	17417622	17417681
48	7	21767	GAGCCCATAACTACGGAGAAAGC	CGAGAGGAAGAAGAAGAACAAGACG	20991582	20991605
49	7	22020	CGTTGAGACGGTCACCTAATGC	GCTCAAATGTTTGACACGAAGC	26232603	26232642
50	7	22174	CTGAGGTTTGCCGACGAGTTCC	GGGCCATTCTTTCCTTGGTTGG	29477616	29477635
51	8	22198	AATTGCCCAACGAGCTAACTTCC	TGAGCTGTTTGTGCTCTTCTACTTCG	119997	120022
52	8	22679	CCTCATCATCGTCTCCGACTGC	TCTCAAACCCTCAACAATGTCTCG	8573757	8573800
53	8	22835	ATGTTGCACTCCTTCAATGTCC	CATGAGACAATGCCAGAAAGC	12288130	12288159
54	8	22837	ACCTGGGTCAGATGTCTGTTTGG	GGTAGAGCTCCATCCATCTTAGTGC	12369866	12369913
55	8	23121	CGAGATACGGTGAATGACACACG	ACCAATCAGAGAGCACACACAGG	20279094	20279115
56	8	23312	CTGCAACCTGCACGAGTACACC	GTCCCATTGGATAGAATCCCAGAGC	23551198	23551242
57	8	23468	GCGGTAACAAACCAACCAACC	AAAGCAGGACACAGTCACACAGG	26140988	26141035
58	8	23613	TGGATCCATTCCATTCCATTCC	CCGAGAATACGCATAAAGCTTGC	28070747	28070766
59	9	23793	GCATCTCAGCAAACAAGAACAACC	GAGGCCATCAAGCAGTTCTCC	4300565	4300621
60	9	23982	CAAGCAGGAGCTCTTCTATTCC	GCAGAAGCGTGATGTTAATCG	8490799	8490844
61	9	24353	GTTCTTCCCTACTGTGAACATGG	GTGAAATATCCCTATGCTCTCACC	15451407	15451468
62	9	24569	GCCTTTCGGATGGAAGTAATTAGG	CCGTAACCCTGACAATATGAAACG	18619891	18619936
63	9	24844	CCACTACACCAAATCGTCATTGC	ATGGATCTCCAAGTGATGGATGG	22546728	22546747
64	10	24985	CACCTCCTCTCTCCTCTCTATCTTCC	TATAAGGGCTTTCGGATGGATGG	2599545	2599566
65	10	25003	GATTGATCCGAGAGACAAATCC	TCGATCAATAGTAGCAGCAGTAGG	3336416	3336439
66	10	25181	AAAGAGCTTCCCTAATGGCTTCG	GAGAGAATGACCTCTCCCAAGACC	8444115	8444180
67	10	25319	AGGGTAGAGTATGTCGGTGTTTCC	CCGTGGCAGTAGCAGTAGGC	12249262	12249289
68	10	25379	CCCTTAGATTCCTGCAGCTTTCC	AAGAGGGAATTGGAGGAATGAGC	13284549	13284604
69	10	25882	GACATGCGAACAACGACATCC	CTATAGTTGTTGCACATGCGATCC	21982177	21982224
70	11	26051	ATCAGCATATCTCCCTGCAAAGC	GGTTCTTCACCCGCCATATTCC	1981423	1981464
71	11	26303	CATGCCACACAGATCACTCTATATGC	CGATGGTTATGTACGTTGCATGG	6992773	6992792
72	11	26466	GACCAAAGGAGAGGCTGAGAGG	AAACACCGTAGCGAAGCATGG	10268026	10268085
73	11	26668	ACGAAGCGTCCTGCATCATCC	CCTTCCAACCGAAATTTGAGAGG	15426228	15426248
74	11	26892	GCATATCACACACCCTTCTCTCG	TTTAGGTGCAAGTGAACCTTGG	19554816	19554837
75	11	27377	CGATTGCGATGGCGATTAGG	GAATCGGTGGAAGAGGTGACG	28301696	28301734
76	12	27507	TGTATAGCCCGGAAGTATGATCC	TCTGGTCTCGTCTCTCATGTGC	2200847	2200945
77	12	27812	TACCAACCGATGCAGATCAGAGC	CAAACCCGGACTCCAAATTAAACC	7426407	7426481
78	12	27869	AAGTAGTGGTCGAAGTGTGTATCG	GGTGAATGGTCAAGTGACTTAGG	8828359	8828408
79	12	28102	CACTAATTCTTCGGCTCCACTTTAGG	GTGGAAGCTCCGAGAAAGTGC	15955709	15955748
80	12	28367	CGTATCTCCACCTCCCGAGAAGC	GCCAAATCTCACGGATCGAAGC	21242611	21242652
81	12	28819	GAACGTCTCGTTCCCTATCACG	TCCACTCACTCATCTCTCCTTGC	27397433	27397453

3.3.2 *TFS1* was located on chromosome 3 by primary mapping

The F2 population used for map-based cloning was generated from the cross between *tfs1* mutant and *nipponbare*. Statistic analysis has been performed in F2 population with 20000 individuals and the sterility segregation is always accompanied with abnormal hull, the segregation ratio was very close to 3:1 (fertility: sterility), we also crossed *tfs1* with many other varieties including *indica*, *japonica* and hybrid rice, the sterile phenotype was never separated with abnormal hull, suggesting both phenotypes were controlled by the same locus, a single recessive nuclear gene.

40 individuals of mutant phenotype and 40 individuals of 4266's phenotype were pooled respectively and were used for primary mapping. Both pools were examined by 81 polymorphic SSR markers which distribute evenly on 12 chromosomes. The SSR marker 15576 near the centromere region, which was flanked with 15295 and 15861 on chromosome 3, was linked to the target phenotype, suggesting the mutant locus was on chromosome 3 (Fig.3-7).

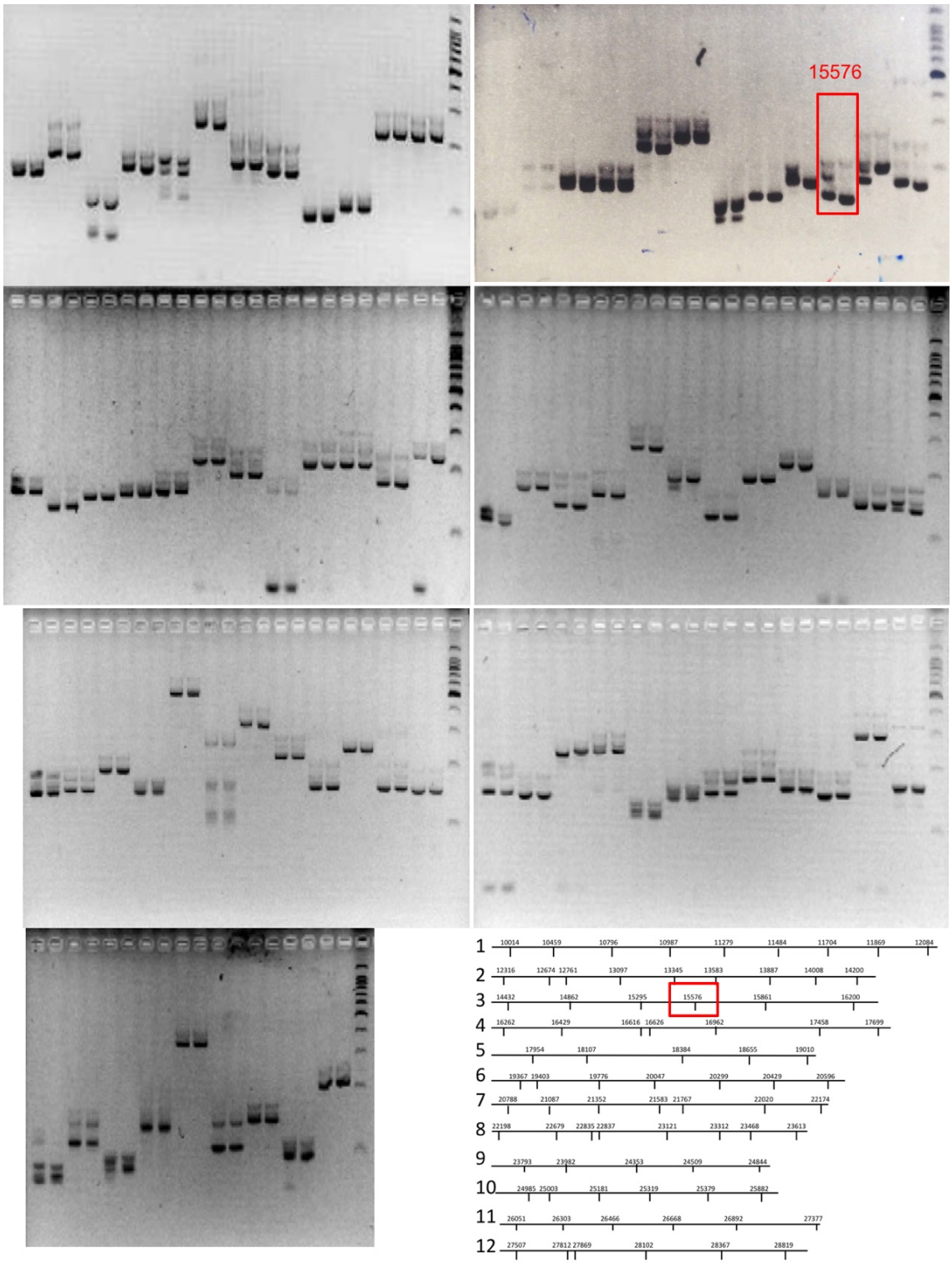

Figure 3-7. Primary mapping result by using SSR markers

DNA from two parents *nipponbare* and *tfs1*, and two DNA mixed pools from 40 individuals of fertility plant and female stertility plant, respectively, were used for the examination of selected SSR markers. The SSR marker, 15576 which locates on chromosome 3, was found to linked to the phenotype.

3.3.3 Fine mapping for the *TFS1* locus

DNA from 40 individuals with mutant phenotype were examined singly by more SSR markers between 15295 and 15861, the number of recombinants were decrease gradually from 15861(21/40) to 15295(0/40), indicating the target gene locates before marker 15295 (Fig.3-8A). Then more SSRs before 15295 were used to examine those 40 individuals again, which located the target locus between 15236 and 15353 (Fig.3-8 B).

To narrow down the region, we generated 47 SSR markers (F1~F47) between 15236 and 15280, 85 SSR markers (R1-R85) (supplementary table 3-1) between 15353 and 15316 to examined 1000 more F2 individuals of female sterility phenotype and located the gene between F47 and R45(supplementary table 3-1), a 2.17M region. However, because the recombination rate is very low, we have used 4000 more F2 individuals of defective phenotype for fine mapping but failed to narrow down the target gene in a smaller region.

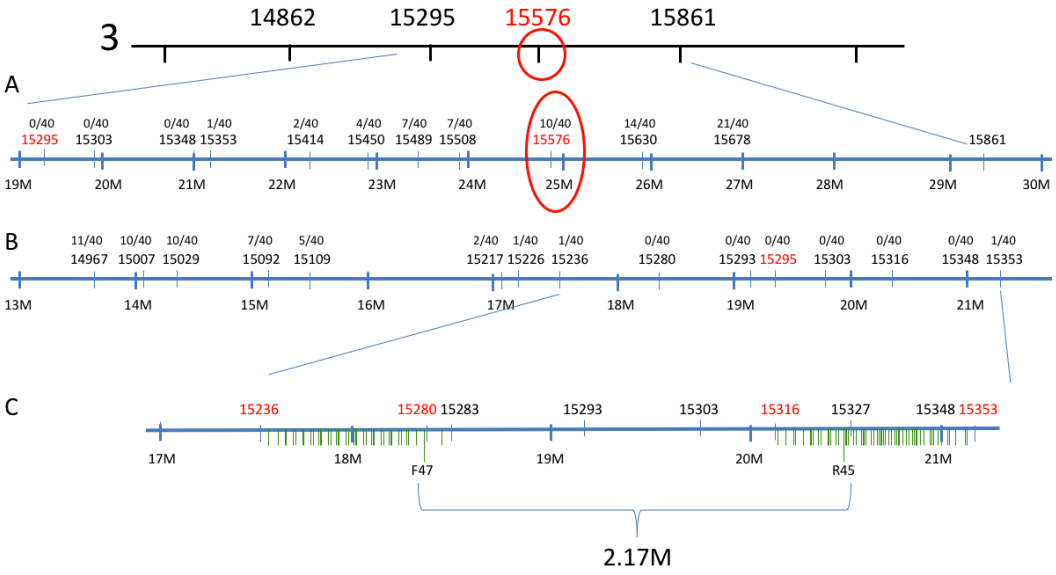

Figure 3-8. Location of SSR markers and mapping regions of fine mapping.

(A) SSR markers between 15295 and 15861 were used for cross over calculation among 40 individuals, and target locus localized before marker 15295. (B) More SSR markers used for cross over calculation before marker 15353 and locate the target locus between 15236 and 15353. (C) More SSR makers were used for large population fine mapping, and narrow the gene in a 2.17M region between marker F47 and R 45. The green short lines show the location of SSR markers.

Table 3-2. The name and the sequence of selected SSR markers for fine mapping

No.	Chromosome	SNP Marker	Forward Primer	Reverse Primer	SSR start	SSR end
1	3	15303	GAATCGGGTCTACGGTTTAGG	AAAGGAAGAGAAGAGGCAACG	19700251	19700295
2	3	15348	CGGCGTAAGTGGTGAGTAGTGC	CCCATTCTTGGTTTCTCACTAAGC	21015101	21015148
3	3	15353	GGGTCAGCTGCACACATAGAACC	TCTTCAGCTCTCCCTCTCCTTCC	21193831	21193876
4	3	15414	TCCATCATATGCTCTGCTCTCTGC	CCTCCCTTCCTCCAGATCACC	22339907	22339948
5	3	15450	AATGGGTGCAGTTAACCAAACG	TGATGGTCAAAGGATCATCAGG	22934092	22934123
6	3	15489	CACCTTCTTGAGAAGCTCCTTCG	GACATCGAGAGCGAGGACAGC	23624397	23624423
7	3	15508	GCACCAGCAGCAGCTAGAACG	AGACGGACAGAGTTTCTCTCCTTAGC	23870678	23870709
8	3	15630	AACTCGAAGGATCTCGCCCAACC	ACCCACCTCCTCACGCTGTACG	25880945	25881054
9	3	15678	CCTTAGCACACGTGACACATACTAGC	TCTTTGGATTAAGGGTGGTTGC	26867396	26867421
10	3	14967	TCCTATACTTGTCTGTCCATCGATCC	GTTCCTGCACAACAACAACAACC	13574545	13574572
11	3	15007	CAACAGCTATCTGCTAGGTTCG	CTCATCAAGTCAGACAGAGTGG	14075551	14075580
12	3	15029	GATTCCTATGGAGGATTGTTGC	AACTCCACCGGTGTTAAGAAGG	14341066	14341092
13	3	15092	GCATGCTACGGAACAATTAACACG	ACGGTGGCAAGATCACATCG	15345397	15345417
14	3	15109	ATCGCTCGATTCCTTTCCATCC	CTGCAGTTCAAGGCGGTAGTACG	15494688	15494708
15	3	15217	AAGAACCCACCTGCGGTTAGC	CTACAGCTTTCTTGATTCGCTTGG	17098286	17098357
16	3	15226	CCCTACACTACGCTCACAAACAAACG	TAGAGCGCCGATATCCCGAAGC	17264450	17264501
17	3	15236	CACTCCCTCCTCTCTCCTCTCC	GTTGGTTGGTCGGTTGCTTACC	17520725	17520748
18	3	15280	AGTTGAAGTGTTGACCGCAATCC	GTAGGTAGAGGCTCTGGCAGTCG	18441927	18441947
19	3	15293	ATCGGCAAGCAATGTCATCAAGC	GTGCATGTGCAAGACACGTTCC	19218577	19218626
20	3	15295	GGAGGGCGTACATGTCAATAAGG	CCAGACCGGGACTTCTCATAGG	19324752	19324783
21	3	15303	GAATCGGGTCTACGGTTTAGG	AAAGGAAGAGAAGAGGCAACG	19700251	19700295
22	3	15316	GTCCATACCAACTCACATCAATCC	CAGTTCCAAGTTGTATGGTAGACTGG	20234357	20234376
23	3	15283	GCTACAAATAGCTGCAAACTGC	TTGGACTAGCCTTTGACTGAGG	18522253	18522280
24	3	15327	CTGTGTGTAGTGGCATGTTACC	AACATTCCCACCTCTATCTAGC	20604056	20604087

3.3.4 MutMap reveals the mutation locus

We plotted SNP-index distributed on all 12 chromosomes and focus on which the SNP-index is around 1 (Table 3-3). In pool 1 (Fig.3-9), two SNPs at chromosome 3 locate at centromere region between 19M to 21M consistent with the mapping region of Map-based cloning. The plot between 19M to 20M is a substitution of C-to-A in *tfs1* and all the individuals in pool1, lead to a non-synonymous mutation in an AGO gene. The other plot between 20M to 21M is locate at gene intergenic region. The SNP-index results of pool 2 (Fig.3-10) and pool3 (Fig. 3-11) furthermore confirm the result, thus locate the mutation locus as the candidate gene, which is an Argonaute gene. However, there were two regions between 26Mb to 28Mb region and 23Mb to 25Mb region on chromosome 8 and chromosome 11 respectively, had obvious cluster plots in 3 pools, but not means that this region is linked with the mutative phenotype since many plots in this two region have very low SNP-index. According to our Map-based cloning result, we can exclude these plots.

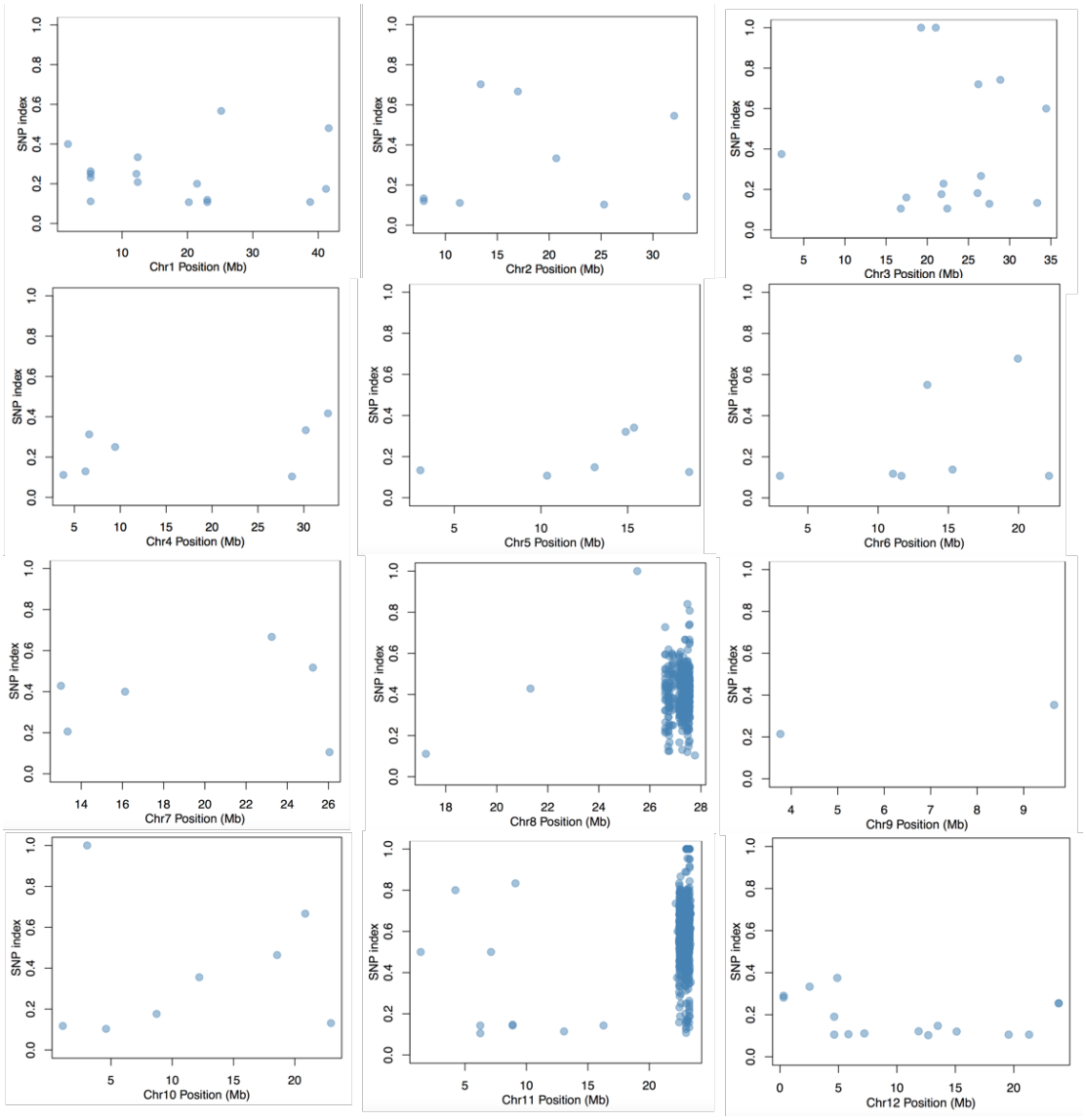

Figure 3-9. SNP index plots distribute on 12 chromosomes in pool1.

In pool 1, two SNPs locate at chromosome 3 with SNP index around 1. One is between 19M to 20M, another one is between 20M to 21M consistent with the mapping region of Map-based cloning, a substitution of C-to-A in *tfs1* and all the individuals in pool1, lead to a non-synonymous mutation in an AGO gene. The other plot between 20M to 21M is locate at gene intergenic region.

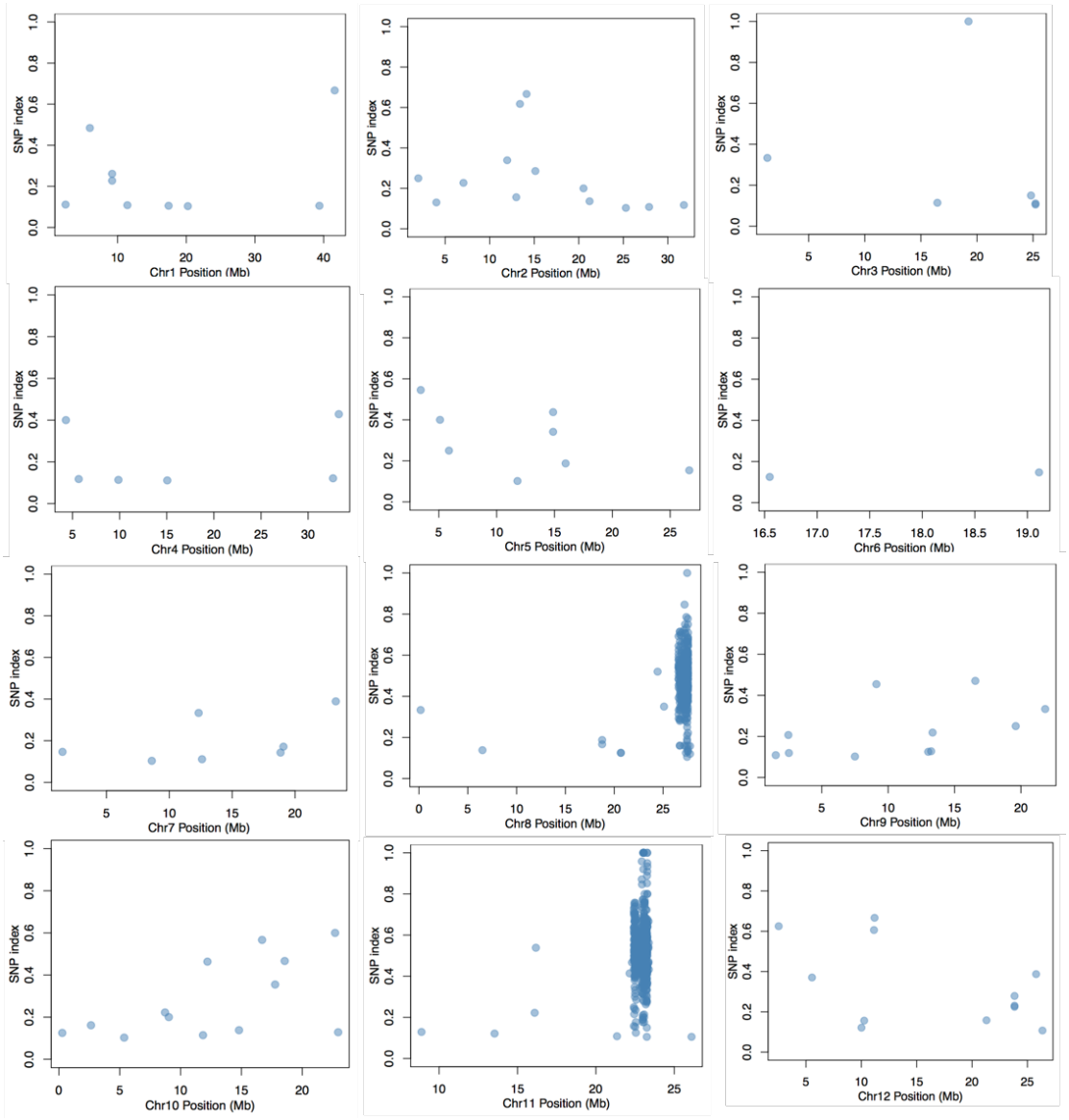

Figure 3-10. SNP index plots distribute on 12 chromosomes in pool 2.

In pool 2, a SNPs locate at chromosome 3 with SNP index around 1, between 19M to 20M, within the mapping region of Map-based cloning, a substitution of C-to-A in *tfs1* and all the individuals in pool2, lead to a non-synonymous mutation in an AGO gene.

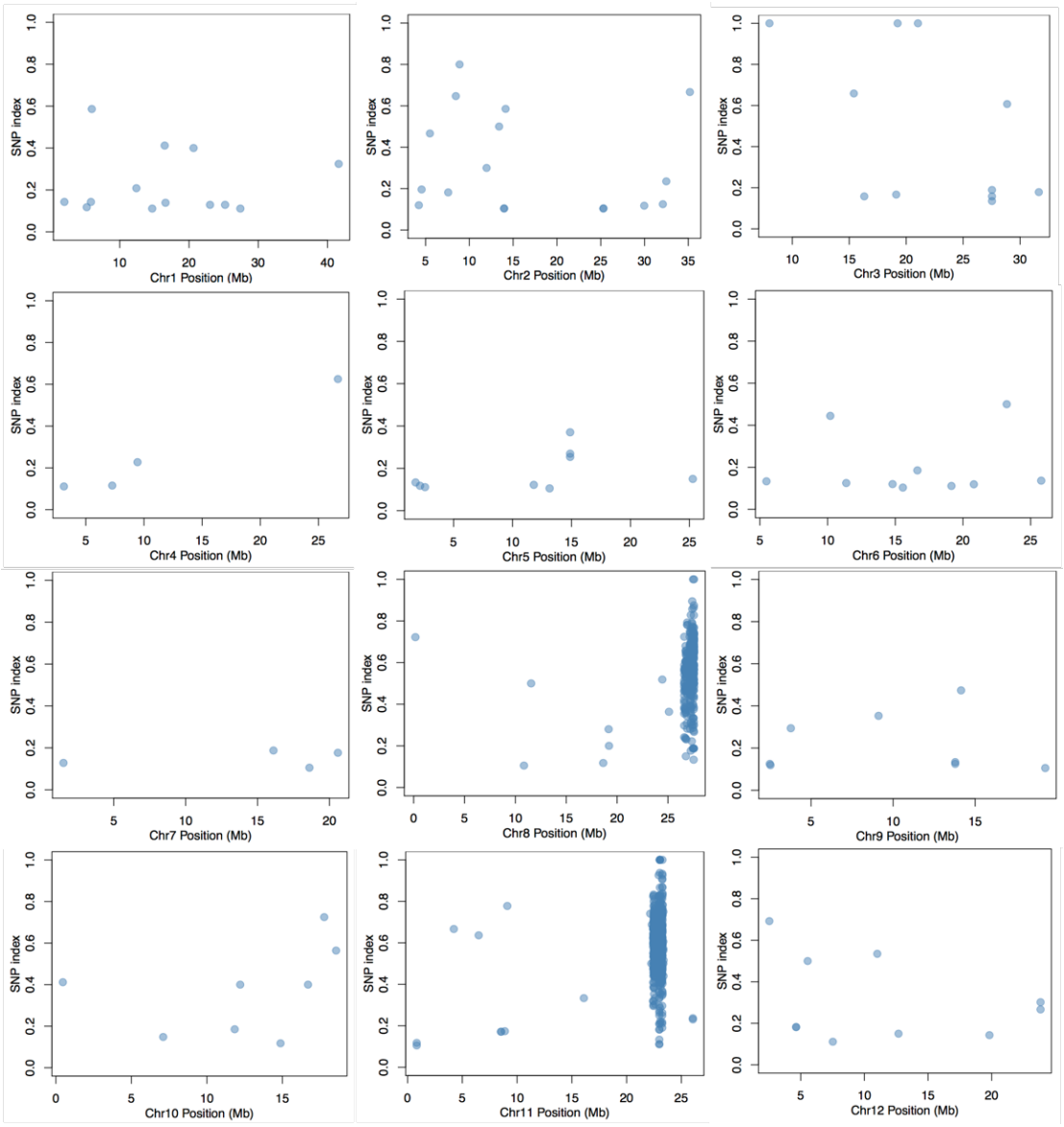

Figure 3-11. SNP index plots distribute on 12 chromosomes for p3

In pool 3, two SNPs at chromosome 3 locate with SNP index around 1. One is between 19M to 20M, another one is between 20M to 21M consistent with the mapping region of Map-based cloning, a substitution of C-to-A in *tfs1* and all the individuals in pool3, lead to a non-synonymous mutation in an AGO gene. The other plot between 20M to 21M is locate at gene intergenic region.

Table 3-3. The list of selected plots with SNP index exceeding 0.9

Chr	Position	Index	REF	ALT	SNP Effect

3	19-20Mb	1	C	A	EFF= NON SYNONYMOUS CODING (MODERATE| MISSENSE| Ctt/Att | L603I | | LOC Os03g33650 | | | LOC Os03g33650 .1|3|1)
3	21-22Mb	1	G	A	EFF=DOWNSTREAM(MODIFIER | | 1 1 7 6 | | | LOC Os03g37864 |||LOC Os03g37864.1||1) DOWNSTREAM(MODIFIER | | 3 9 3 8 | | | LOC Os03g37869|||LOC Os03g37869.1||1) DOWNSTREAM(MODIFIER | | 4 6 8 5 | | | LOC Os03g37878|||LOC Os03g37878.1||1) INTERGENIC(MODIFIER | | | | | | | | | | 1)
8	25-26Mb	1	C	A	EFF=DOWNSTREAM(MODIFIER | | 1 6 3 | | | LOC Os08g40290 |||LOC Os08g40290.1||1) DOWNSTREAM(MODIFIER | | 2 9 8 9 | | | LOC Os08g40280|||LOC Os08g40280.1||1) INTERGENIC(MODIFIER | | | | | | | | | | 1)
8	27-28Mb	1	C	T	EFF=DOWNSTREAM(MODIFIER | | 1 6 0 3 | | | LOC Os08g43430|||LOC Os08g43430.1||1) INTERGENIC(MODIFIER | | | | | | | | | | 1)
10	30-31Mb	1	C	T	EFF=INTERGENIC(MODIFIER | | | | | | | | | | 1) UPSTREAM(MODIFIER||1659||| LOC Os10g06030|||LOC Os10g06030.1||1)
11	22-23Mb	1	A	G	EFF=INTERGENIC(MODIFIER | | | | | | | | | | 1)
11	23-24Mb	1	C	A	EFF=DOWNSTREAM(MODIFIER | | 1 0 2 2 | | | LOC Os11g38750|||LOC Os11g38750.1||1) INTERGENIC(MODIFIER | | | | | | | | | | 1)
11	23-24Mb	1	C	G	EFF=INTERGENIC(MODIFIER | | | | | | | | | | 1) UPSTREAM(MODIFIER||2600||| LOC Os11g38760|||LOC Os11g38760.1||1)
11	23-24Mb	1	G	A	EFF=INTERGENIC(MODIFIER | | | | | | | | | | 1) UPSTREAM(MODIFIER | | 8 7 2 | | | LOC Os11g38760|||LOC Os11g38760.1||.1)
11	23-24Mb	1	C	T	EFF=INTERGENIC(MODIFIER | | | | | | | | | | 1)
11	23-24Mb	1	G	A	EFF=INTERGENIC(MODIFIER | | | | | | | | | | 1)
11	23-24Mb	1	C	G	EFF=INTERGENIC(MODIFIER | | | | | | | | | | 1)
11	23-24Mb	1	C	T	EFF=DOWNSTREAM(MODIFIER | | 4 2 8 | | | LOC Os11g39070|||LOC Os11g39070.1||1) INTERGENIC(MODIFIER | | | | | | | | | | 1) UPSTREAM(MODIFIER||2087||| LOC Os11g39060|||LOC Os11g39060.1||1) UPSTREAM(MODIFIER||4843||| LOC Os11g39050|||LOC Os11g39050.1||1)TFS
11	23-24Mb	1	T	C	EFF=INTRON(MODIFIER | | | | | LOC Os11g39079 |||LOC Os11g39079.1||1|1) UPSTREAM(MODIFIER||1432||| LOC Os11g39090|||LOC Os11g39090.1||1)
11	23-24Mb	1	A	G	EFF=DOWNSTREAM(MODIFIER | | 1 6 0 0 | | | LOC Os11g39110|||LOC Os11g39110.1||1) INTERGENIC(MODIFIER | | | | | | | | | | 1) UPSTREAM(MODIFIER||1018||| LOC Os11g39100|||LOC Os11g39100.1||1)
11	23-24Mb	1	T	A	EFF=DOWNSTREAM(MODIFIER | | 1 3 3 7 | | | LOC Os11g39110|||LOC Os11g39110.1||1) INTERGENIC(MODIFIER | | | | | | | | | | 1) UPSTREAM(MODIFIER||1281||| LOC Os11g39100|||LOC Os11g39100.1||1)
11	23-24Mb	1	G	A	EFF=INTERGENIC(MODIFIER | | | | | | | | | | 1) UPSTREAM(MODIFIER||1537||| LOC Os11g39120|||LOC Os11g39120.1||1) UPSTREAM(MODIFIER||4719||| LOC Os11g39130|||LOC Os11g39130.1||1) UPSTREAM(MODIFIER||4719||| LOC Os11g39130|||LOC Os11g39130.3||1) UPSTREAM(MODIFIER||4719||| LOC Os11g39130|||LOC Os11g39130.4||1)

3.3.5 Confirmation for candidate gene by complementation transgenic plant

To confirm the candidate gene, we have constructed two kinds of vectors for transgenosis. One is the overexpression of *TFS1* gene in *tfs1* mutant, and the other is the complementation experiment by using the whole genomic DNA of *TFS1* gene in *tfs1* mutant. Also we have tried to knock out the candidate gene by CRISPER/Cas9 system. After the transgenosis, we have got 11 RNA over expression transgenic lines in the background of *tfs1* rice. However, none of them has been restored. To find out the reason, we have amplified the DNA fragment which covers the mutative site by PCR and sequenced the PCR products. Results shown that all the OE line contain two kinds of base, one from the *tfs1* itself and the other from transferred cDNA from 4266 plant. However, the OE lines don't have a higher expression level of *TFS1* gene may due to the co-suppression in transgenic plant (Fig.3-12). These results may explain why the phenotype of *tfs1* is not restored by cDNA overexpression.

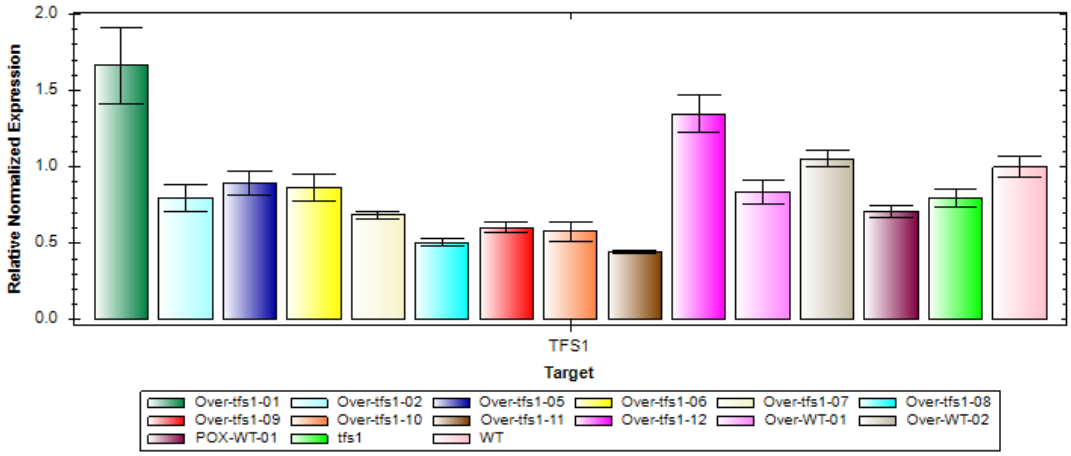

Figure 3-12. Expression level of *TFS1* gene in overexpression transgenic lines.

Expression level were detected by the primer near the 3'UTR region. Leaves from individual plant of those transgenic lines were collected for RNA extraction. Experiments were performed for three times.

Then we have got several CRISPER/Cas9 transgenic lines about *TFS1*, which was designed to knock out the target gene near the mutative position. Interestingly, one of the lines displayed a similar phenotype with *tfs1* mutant (Fig. 3-13 D). DNA sequencing at the SG position for CRISPER/Cas9 shows that the other lines without any phenotype were heterozygotes, which might have phenotype in its progeny because of segregation. However, for the line with similar phenotype, the allele of target gene shown a different

mutation patterns at the same gene after CRISPER, one is a base insertion while the other is an amino acid deletion near the mutant position. In this plant, only a TFS1 protein with one amino acid deleted was produced, and thus bearing a similar phenotype with *tfs1* mutant. This result primarily proves that the candidate gene has a big possibility to be *TFS1*, which need to be verified by complementation lines for this gene.

Since the whole gene of *TFS1* is 10 kb, including the promoter and introns, it was much harder to amplify this gene than the cDNA. So that it took a little bit longer to get the transgenic line for complementation lines. Fortunately, we have found that most of complementation lines had been restored from the *tfs1* mutant phenotype. As shown in figure 3-13 C, the size of the panicle form transgenic plant is similar with *tfs1* (Fig. 3-13 B) whereas both the defective hull and fertility were restored. Our results indicate that the candidate gene is *TFS1* gene.

Figure 3-13. The panicle form 4266, *tfs1* plant, and transgenic plants under normal condition.

(A) Panicle form 4266 plant shows normal fertility. (B) Panicle form *tfs1* plant shows total sterile with defective hull. (C) Panicle form transgenic complementation plant is restored and shows similar phenotype with 4266 plant. (D) Panicle form CRISPER plant show similar phenotype with *tfs1*, also sterile with defective hull.

3.3.6 TFS1 was located in cytoplasm

In order to study the biological function of *TFS1* gene, a fusion protein with TFS1 and GFP were prepared. The subcellular localization of TFS1-GFP fusion in transformed protoplasts was shown in figure 3-14, A1-D1 show the signal of TFS1 from 4266 plant while A2-D2 show the signal of TFS1 from *tfs1* plant. A1 and A2 show that the GFP signal is localized at cytoplasm, while no signal is detected in both the plasma membrane and nuclear. It is clear show the chloroplast auto-fluorescence in both B1 and B2. C1 and C2 show the merged image of GFP and chloroplast auto-fluorescence. The signal of GFP has no overlap with chloroplast auto-fluorescence. These results demonstrated that TFS1 is located in cytoplasm.

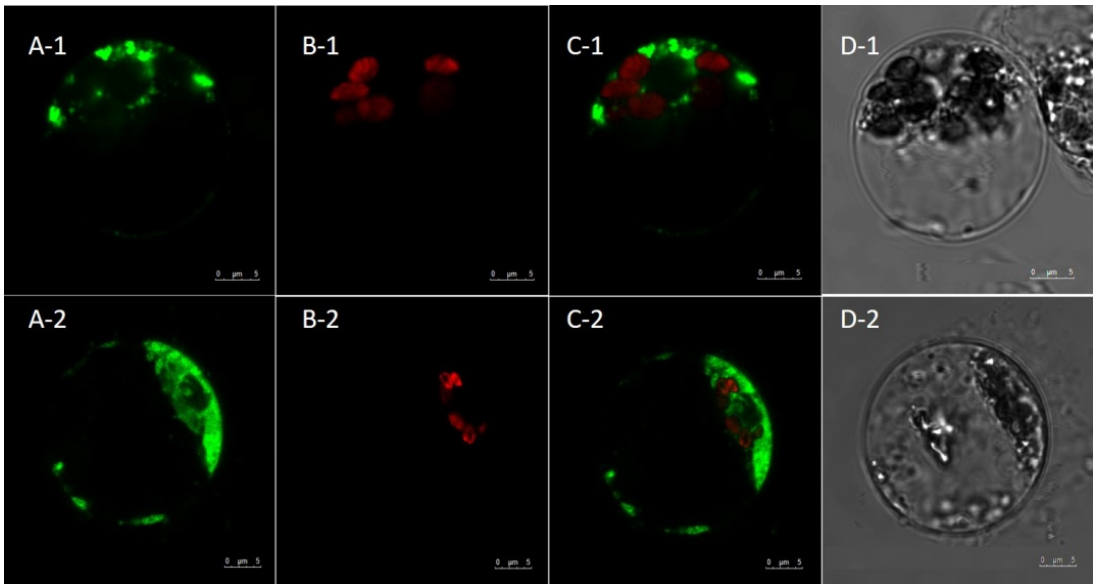

Figure 3-14. Subcellular localization of TFS1-GFP fusion proteins in transformed protoplasts.

(A1-D1) The signal of TFS1 form 4266 plant. (A2-D2) the signal of TFS1 form *tfs1* plant. (A)The fluorescent TFS1-GFP signal (green) localized at cytoplasm, no signal detected at nuclear region and membrane. (B) Chloroplast auto-fluorescence (red). (C) Merged signal of GFP and chloroplast auto-fluorescence show the TFS1-GFP signal (green) has no overlap with chloroplast auto-fluorescence (red). (D) Light microscope view.

3.4 Discussion

Map-based cloning is an effective forward genetic technique to identify genes underlying phenotypic variations. Map-based cloning which relies on the construction of segregation population and the molecular markers for linkage analysis has been widely applied to different species and many functional genes have been reported in the past decades. In rice, many important characteristics of great agronomic application are identified by this method. As mentioned above, the cloning of photo- and thermo-sensitive male sterility gene is a great breakthrough in hybrid rice (Ding et al., 2012; Zhou et al., 2012). Recently, Wang et al (2012) has ingeniously applied this technique in cloning a gene, *OsSPL16*, that control of grain size, shape and quality in rice by using a NIL line. After the identification of *OsSPL16*, the author further cloned a new gene, *GW7* also by using map-base cloning, which acts with *OsSPL16* as a regulatory module to determine grain shape and simultaneously improves rice yield and grain quality (Wang et al, 2015). Their results were highly praised by Professor McCouch from Connell University owing to the simultaneously improves yield and grain quality, which is a pair of conflit trait that makes the improvement of yield and grain quality at the same time a very difficult problem. More and more important gene for plant breeding are cloned by map-base cloning indicating it an effective forward genetic technique in identifying new genes, although it has the disadvantages of time-consuming and labor demanded. Moreover, a dense genetic marker developed for the nutation locus will be of great help for cloning of the mutant gene, thus the molecular marker has been greatly developed and applied to map-base cloning, such as SSR, SNP and Indel.

In this study, 81 SSR markers well-distributed in rice genome were selected to the primary cloning. A small F2 population with 40 individuals was used to screen these SSR markers and result shown that one of the markers from chromosome 3 was linked to the mutant phenotype, indicating the mutant locus was in chromosome 3. More SSRs around this marker was developed to primarily locate the mutant locus, and two markers, 15236 and 15353, were found to link with mutant locus. Unfortunately, the centromere is included inside the region between these two markers, which make our map base cloning more difficult owing to the extremely low recombination rate at centromere region (Hou et al., 2010; Lukowitz et al., 2000). Thus, for the fine mapping, we selected

another 132 SSRs between these two markers to examine a large scale F2 population. Results shown that we can only locate the mutant locus to a 2.17M DNA region between R45 and R47 even though we have enlarged the F2 population to 20,000 individuals (Fig. 3-8), further prove that the recombination rate in centromere region is extremely low.

Next generation sequencing has been widely used in gene cloning a few years ago, which has greatly accelerated the cloning progress and improved the accuracy. Different strategies have been developed for various mutant phenotypes. Among those frequently used sequencing techniques, MutMap is a recently developed sequencing technique, which is also efficient and affordable compared with traditional methods. Base on whole genome sequencing (WGS), the MutMap has been applied in many species for loci studies. It is an efficient cloning technique without the construction of mapping population, not depending on genetic linkage map and the development of huge number of molecular markers. By using SNP analysis, Abe et al (2012) first time used MutMap to locate 7 mutant genes from 7mutnat lines which were crossed their parent line and create three DNA pool with around 30 individuals' DNA according to the mutant phenotype in F2 population. In their experiment, SNP-index has been applied, in which all reads that cover a specific DNA region containing the same SNP which is different with the reference genome was set as 1, whereas only half reads containing this SNP will be set as 0.5. With this SNP-index, the analysis result can be plotted and that the SNP near the mutant location are usually higher than others. The MutMap has been successfully used in many species and cloned many important genes, such as some important agronomic trait genes, rice male sterility gene, salt-tolerant gene and dwarf gene.

In our study, we have located the mutant position into a 2.17M DNA region on chromosome 3. Due to the low cross over rate in centromere region, we could not make any progress in reducing the distance. Then the MutMap was used to clone the target gene, result shown that those SNP-index of 1 were found only in part of the 12 chromosomes (Fig. 3-9), which was also listed in table 3-3. As we can see although the chromosome 11 has the most candidate of SNP, two SNP with index of 1 were found in near the centromere region, however, only one of which was contained in the DNA region of our map-base cloning result (Fig. 3-8). By combination map-base cloning

with MutMap cloning, we have accelerated our mapping progress and cloned a candidate gene for *TFS1*, which is an Argonaute gene. Since it is only a SNP between *tfs1* mutant and 4266 plant in this candidate gene, we constructed two kinds of transgenic plants to verify the candidate, one is complement line with whole genome DNA of target gene, including the intron and promoter, the other is over-expression line with cDNA of candidate gene, both recombined vectors were transfer into *tfs1* mutant rice. Unfortunately, none of the OE lines has restored the sterility phenotype of *tfs1* plants. The transgenic plants were examined by sequencing the DNA around the mutative position. Results shown that all OE lines contain two kinds of DNA from *tfs1* mutant and 4266 plant, suggesting the successful transgenic plants were produced. To find out the reason why over expression of *TFS1* can't restore the phenotype, we then examined the expression of candidate gene by qRT-PCR, result shown that almost all of the OE lines had a low expression level of *TFS1* gene when compared with 4266 plant (Fig. 3-12), which may explain why those OE lines didn't display a restored phenotype. In transgenic plant, over expression often result in a co-suppression effect (Jorgensen, 1995). The same phenotype of these OE plants with *tfs1* mutant might be resulted by co-suppression. Then we have got a CRISPER line which displayed a similar phenotype with *tfs1*. DNA sequencing revealed that this line can only produce an amino acid deletion TFS1, which resulted in a similar phenotype with the mutant line, indicating that this candidate gene could be the mutant gene *TFS1*. Later we have got several complement lines with the whole candidate gene from 4266 plant. As we expected, the female sterility phenotype has been restored in these complementation lines, which prove that the candidate gene is *TFS1* gene, and that the mutant gene is cloned.

Argonaute proteins have a complex structure and the structure has been well documented in animals, which is mainly contain six domains. Many mutations in these domains have severe phenotypes such as growth and development as well as the responses to abiotic and biotic stresses. Further studies reveal that AGO protein is not only regulate gene expression, but also participate in miRNA biogenesis (Mi et al., 2008). In plant, it has been reported that guide RNAs mediates the cleavage of target mRNA relies on high degree sequence complementary to their target mRNA. According to recorded functions of AGO protein, the following researches will be focus

on the gene expression variation and the level of small RNA between the *tfs1* and 4266 plants under normal and low temperature conditions.

Chapter 4 Transcriptome and sRNA analysis of *tfs1*

4.1 Introduction

Female sterility is widely spread in plant world and may be caused by various reasons. Ovule development and double fertilization have been a research focus for many years, and many genes that control these two processes have been identified in Arabidopsis (Shi and Yang, 2011). However, the mechanisms that control ovule development are still far from being illustrated. In rice, although the research on male sterility studies has been of great achievement(Breakfield et al., 2012), little is still known about the female sterility, especially the photo/thermo-sensitive female sterility (P/TMS). Fertility alternation of plants in response to environmental cues is a complicated process, which has been first reported in Professor Qifa Zhang's lab (Ding et al. 2012, Zhou et al. 2012). After the reports from these two groups, no other research on environment sensitive line of rice has been reported. Thus it is of great importance to further explore the controlling mechanism by which plant will change its fertility in response to environmental cues.

Due to the high through output and accuracy, the next generation sequencing technology has been applied to all aspects in science researches. As the cost of this technology goes down, it became a popular research method all over the world. Among those technologies related to sequencing, RNA-seq is the most popular and frequently used technology in researches of life science. In recent studies, RNA-seq can be used to verify transcription event happened in the cell. For instance, it's verified that RNA-seq reveals RNA transcript alternative splicing during the transcription process. Furthermore, RNA-seq can be applied to study the modification at the level of post-transcription. Besides, gene expression difference of different sample or treatment comparison can be analyzed by RNA-seq. Additionally, since RNA sequencing can detect almost all kinds of RNAs, it is also used to examine the different populations of RNA, including the total RNA, small RNA, mRNA, lncRNA, and so on. Thus it is a powerful technology for the study of RNA and an ideal method to reveal the effect of mutation gene on a global RNA profile.

Argonaute proteins have been well studied due to its fundamental function in different biological processes (Tiwari et al., 2003). It has been reported that AGOs are involved

in affecting the growth and development as well as in response to abiotic and biotic stresses (Douglas et al., 2010). Further studies reveal that AGO proteins not only regulate gene expression, but also participate in miRNA and siRNA biogenesis (Mi et al., 2008). Gene expression regulation has TGS and PTGS which are associate with different AGO members The former complexes involve AGO4, AGO6, and AGO9 and their associated 24-nucleotide small RNAs whereas the latter complexes is composed of the cleavage-competent, AGO1 and AGO7 proteins, as well as their associated 21- to 22-nucleotide small RNA (Nagasaki et al., 2007). In plants, AGO2, AGO4 and AGO10 also have been proved have slicer activity which may have function in PTGS (La Rota et al., 2011). For the ovule morphogenesis, various studies have document that RNA interference pathway is indispensable. Besides, small RNA is also an important regulating factor for the ovule development (Harper et al., 2000; Achard et al., 2006).

In this study, we have previously identified a sterility germplasm. Primary studies indicated it is a temperature sensitive female sterility line and controlled by a recessive nuclear gene. After several years hard work, we have cloned this mutant gene, which belongs to AGO gene family. To further explore the controlling mechanism for this mutant gene on female fertility, we have used the RNA-seq, including transcriptome sequencing and small RNA sequencing, to investigate the RNA profiles in *tfs1* and 4266 plants at different growth stages. By transcriptome and small RNA analysis, as well as the correlation analysis, dynamic gene expression was used to establish a new insight into the mechanism of female development. In our study, the transcriptome and small RNA sequencing were applied to find the candidate small RNAs which have significant different expression between 4266 and *tfs1* plant at flowering stage. Thus figure out which step is interrupted during double fertilization and the regulating mechanism of temperature sensitive. By analyzing the data of small RNA sequencing and RNA sequencing, we identified the miRNA and siRNA in our sample, and analyzed the abundance of different siRNAs and miRNAs in all the samples to find the sRNA that have correlation with the mutation phenotype, and then found their targets, combined the results of RNA sequencing data, to do GO analysis, to find out which pathway is related to the mutant phenotype.

4.2 Materials and methods

4.2.1 Plant materials

Plant materials and growth condition were described as above. *tfs1* and 4266 grow under lower temperature condition (25-27°C) were used for sample collection. Florets before and after fertilization were sampled separately.

4.2.2 Sample preparation and RNA extraction

Six samples with 3 duplications: 4266 LT-uf (unfertilized 4266 florets sampled in lower temperature: 25-27 ℃), 4266 LT (fertilized 4266 florets sampled in lower temperature: 25-27 ℃), MLT-uf (unfertilized *tfs1* florets with mutant phenotype sampled in lower temperature: 25-27 ℃), MLT (fertilized *tfs1* florets with mutant phenotype sampled in lower temperature: 25-27 ℃), MR-uf (unfertilized *tfs1* florets with normal phenotype sampled in lower temperature: 25-27 ℃), MR (fertilized *tfs1* florets with normal phenotype sampled in lower temperature: 25-27 ℃) were collected for total RNA extraction, each sample composed of about 30 Florets collected from 10 panicles. The unfertilized florets refer to the 1day before pollination and the fertilized florets refer to 1hour~25 hours after pollination. Total RNA extracted for mRNA and sRNA sequencing by using Purelink plant RNA reagent (Thermo Fisher) accord to the manufacturer's instructions, DNA was removed by RNase-Free DNase Set (Qiagen).

4.2.3 mRNA sequencing and analysis

For RNA-seq, we prepared the libraries for each sample with 1 μg total RNA. Using the standard Illumina protocol to prepare the sequencing library. Fragmentation reagent was used to break the total mRNA into small pieces. First-strand cDNA was generated using random hexamer-primed reverse transcription, then a second-strand cDNA was synthesized and followed by end repair. Then ligated with adaptors at adenylated 3' end. This process was to amplify the cDNA fragments with adaptors from previous step by using the SPRI beads, and the product was dissolved in EB solution. The double stranded PCR products were heated to denatured and circularized by the splint oligo sequence, which was constructed as the library and were sequenced on HiSeq x ten (Illumina), 6G raw data for each sample was collected.

For RNA sequencing data processing, the raw data were trimmed to remove adaptor sequences, clean reads were mapped to ensemble rice genome reference which is dating from March, 2017 (http://plants.ensembl.org/Oryza_sativa/Info/Index) by hisat2 (Kim et al., 2015). Since our sample is the progeny of *indica* and *japonica*, we use both genome of *indica* (ASM465v1-release33) (data not shown) and *japonica* (IRGSP-release 33) as reference genomes. The gene expression metrics were calculated by using FPKM by Stringtie (Pertea et al., 2015) based on the hisat2 alignment. Differentially expressed genes between samples were detected by R Bioconductor package DESeq2 with the random sampling model based on the read count for each gene, We used $\log_2$Ratio$\geqslant$1and q<0.05 as the threshold for judging the DEGs (Love et al., 2014). Correlations between samples and PCA analysis were calculated with in-built R functions.

4.2.4 Small RNA sequencing and analysis

the same RNA sample was used with RNA-seq, microRNAs at length 18 to 30nt were isolated from 1 μg total RNA for each sample by gel electrophoresis. The library preparation is followed the standard protocol and the sequencing is on BGISEQ-500RS (Illumina) platform by sequencing company (BGI, Hongkong). More than 50 million clean reads were obtained in each of the pool.

miRNA annotation The raw data was trimmed to remove adaptor sequences and filtered, the length of these clean tags which was between 18nt and 30nt were mapped to miRBase version 21 (http:www.mirbase.org) dating from July 3, 2014 and reference genome by AASRA (Raghavan, 1988; Tang et al., 2017). The small RNA expression level is calculated by using TPM (Transcripts Per Kilobase Million) (t Hoen et al., 2008).

miRNA prediction We used RIPmiR for miRNA prediction as described (Breakfield et al., 2012). Both genome of *indica* (ASM465v1) and *japonica* (MSU7) as reference genomes were used.

miRNA target prediction: *sPARTA* package (Patel et al., 2015) was used for target gene prediction meanwhile the PARE (parallel analysis of RNA end) data (German et al., 2009) was used to identify miRNA-directed cleavage site.

Phasing analysis: All the sRNA reads of 18 libraries were mapped to rice genome for Phase loci prediction, 21-nt siRNA phased loci were predicted as previous method with a threshold P value ≤ 0.001(Arikit et al., 2014; Xia et al., 2013).

4.3 Results

4.3.1 Global gene expression trends and DEGs identification

In our study, the *tfs1* exhibit a phenotype of temperature sensitive female sterility, which can be used to investigate the molecular mechanism of ovule development and double fertilization. The RNA sequencing approach was used to perform transcriptome-profiling analysis of the floret at two developmental stages, before and after fertilization in 4266 and *tfs1*. Those genes with low confidence expression values were excluded, 13,384 genes with an FPKM value >10 at least in one sample designated as expressed, and less than 10% counts were filtered. The biological identity of tissues was well reflected in their respective transcriptomes as revealed by hierarchical clustering of Pearson's correlations among different samples based on the FPKM value. The samples clustered together based on their developmental stage and provided an obvious glimpse into the samples. For the three different duplications, the correlation value is all over 0.95 (Fig.4-1, A). The PCA analysis among the 6 samples shows that the clusters of the unfertilized samples are separate with fertilized samples, and the three duplications of a sample cluster together (Fig.4-1, B). The Pearson correlation and PCA analysis revealed that the biological background and the duplications of those samples are both reliable, indicating the data can be used for further analysis.

We identified a number of differently expressed genes (DEGs) in two groups separately and the expression of RNAs are compared by FPKM value. MR has the same phenotype with 4266LT, both show normal closed hull and can generate normal seed. MLT has the same background with MR, but showed sterility with defective hull. Group 1 is among 4266LT-uf, MR-uf and MLT-uf, which may play crucial roles in floret development, including the development of female gametophyte and hull. We selected 758 genes have DE in 4266LT-uf vs MLT-uf overlapped MR-uf vs MLT-uf, but exclude with 4266LT-uf vs MR-uf (Fig.4-1, C, Supplementary table 4-1). Group 2 is among 4266LT, MR and MLT, similar with group 1, which may be influenced by

double fertilization, including the pollen tube tip rupture, PCD of the synergid cell, and the development of zygote and endosperm. We selected 447 genes have DE in 4266LT vs MLT overlapped MR vs MLT, but exclude with 4266LT vs MR (Fig.4-1, D, Supplementary table 4-2). The heat-map in Figure 4-2 (https://software.broadinstitute.org/morpheus/) shows the expression difference among samples of those genes. Figure 4-1A shows genes expression selected in group 1, stage before fertilization, figure 4-1B shows genes heat-map selected in group 2, stage after fertilization. The color clearly shows that 4266 and MR has similar expression level at different stage.

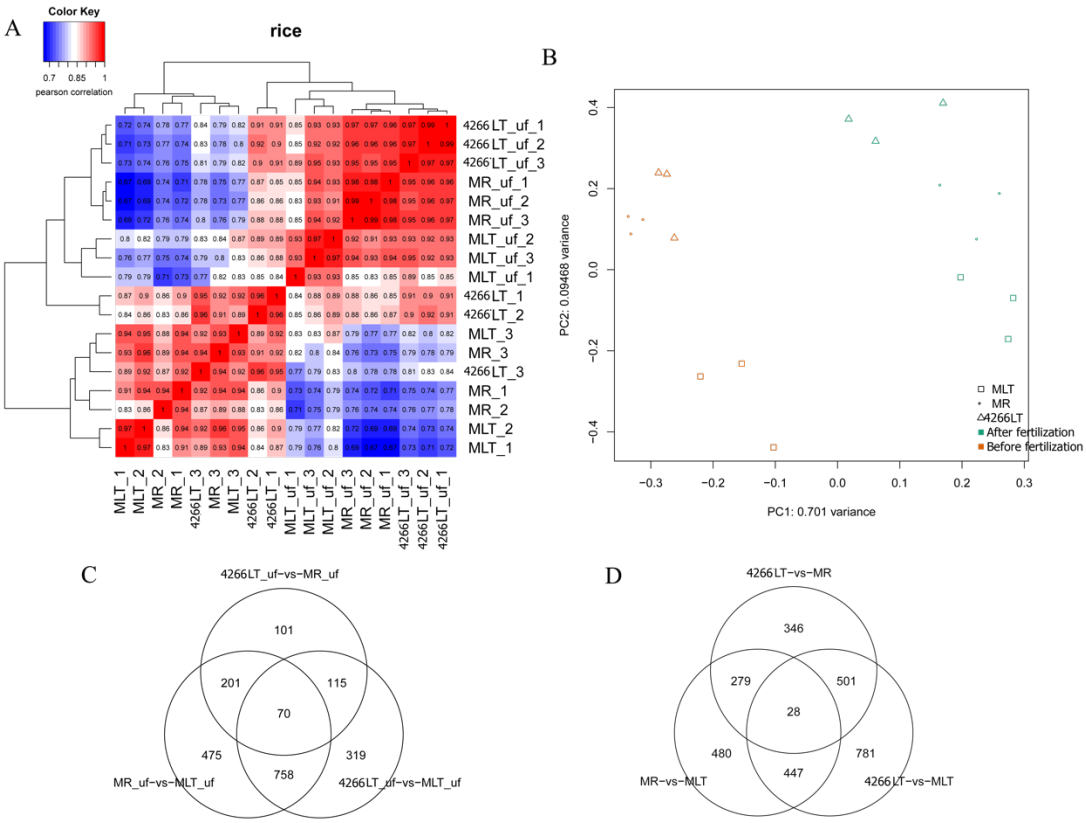

Figure 4-1. Correlation analysis of samples and DE analysis.

(A) Pearson's correlations among different samples based on the FPKM value, shows that among three duplications, the correlation value is all over 0.95. (B) PCA analysis among the 6 samples shows that the clusters of the unfertilized samples are separate with fertilized samples, and the three duplications of a sample cluster together. (C) Venn diagram shows the number of DE genes of different comparisons before fertilization. (D) Venn diagram shows the number of DE genes of different comparisons after fertilization.

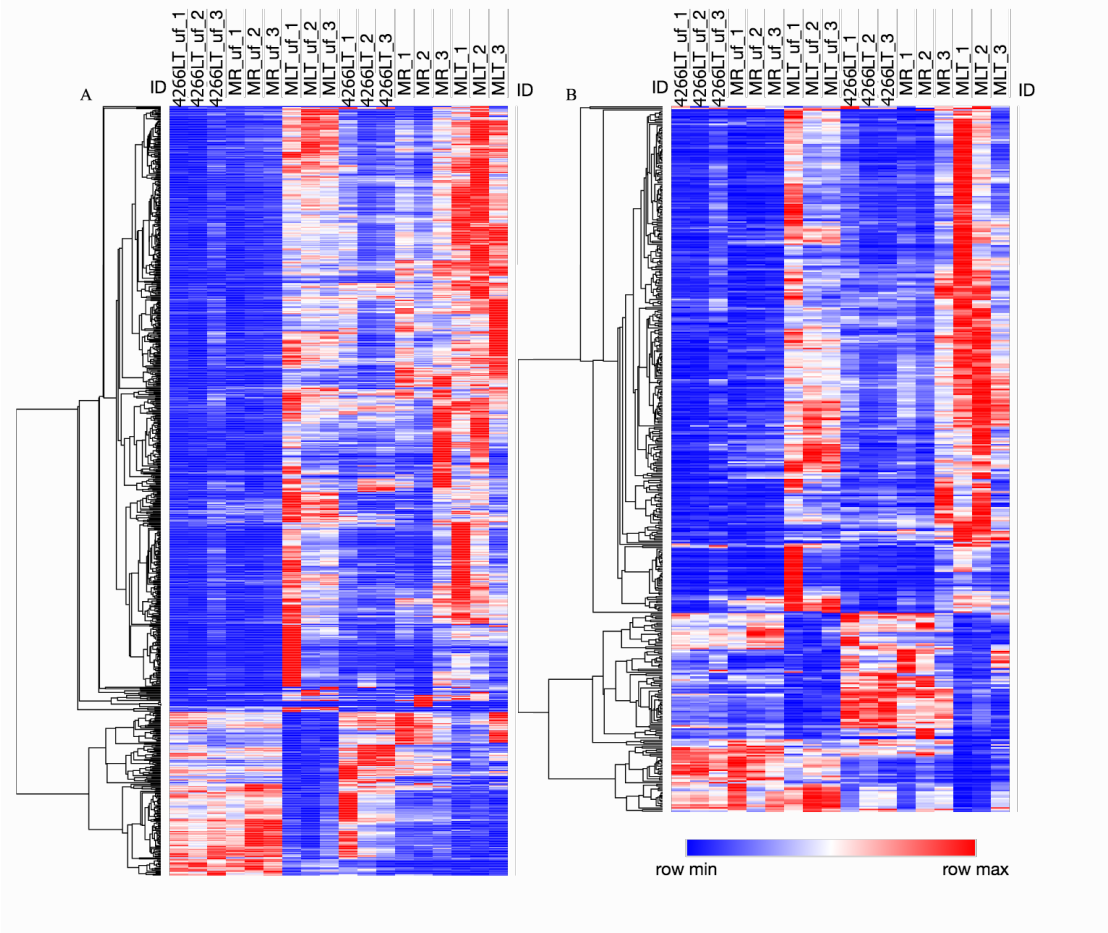

Figure 4-2. Heat map of selected genes expression level.

(A) Selected 758 genes have DE in 4266LT-uf vs MLT-uf overlapped MR-uf vs MLT-uf, but exclude with 4266LT-uf vs MR-uf. (B) Selected 447 genes have DE in 4266LT vs MLT overlapped MR vs MLT, but exclude with 4266LT vs MR.

Based on the annotation of those selected genes, we did GO enrichment analysis on agriGO (http://bioinfo.cau.edu.cn/agriGO/analysis.php) by using both *japonica* genome (MSU7) and *indica* genome (ASM465v1) (data not shown). Those 758 genes have different expression level before fertilization and 447 genes after fertilization were selected for GO enrichment. GO enrichment analysis of biological process revealed that those genes are significantly enriched in many GO terms at stages before fertilization (Fig.4-3), which is related to the mutative phenotype: such as signal transduction, embryonic development, cell differentiation, transcription and protein modification process. Those gene belong to different GO terms are not work independently, and we can also find some clue from the cellular component GO enrichment results. Embryonic development is of course related to our mutative

phenotype, many genes related to abiotic stimulus and endogenous stimulus also enriched in this stage. It is a complicated process which is regulated by many transcription factors, such as MADS-box family, HD-Zip family and ARF family. Previous studies have revealed that these TFs are regulated by Argonaute genes.

The expression level of these genes form different enriched GO terms are showed with heat maps (Fig.4-4). Signal transduction, protein modification transcription and cell differentiation have similar characteristic, only several genes have higher expression level in 4266 and MR at both before and after fertilization, most genes have higher expression in MLT-uf, which has the defective phenotype. While more genes enriched in GO term response to stress and endogenous stimulus has higher expression level in 4266 and MR which may explain why lower temperature can store the phenotype of *tfs1*. And most of those genes expression level gradually increased in 4266, MR and MLT, which may explain why only part of fertility can be restored in *tfs1* under lower temperature. Many of these genes are transcription factors, such as ARF, HD-zip, which has been reported regulated by Argonaute genes. Another large proportion of these genes are enzyme, such as LRR (Leucine-rich repeat protein kinase family protein), Leucine-rich repeats are frequently involved in the formation of protein–protein interactions; Cytochrome P450, the terminal oxidase enzymes in electron transfer chains; CRCK (calmodulin-binding receptor-like cytoplasmic kinase), which has been proved to have important function of signal transduction during development.

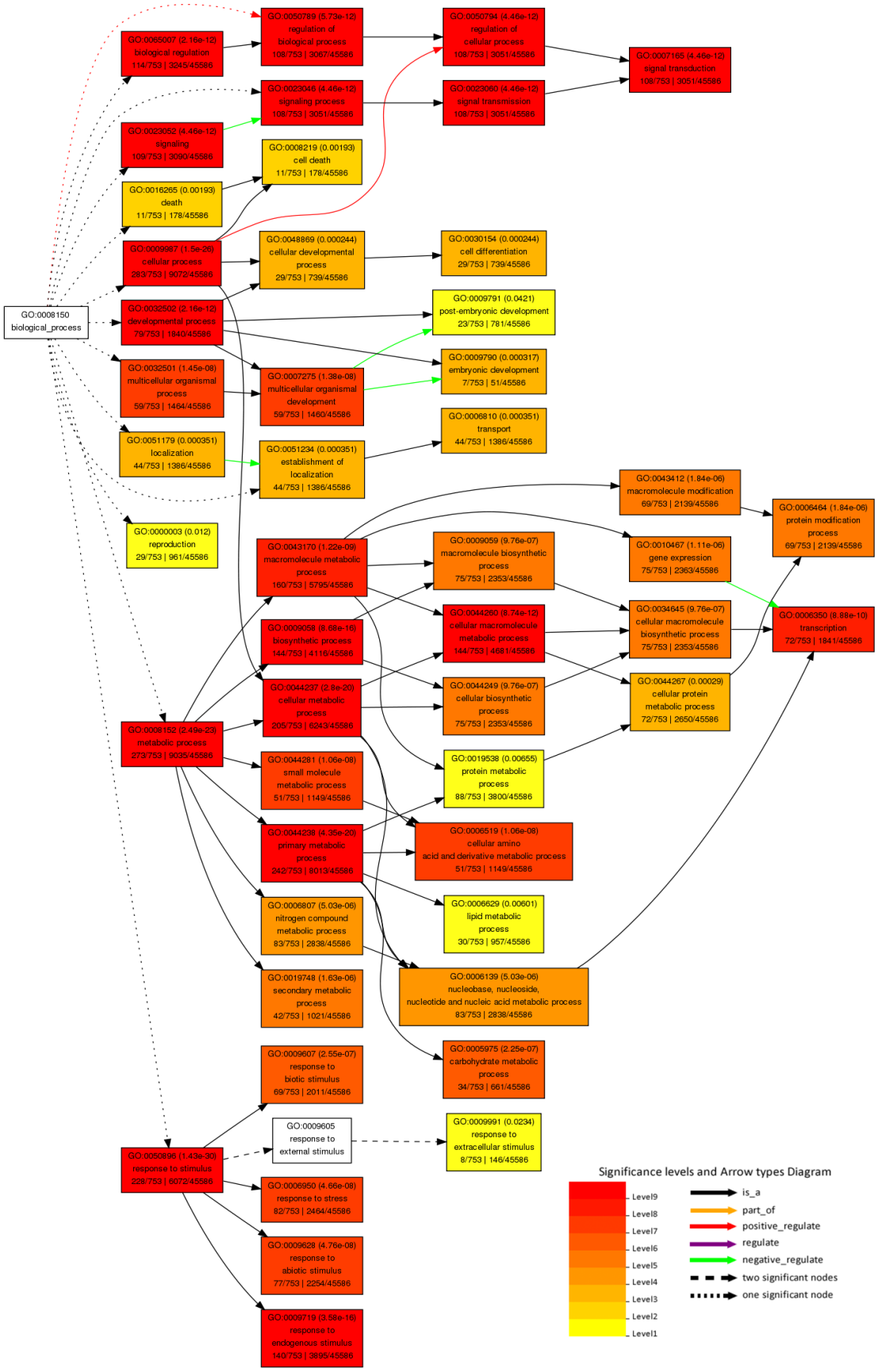

Figure 4-3. Biological process GO enrichment of selected DE genes among 4266LT-uf, MR-uf and MLT-uf, at stage before fertilization.

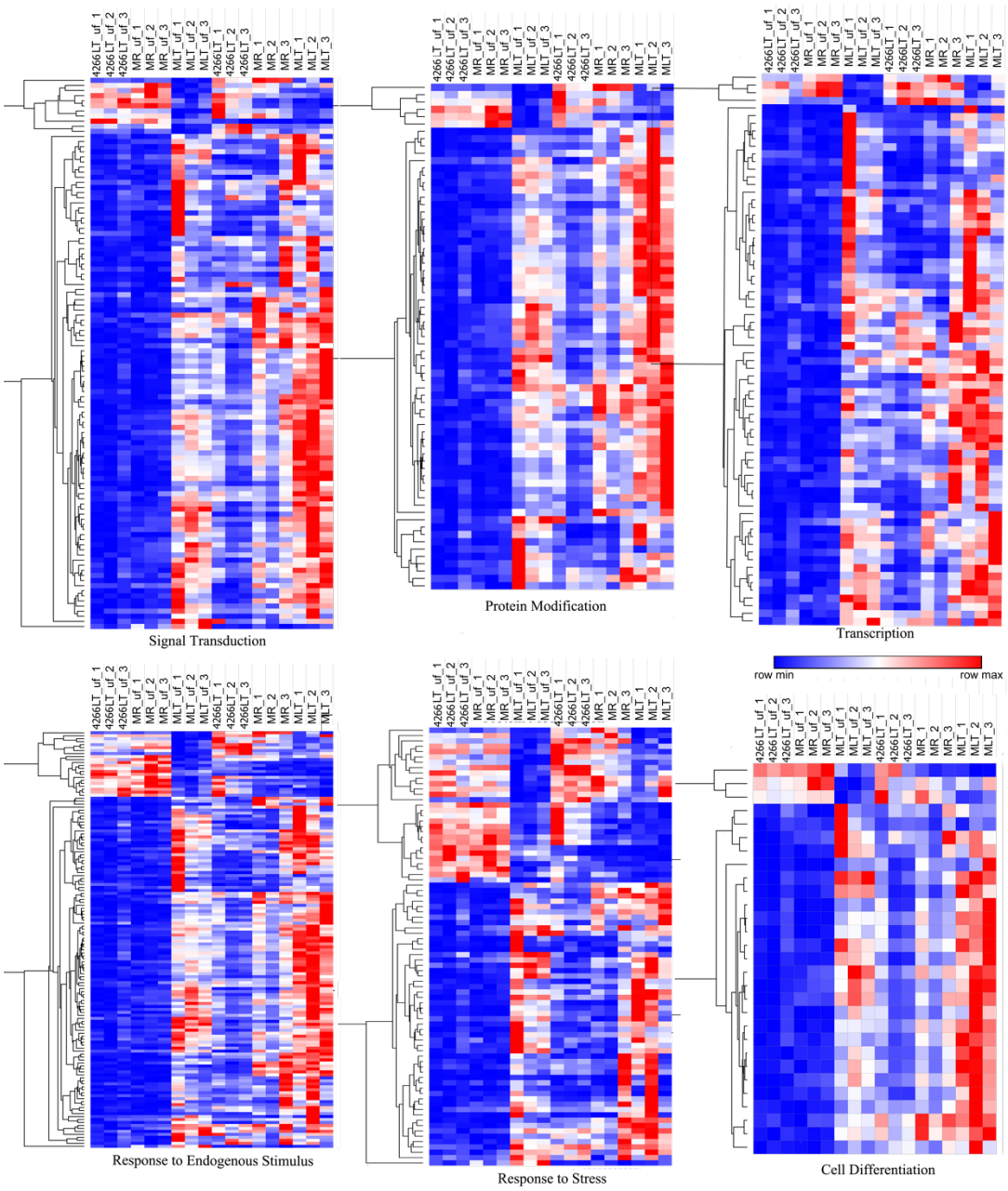

Figure 4-4. Heat map of selected GO terms' gene expression level before fertilization.

447 genes have different expression level after fertilization were selected for GO enrichment. Enrichment of biological process analysis revealed that those genes are significantly enriched in signal transduction, nucleic acid-metabolic process and protein modification process, which is similar with that before fertilization (Fig.4-5). Nucleic acid metabolism is the process that nucleic acid synthesized and degradation, which is a fundamental cellular process in plants and play an important role in PCD. PCD happens during the whole life cycle of plants, especially in plant reproductive progress, such as the disappearance of tapetal layer during pollen development and

degradation of synergid cells during double fertilization. In *tfs1*, the PCD of synergid cell is failed during fertilization, consistent with the results of GO enrichment,

The expression difference of these enriched GO terms showed with heat maps (Fig.4-6). Signal transduction, protein modification transcription and response to endogenous stimulus, some genes have higher expression level in 4266 and MR before fertilization, whereas some genes have higher expression level in 4266 and MR after fertilization. Only several genes enriched in GO term response to abiotic stimulus and nucleic acid-metabolic process has higher expression level in 4266 and MR.

Similar to stage before fertilization, many of these genes are transcription factors, and enzyme. Beside of those genes we mentioned above, there are also many other kinds of TF. MFS (Major facilitator superfamily protein) functions in the transportation of drugs, metabolites, oligosaccharides, amino acids and oxyanions, MFS protein energetically drive transport utilizing the electrochemical gradient of the target substrate (uniporter), or act as a cotransporter where transport is coupled to the movement of a second substrate. A histone gene which has lower expression in MLT before and after fertilization, whereas an osmotin has higher expression in MLT before and after fertilization, besides, a sugar transporter gene has dramatically high expression level in MLT.

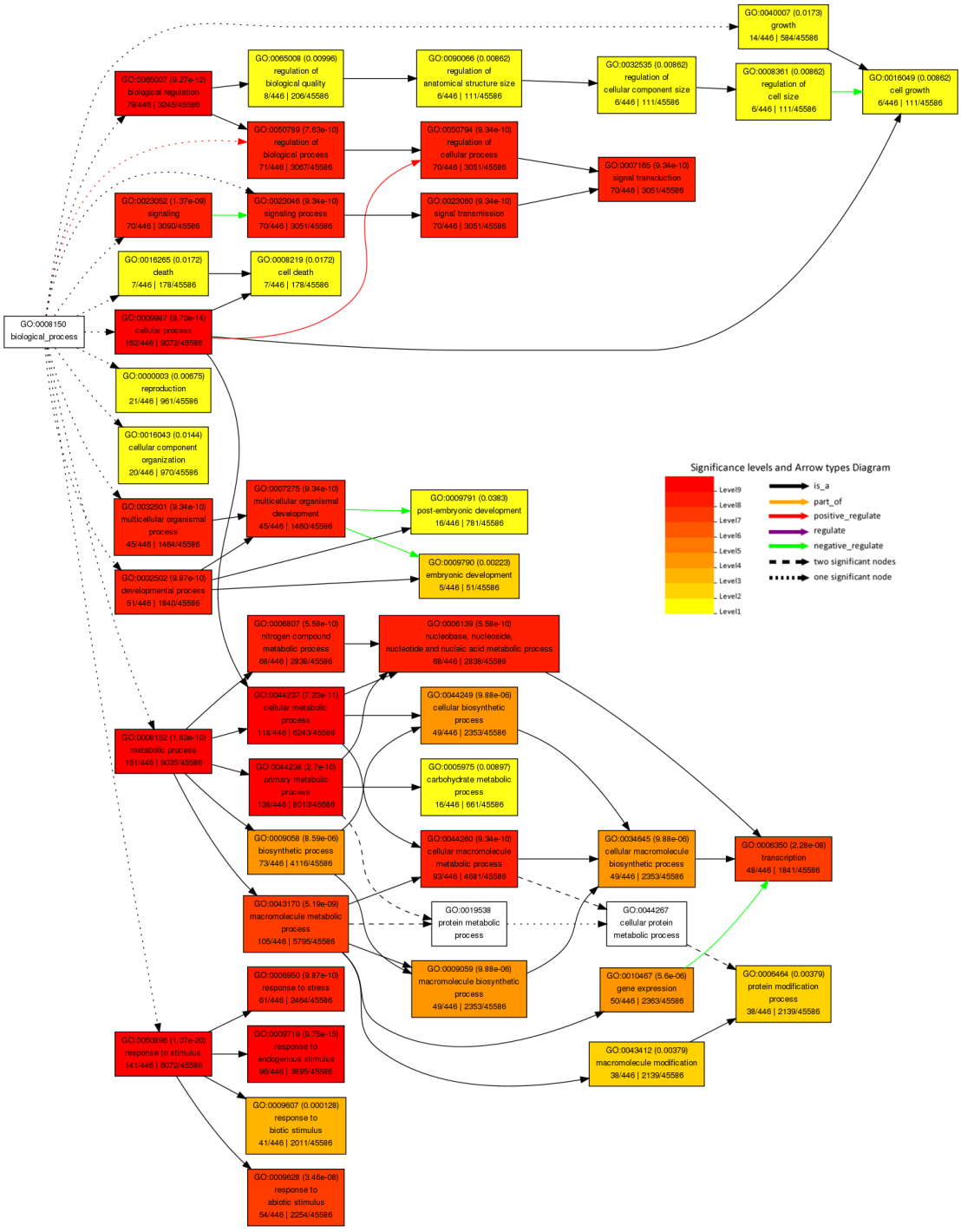

Figure 4-5. Biological process GO enrichment of selected DE genes among 4266LT, MR and MLT, at stage after fertilization.

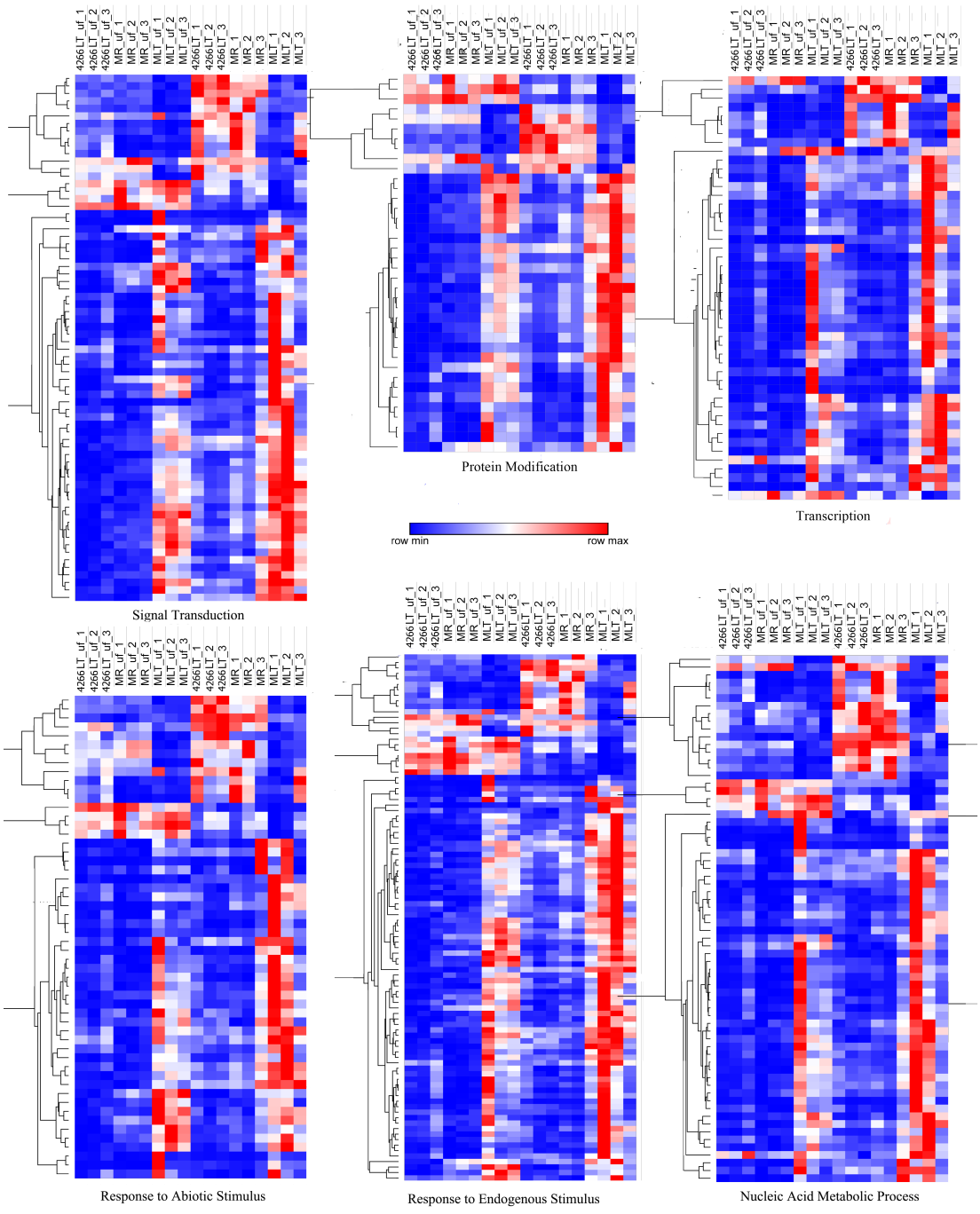

Figure 4-6. Heat map of selected GO terms' genes expression level after fertilization.

Beside of DE analysis at different stages, we selected genes have different expression level at both stages, and the number is 203. GO enrichment revealed that those genes are significantly enriched in response to stimulus, especially response to endogenous stimulus (Fig.4-7). Among those genes enriched in this GO term, we found some of them are related to ABA pathway and some are related to ethylene pathway, which

have been reported that play important roles in PCD regulation during endosperm development in maize.

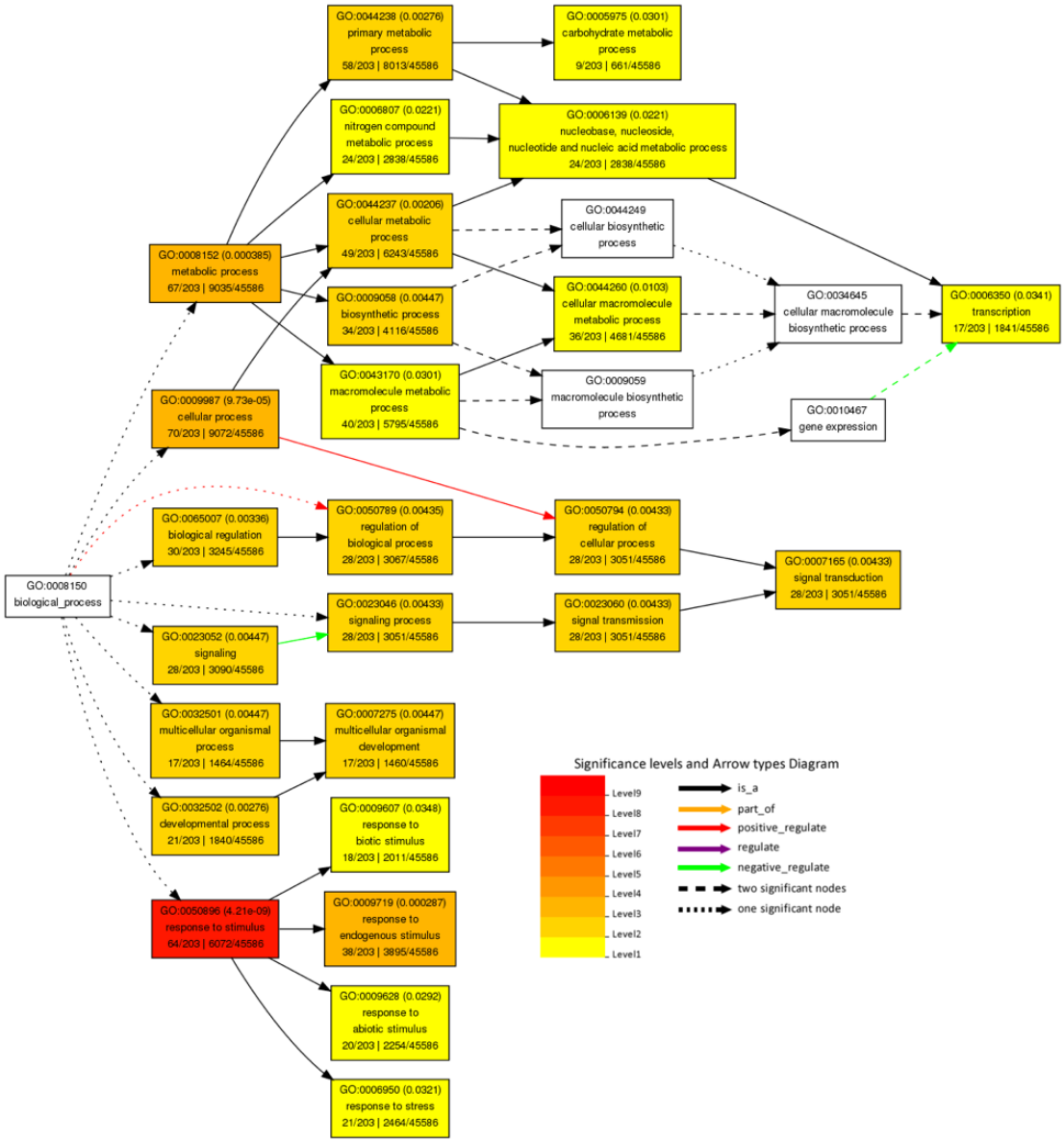

Figure 4-7. Biological process GO enrichment of selected DE genes at both stages among 4266LT, MR and MLT.

4.3.2 Identification of Novel miRNAs and siRNA phased loci

For miRNA annotation, after mapping, more than 88% of the clean reads can be mapped (Table 4-1), while those unknown tags were used for novel sRNA prediction. Among all the *MIRNA* genes for more than 70 species in miRBase database, 713 mature miRNAs derived from 592 precursors have been registered for rice (*japonica*, MSU7). We used RIPmiR for miRNA prediction as described (Breakfield et al., 2012). Since

our sample is the progeny of *indica* and *japonica*, we use both genome of *indica* (ASM465v1) (data not shown) and *japonica* (MSU7) as reference genomes, and the numbers of detected sRNA for each sample shown in Table 4-1. Statistics show that 972 miRNAs include 339 novel miRNAs and 633 registered miRNAs were identified.

Table 4-1. Summary of sequencing data that alignment statistics of tags to reference genome and detected sRNA for each sample.

Sample name	Raw tag count	Clean tag count	Percentage(%)	Clean tag count	Mapped tag	Percentage(%)	Known miRNA count	Novel miRNA count
MLT-1	108628420	100855178	92.84	100855178	94948203	94.14	506	334
MLT-2	85732815	78449490	91.5	78449490	73002584	93.06	488	329
MLT-3	78811315	66607113	84.51	66607113	61822030	92.82	484	328
MLT-uf-1	93188688	86141576	92.44	86141576	80737209	93.73	516	335
MLT-uf-2	76480628	70843040	92.63	70843040	62415261	88.1	499	334
MLT-uf-3	91872417	84651518	92.14	84651518	79900958	94.39	512	336
MR-1	67649592	62590982	92.52	62590982	58521362	93.5	509	323
MR-2	93039538	83006946	89.22	83006946	78356592	94.4	501	332
MR-3	72057979	67548108	93.74	67548108	63393019	93.85	491	327
MR-uf-1	56907798	52462445	92.19	52462445	49064825	93.52	489	333
MR-uf-2	73877868	65548407	88.73	65548407	61609683	93.99	490	333
MR-uf-3	85777952	79055842	92.16	79055842	73961929	93.56	510	334
4266 LT-1	82807620	75585073	91.28	75585073	70668478	93.5	503	333
4266 LT-2	76208344	69754894	91.53	69754894	65869397	94.43	497	328
4266 LT-3	55902694	52255568	93.48	52255568	49346551	94.43	474	327
4266 LT-uf-1	63290993	56931041	89.95	56931041	53134472	93.33	500	324
4266 LT-uf-2	79967832	72775756	91.01	72775756	68134740	93.62	501	332
4266 LT-uf-3	55249901	50265712	90.98	50265712	47153645	93.81	498	329

Small interfering RNAs can regulate gene expression by PTGS which is common in higher plants. Previous study revealed that siRNAs generation is triggered by certain AGO-loaded miRNAs, followed by synthesis of dsRNA and finally be processed to 21-nt siRNA. And the loci which can produce siRNAs are PHAS loci. PHAS loci are widely spread on the whole genome, including the intergenic region and intragenic region. All the sRNA reads of 18 libraries were mapped to rice genome for Phase loci prediction, 21-nt siRNA phased loci were predicted with a threshold P value ≤ 0.001 as previous method and a total of 3554 PHAS loci (supplementary table 4-3) were predicted. Including 627 (17.9%) overlapped intragenic region (supplementary table 4-4) and 2927 locate at intergenic region. Then we integrated the miRNAs and PARE data to identify the miRNA triggers of the PHAS loci. This identified 27 miRNA triggers of 344 PHAS loci (Supplemental table 4-5), 14 PHAS loci triggers by miRNA are overlapped annotated protein-coding genes (Fig. 4-8, Supplemental table 4-6).

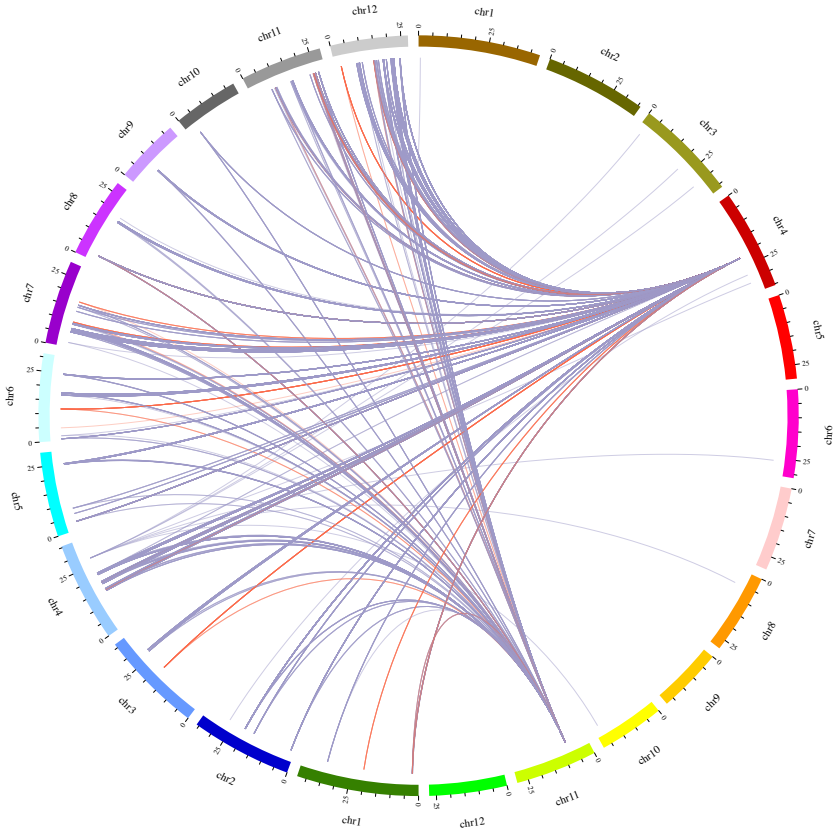

Figure 4-8. Position of 21-nt siRNA Phase loci and triggered miRNA.

The left side is the position of 344 21-nt siRNA Phase loci on rice 12 chromosomes. The right side is the position of 27 trigger miRNA Phase loci on rice 12 chromosome.

4.3.3 Differentially abundant miRNAs and target genes

The abundance of miRNAs were also compared in two group separately, the same with RNA-seq analysis talked above. Group 1 included 4266LT-uf, MR-uf and MLT-uf, the expression level was similar between 4266LT-uf and MR-uf, but has significant different with MLT-uf or the expression level in MR-uf that was between 4266LT-uf and MLT-uf were selected. Which means the miRNAs related to the mutative phenotype is existed before fertilization, such as the defective hull and nonfunctional female gametophyte. Group 2 included 4266LT, MR and MLT, same with group 1, the expression level was similar in 4266LT and MR, whereas had significant difference with MLT or the expression level in MR that was between 4266LT and MLT were selected. These results indicated the miRNAs related to the mutative phenotype was

triggered by fertilization or express after fertilization, such as the degradation of synergids or the PCD of other cells which were related to double fertilization.

After statistic of two groups comparison result, we selected 77 which has similar expression in 4266LT and MR, whereas has different expression level in MLT, including 28 novel miRNAs and 49 registered miRNAs (supplementary table 4-7). To better understand the development processes they may involve in, we use a heat map (https://software.broadinstitute.org/morpheus/) to show the expression difference of these miRNAs (Fig.4-9). For some miRNAs, their expression level was slightly higher in MLT-uf than 4266LT-uf and MR-uf, or shown no significant different, whereas their expression level in MLT was much higher than in 4266LT and MR. miR159 in previous study was predicted to target MYB and TCP transcription factors. miR160 was predicted to target ARF (Jones-Rhoades and Bartel, 2004) and miR166 was target HD-ZIP (Douglas et al., 2010). osa-miR530 was predicted to target Plus-3 domain protein, 2,3-diketo-5-metylthio-1-phosphopentane phosphatase and transposable element protein (Lu et al., 2008b). Expression of osa-miR444d in MLT-uf was much lower than that in 4266LT-uf and MR-uf, but in MLT it was much higher than that in 4266LT and MR. osa-miR444d was predicted to target ANR1-like MADS box protein, which is also an important TF that regulate the development of pistil (Lu et al., 2008a; Morin et al., 2008). osa-miR164a, osa-miR156i, osa-miR1863c had an opposite expression trend with osa-miR444d. previous study reported that the balance between the miR164A and *CUC2* genes controls leaf margin serration in Arabidopsis (Nikovics et al., 2006), miR156 and 168 were predominantly associated with AGO1a and AGO1b, not with AGO1c; miR160 and 167 were mainly recruited by AGO1c, not by AGO1b (Wu et al., 2009), miR167 have similar expression level before fertilization, but in MLT is much lower. For another pattern of miRNAs, their expression levels were much higher in MLT-uf, but had no significant difference in three samples after fertilization. Researches in Arabidopsis revealed that the levels of miR168 and AGO1 may be fine-tuned by an autoregulatory loop (Vaucheret et al., 2006) while miR399 were predicted to target phosphate transporter. miR162 has a bulged nucleotide as it base pairs to the mRNA of DCL1. The target of miR1861 belongs to structural maintenance of chromosomes (Zhu et al., 2008). More miRNAs predicted targets were analyzed by GO enrichment.

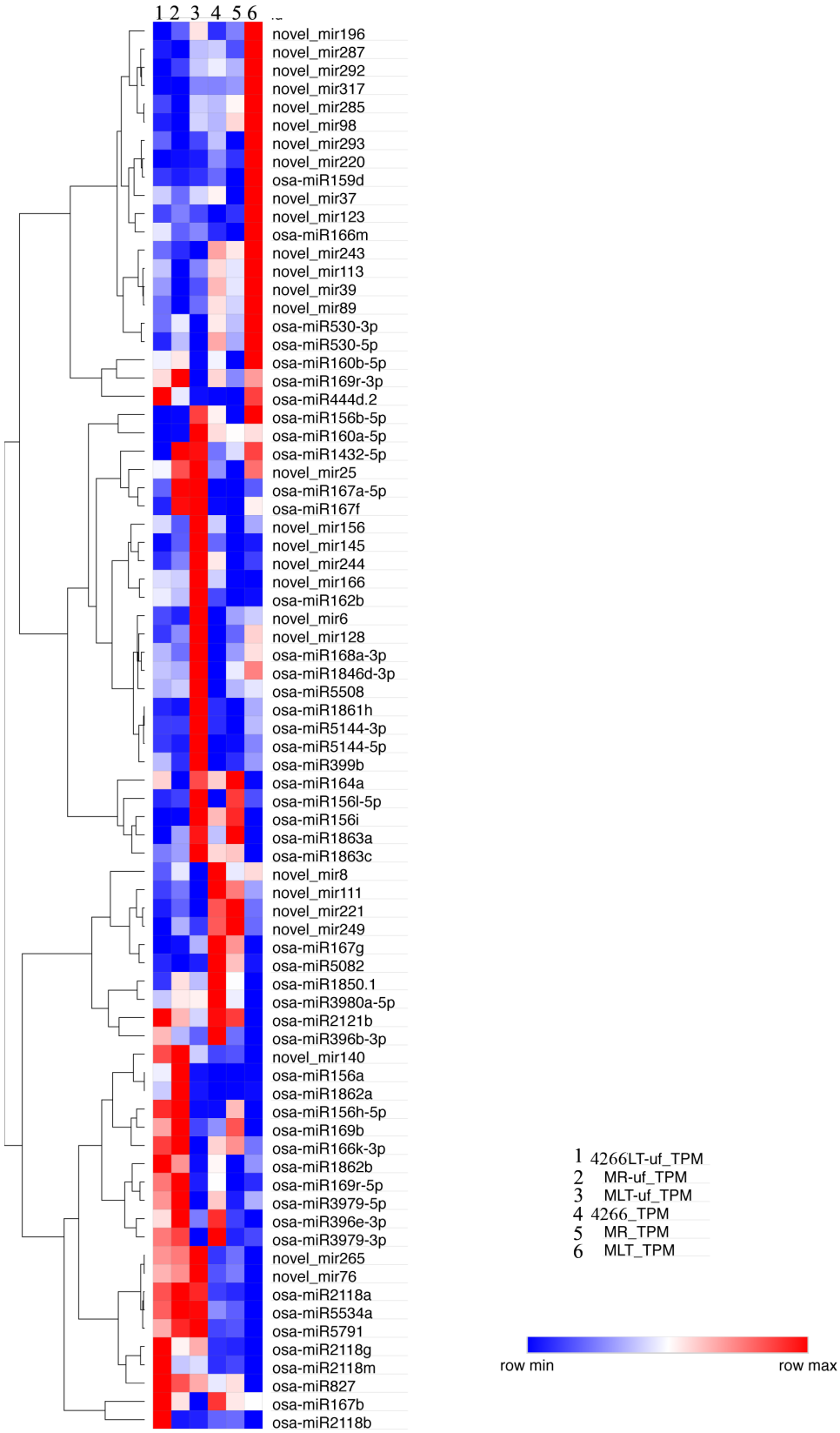

Figure 4-9. A heat map showing miRNA expression in different samples.

For more detailed comparison, we did cluster analysis of their expression level showed by TPM (supplementary table 4-7). To show the tendency of different miRNAs among

the 6 samples, we use a scaled log transformed TPM to show the value of TPM at relatively uniform value, thus compare the difference among 6 samples (supplementary table 4-8). The formula is shown as following:

$$\log_2(\mathrm{TPM})=X$$

$$\text{Scaled } \log_2 (\mathrm{TPM}) = \frac{X - \bar{X}}{\sigma}$$

There are 11 clusters of miRNA, and some miRNAs can't be clustered with others, but their expression ranges among 6 samples are interesting, which are involved in GO enrichment analysis. In cluster A, D, G, K, the expression level between 4266LT-uf and MR-uf has no significant difference, whereas in MLT-uf is much higher. In cluster B, C, E and H, the expression level between 4266LT-uf and MR-uf has no significany, whereas in MLT-uf is much lower (Fig. 4-10).

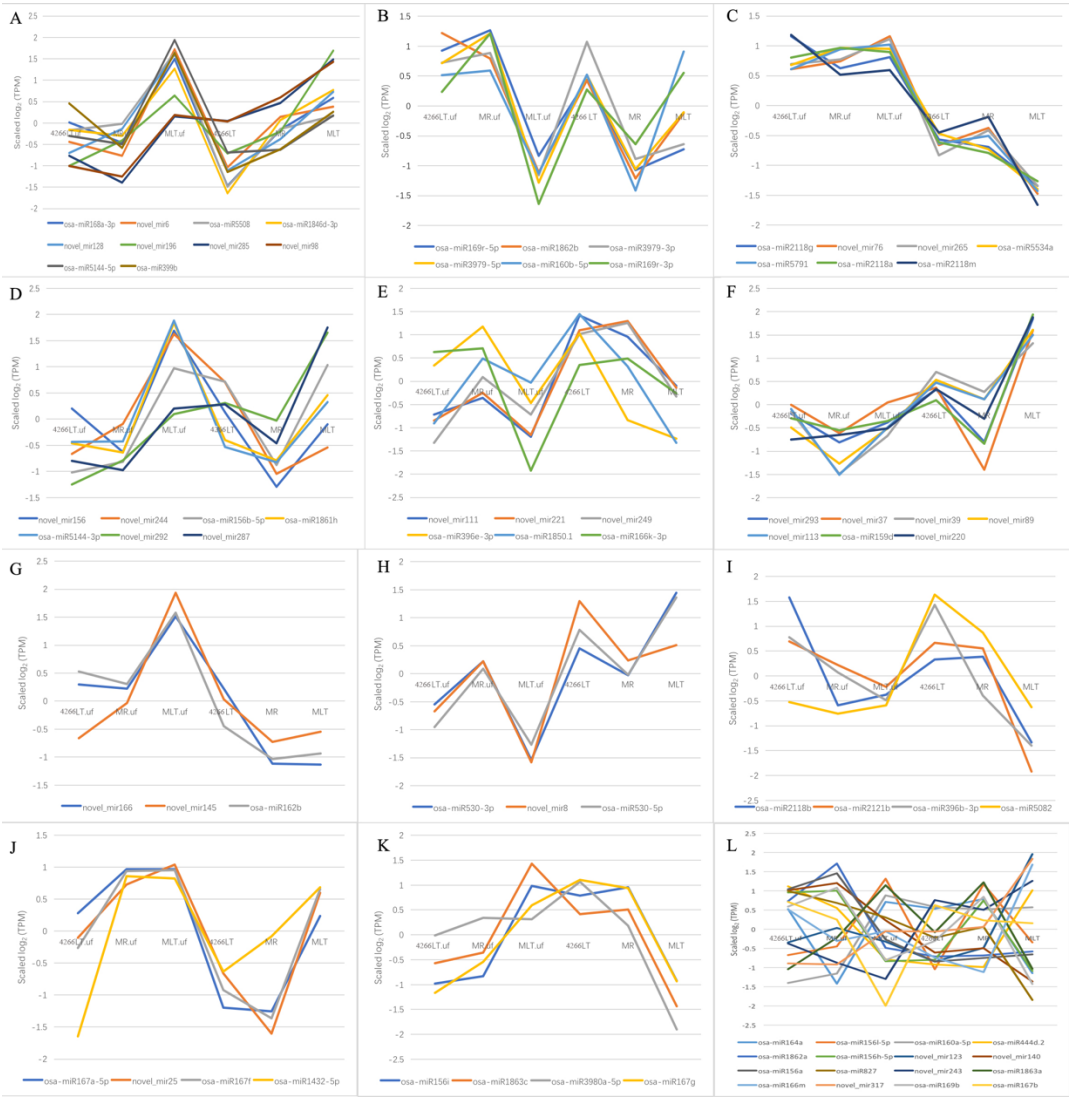

Figure 4-10. Cluster analysis of 77 miRNAs expression pattern among 6 samples.

Prediction of miRNA targets is a critical step in identifying miRNA: mRNA target interactions for experimental validation. Researches revealad that miRNAs function via base-pairing with complementary sequences within mRNA molecules. We use PARE (parallel analysis of RNA end) data to verify miRNA-directed target cleavage. For the analysis of using *japonica* (MSU7) as reference genomes, 21 miRNAs found 65 targets (supplementary data Table 4-9), which participated in various biological process and were mainly divided into two branches, one participated in reproductive process and finally regulate flower development, the other branch included many processes such as nitrogen compound metabolic, macromolecule biosynthetic and gene expression, thus control the transcription (Fig. 4-11). The molecular function analysis also shows those genes function is related to transcription factor activity, which is consist with the former report by many researchers, Argonaute can regulate the expression or translation of some transcription factors through small RNAs directed pathway.

The list of the genes of GO:0009908 and GO: 0006350 are shown in table 4-2, the annotation of those 33 genes from ensemble (http://plants.ensembl.org). There are 7 ARFs, 7 SPBs, 2 MYB genes, 3 NAC-domain containing proteins, 4 nuclear factors, 2 AGL16s (Agamous-like MADS-box protein AGL16) and 1 TCP2 gene. All those genes are famous transcription factors and have been reported to be related to plant reproductive development. As mentioned above, ARF, Myb, MADs-box and NAC domain protein are all involved in developmental processes and are also related to the phenotype of our mutant. Nuclear factor is a family of closely related transcription factors which constitutively bind as dimers to specific sequences of DNA with high affinity. Previous study shows that TCP2 plays a pivotal role in the control of morphogenesis of shoot organs by negatively regulating the expression of boundary-specific genes such as CUC genes, probably through the induction of miRNA.

The analyses on the differentially expressed genes and miRNAs are independently, while the target genes of selected miRNAs were consistent with the DEGs found from transcriptome analysis. This result is consistent with previous study and further suggested that the defective phenotype of *tfs1* was controlled by a complicated network, including both miRNA and genes. We found their expression level is not obvious that the genes belong to the same family have similar pattern, which may illustrate that the function of different family members are various.

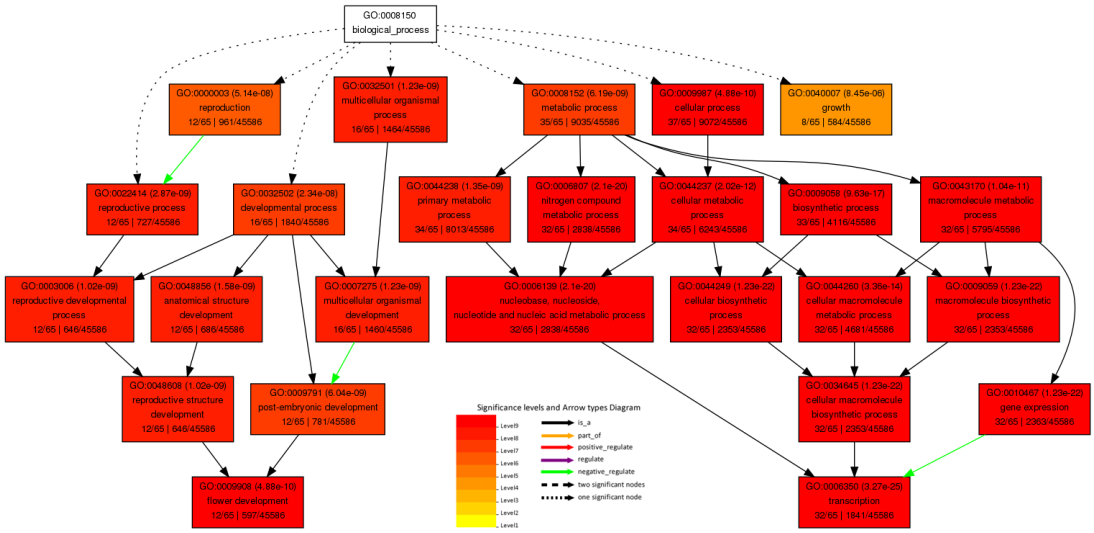

Figure 4-11. GO enrichment analysis of 65 target genes of 21 DE miRNAs.

Table 4-2. The list of the target genes enriched in GO term flower development and transcription, and their FPKM value.

			FKPM_mean					
Loci	arabi-symbol	arabi-defline	4266 LT_uf_average	MR_uf_average	MLT_uf_averge	4266 LT_average	MR_average	MLT_averge
LOC_Os01g59660	ATMYB65,MYB65	myb domain protein 65	18.94	21.60	21.81	11.32	12.02	8.69
LOC_Os01g69830	SPL	SBP domain protein-like transcription factor family protein	16.94	18.11	21.18	31.05	22.93	30.46
LOC_Os02g04680	SPL2	squamosa promoter binding protein-like 2	13.42	16.53	7.45	20.93	18.77	14.09
LOC_Os02g06910	ARF6	auxin response factor 6	30.09	29.04	35.93	33.26	28.19	30.22
LOC_Os02g07780	SPL2	squamosa promoter binding protein-like 2	7.19	9.78	6.34	4.59	7.68	7.04
LOC_Os02g36924	AGL16	AGAMOUS-like 16	8.18	8.68	5.31	1.88	1.58	1.05
LOC_Os02g41800	ARF16	auxin response factor 16	19.91	23.63	20.18	26.19	24.12	23.28
LOC_Os03g07880	ATHAP2C,HAP2C,NF-YA3	nuclear factor Y, subunit A3	5.75	4.50	6.49	6.17	20.29	11.48
LOC_Os03g29760	ATHAP2C,HAP2C,NF-YA3	nuclear factor Y, subunit A3	6.72	5.33	5.50	6.79	9.66	8.39
LOC_Os03g48970	NF-YA9	nuclear factor Y, subunit A9	8.01	5.93	12.74	8.36	16.26	22.07
LOC_Os04g57610	ARF6	auxin response factor 6	46.42	42.21	39.20	44.09	35.29	36.39
LOC_Os06g40330	ATMYB65,MYB65	myb domain protein 65	5.61	6.30	6.22	6.59	10.44	5.89
LOC_Os06g45310	SPL11	squamosa promoter-like 11	5.81	4.94	5.98	8.02	8.26	8.74
LOC_Os06g46270	anac021,ANAC022,NAC1	NAC domain containing protein 1	10.38	7.46	8.61	39.85	14.56	15.14
LOC_Os06g46410	ARF8,ATARF8	auxin response factor 8	21.67	23.51	21.01	22.51	25.87	26.48
LOC_Os06g47150	ARF16	auxin response factor 16	18.96	15.32	18.83	21.29	19.38	20.04
LOC_Os06g49010	SPL2	squamosa promoter binding protein-like 2	17.37	20.48	15.95	29.08	25.43	27.60
LOC_Os07g05720	TCP2	TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 2	16.34	22.73	20.14	20.12	18.07	20.84
LOC_Os07g41720	ATHAP2C,HAP2C,NF-YA3	nuclear factor Y, subunit A3	2.66	1.98	2.01	2.82	5.44	2.97
LOC_Os08g10080	anac021,ANAC022,NAC1	NAC domain containing protein 1	3.75	4.43	3.29	0.76	0.82	0.41
LOC_Os08g39890	SPL9	squamosa promoter binding protein-like 9	2.12	2.09	2.09	2.79	3.02	4.90
LOC_Os10g33940	ARF16	auxin response factor 16	13.18	12.27	11.47	17.02	18.78	13.75
LOC_Os11g30370	SPL	SBP domain protein-like transcription factor family protein	1.59	2.18	1.65	1.86	3.47	3.23
LOC_Os12g41680	anac021,ANAC022,NAC1	NAC domain containing protein 1	12.78	16.14	16.80	13.02	13.99	17.67
LOC_Os12g41950	ARF8,ATARF8	auxin response factor 8	17.84	18.86	16.91	25.85	24.12	20.40
LOC_Os12g42400	NF-YA10	nuclear factor Y, subunit A10	1.38	1.17	0.73	2.09	1.90	1.10

4.3.4 Correlation between miRNA, 21-nt siRNA and target genes

Since transcriptome analysis and miRNA prediction showed coincident results, we are interested in the genes expressions that have correlation with small RNA. Previous study reveals that miRNAs affecting the expression of target genes via translational inhibition or post-transcriptional gene silencing. And the target genes usually down regulated by certain miRNA (Wang and Li, 2009). In plant, miRNA bind on targets genes rely on high extent sequence complementation, among the DE miRNAs and RNAs, we found 4 miRNAs and 5 genes have negative correlation (Table 4-3). Gene SPL2 and SPL9 both target by miR156i, they relate to transcription and flower

development. Genes contain the NB-ARC domain usually relate to disease resistant, while in animals they are relate to cell death.

Table 4-3. The list of the DE genes and DE miRNAs that have negative correlation

miRname	Target	symbol	Annotation	Biological process	Molecular function	Cellular component
osa-miR156i	LOC_Os02g04680	SPL2	Squamosa promoter binding protein-like 2	Transcription, flower development	Transcription factor activity	Nucleus
osa-miR156i	LOC_Os08g39890	SPL9	Squamosa promoter binding protein-like 9			
osa-miR167b	LOC_Os07g29820		NB-ARC domain-containing disease resistance protein	Response to biotic stimulus, Regulators of cell death in animals	Transcription factor activity	Mitochondrion
osa-miR169b	LOC_Os03g48970	NF-YA9	Nuclear factor Y, subunit A9	Transcription	Transcription factor activity	Nucleoplasm
osa-miR528-5p	LOC_Os06g37150		Plant L-ascorbate oxidase	/	Catalytic activity	Membrane

miRNA loading AGO also have function in siRNA generation, and the generated siRNA also can load AGO and regulate genes expression. And the siRNA usually has positive correlation with trigger miRNA. To show the correlation of different miRNAs, triggered siRNAs and target genes among the 6 samples, we use a scaled log transformed TPM and FPKM respectively to show the value at relatively uniform value, thus compare the difference among 6 samples, the formula is the same with above. We found 3 pairs of correlation. We use line chart to present the expression level of target genes and abundance of these correlated miRNAs and siRNAs, which clearly shows the correlations by different color, it is clearly shows that the siRNA has positive correlation with trigger miRNA whereas have negative correlation with target genes (Fig.4-12). The annotation of these genes presented in table 4-4, one is CTP synthesis family protein, which is relate to embryo development. One is TPR super family protein which is related to response to endogenous stimulus. And one is NADH-dependent glutamate synthesis which is relate to plant growth. The function of these three genes is consistent with transcriptome analysis and miRNA target DE genes. Further confirmed that the defective phenotype of our mutant is related to small RNA regulation.

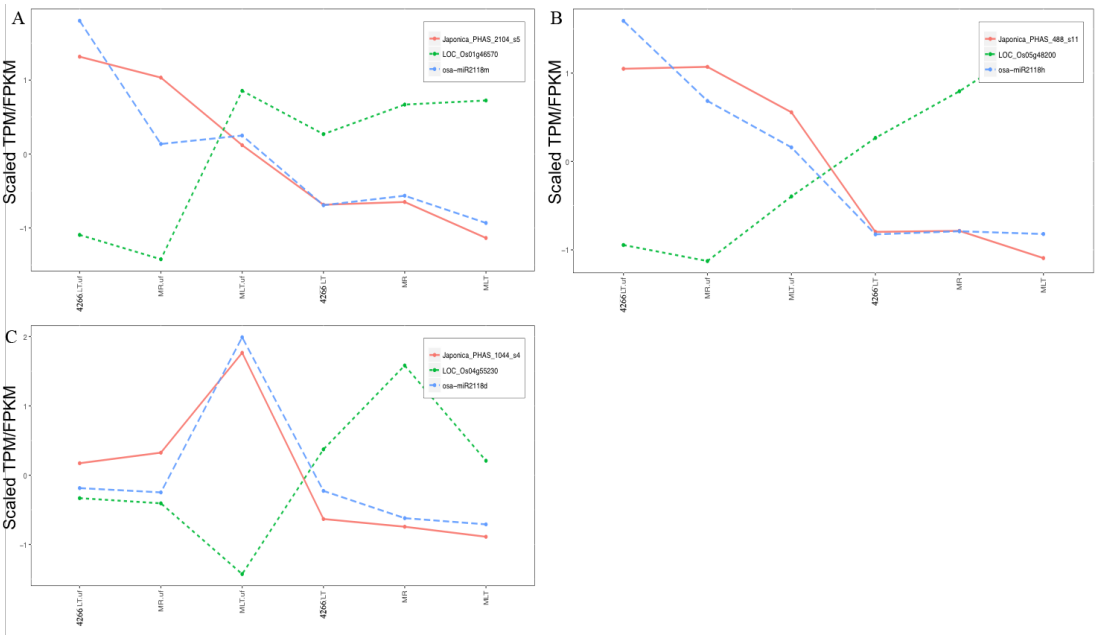

Figure 4-12. Correlation between miRNA and genes

(A) Osa-miR2118m (blue dash line) can trigger the generation of PHAS_2104_s5 (red line), and they have positive correlation, whereas have negative correlation with target gene LOC_Os01g46570(green dotted line). (B) Osa-miR2118d (blue dash line) can trigger the generation of PHAS_1044_s4 (red line), and they have positive correlation, whereas have negative correlation with target gene LOC_Os04g55230(green dotted line). (C) Osa-miR2118h (blue dash line) can trigger the generation of JPHAS_488_s11 (red line), and they have positive correlation, whereas have negative correlation with target gene LOC_Os05g48200(green dotted line).

Table 4-4. The list of the DE miRNAs and triggered siRNA that have positive correlation, and siRNAs' target genes that have negative correlation with certain siRNA.

Trigger miRNA	siRNA ID	Target	Homologous gene	Annotation	Biological process	Molecular function	Cellular component
osa-miR2118m	PHAS_2104_s5	LOC_Os01g46570	emb2742	CTP synthase family protein	embryo development,post-embryonic development,reproduction	catalytic activity	cytosol
osa-miR2118d	PHAS_1044_s4	LOC_Os04g55230		Tetratricopeptide repeat (TPR)-like superfamily protein	response to endogenous stimulus	binding	/
osa-miR2118h	JPHAS_488_s11	LOC_Os05g48200	GLT1	NADH-dependent glutamate synthase 1	cellular process,growth	catalytic activity	plastid

4.4 Discussion

In decades, the structure of Argonaute has been reported in many species, which is mainly composed by four function domains and two linkers. Researches based on mutants revealed that any mutations in these four functional domains have severe phenotypes. However, to date there is no crystal structure of plant Argonaute proteins, since the structure of Argonaute is conserved, we can find some clues from the reported

Argonaute models. But the function is unique for each Argonaute, they can load different kinds of guide RNA and pair with different target strands, which means special mechanism is remained unclear. Researchers usually focus on the function of domains when investigate the function of a protein, our mutant, which has a single SNP mutation, may show a new clue that the intermediate zone is also as important as domains.

TFS1 is an Argonaute, previous study revealed that miRNA is always involved in Argonaute regulation and the target genes are TFs which have comprehensive regulation of plant development. We have performed the small RNA sequencing to study the mutant gene's effect on the small RNA. Screening conditions were applied the same with transcriptome analysis, and many miRNAs have predicted function relative to TFs has been found. miR159 in previous study was predicted to target MYB and TCP transcription factors. miR160 was predicted to target ARF (Jones-Rhoades and Bartel, 2004). miR166 was predicted to target HD-ZIP(Douglas et al., 2010). osa-miR530 was predicted can target Plus-3 domain protein, 2,3-diketo-5-metylthio-1-phosphopentane phosphatase and transposable element protein (Lu et al., 2008b). osa-miR444d was predicted to target ANR1-like MADS box protein, which is also an important TF that regulate the development of pistil (Lu et al., 2008a; Morin et al., 2008). Previous study reported that the balance between the miR164A and *CUC2* genes controls leaf margin serration in Arabidopsis (Nikovics et al., 2006), miR156 and 168 were predominantly associated with AGO1a and AGO1b, not with AGO1c, while miR160 and 167 were mainly recruited by AGO1c, not by AGO1b (Wu et al., 2009). Researches in Arabidopsis revealed that the levels of miR168 and AGO1 may be fine-tuned by an autoregulatory loop (Vaucheret et al., 2006) while miR399 may target phosphate transporter. miR162 has a bulged nucleotide as it basepairs to the mRNA of DCL1. The target of miR1861 belongs to structural maintenance of chromosomes (Zhu et al., 2008).

When *tfs1* mutant is grown under low temperature, part of the floret is restored from the mutant phenotype, which was sampled for RNA-seq analysis. As shown in the RNA-seq results, the DEGs numbers from group of 4266LT-uf vs MR-uf and 4266LT vs MR is significantly lower than that from the other two groups, respectively (Fig. 4-2), especially in plants before fertilization. MR means those florets having same phenotype with 4266 plant, indicating gene expression pattern in MR and 4266 should

be much more similar. Thus our result in RNA-seq is reliable and the following analysis will be of great help in searching the target gene for *TFS1*. To study which biological processes of these DEGs are involved, we have performed GO analysis. Our result revealed that the embryonic development process is highly GO enriched, which is closely relative to the phenotype of *tfs1* mutant. Such as the MADS-box family, MYB, HD-Zip family and ARF family, which are transcript factors and play important roles in regulating this process. Most of these gene families have significant expression difference between these plants. This result is consistent with previous studies showing that these TFs were regulated by AGO protein.

We also have performed the small RNA sequencing to study the mutant gene's effect on the sRNA. Consistent with RNA-seq result, the miRNA profiles in 4266 and MR florets are similar but different with that from MLT both in unfertilized and fertilized floret from these plants. Furthermore, the target gene for these DE miRNAs again points to the TFs that are evolved in flower development and transcription (Fig. 4-11), suggesting the mutant gene playing a key role in this process. To further prove this, a correlative analysis between results from RNA-seq and miRNA-seq has been done independently to discover the relationship between miRNA and gene expression in *tfs1* mutant. And we found some genes expression level have perfect negative correlation with the abundance of miRNAs. One of them is a gene contain the NB-ARC domain, in plant it usually relates to disease resistant, while in animals they are relate to cell death. The gene is target by miR167b, which is considered similar to miR160, can target ARFs. Beside of miRNA and directly targets, we also tried to find siRNAs have positive correlation with trigger miRNA but negative correlation with target genes. The result points to miR2118, and the function of target genes also consistent with transcriptome analysis and miRNA target DE genes. Our sequencing results further suggesting that the *TFS1* gene playing a key role that cause the defective phenotype of our mutant, and miRNA regulation is involved in this process. To further explore the molecular function of *TFS1*, more experiments by reverse genetic techniques will be carried out in the future.

Chapter 5 General discussion and conclusion

Flower development is a fundamental biological process in plant, which is composed of pistil development and ovule development. With its discovery of rice male sterile mutant lines and owing to its great application potential in hybrid rice, the studies on pollen development has become a hot topic for decades in plant reproductive research. Numerous publications on pollen development, pollen germination and growth as well as pollen guidance have been reported worldwide. The identification of thermo- and photoperiod- sensitive male sterility gene, *PMS3* (Ding et al., 2012) or *P/TMS12-1*(Zhou et al., 2012), is a breakthrough in hybrid rice research, which firstly explain the mechanism that rice change its fertility in response to environmental cues, such as temperature and photoperiod (Zhu and Deng, 2012). However, in contrast to male sterility, female sterility in plant is seldom brought to light due to the low application possibility in hybrid reproduction, although it is frequently identified during field practice and in laboratory researches. Except the process of embryo sac development and the guidance role of embryo sac during double fertilization (Chen et al., 2007; Yadegari and Drews, 2004), it's still not clear known about the control mechanism of this process. Besides, thermo- and photo-sensitive male sterility (T/PMS) is more or less common in the world of plants, whereas thermo- and photo-sensitive female sterility is exceptionally rare, not to mention the cloning of a T/PFS gene. Thus, research on female sterility is far behind that on male sterility of rice (Cheung et al., 2013). In hybrid rice, the cost for F1 seed production, especially the labor demand, is relative high because of the inapplicable of mechanization, which has become a bottleneck for the development of hybrid rice research. As an idea material to solve this problem, the utility of a P/TFS recover line in hybrid seed production can significantly decrease the cost and produce high quality F1 seeds. Thus, the discovery of a photo/thermo-sensitive female sterility line is not only of great importance in ovule development research, but also has incredible application potential.

In the present study, we were very lucky to identify a female sterility mutant in the F9 segregation population 4266 of a multi-crossing event. Interestingly, the sterility line was totally infertile under normal conditions for rice growing but partial recovery of the fertility when grown under lower temperature (25-27℃), which suggests that this female sterility line might be a thermo-sensitive female sterility rice. The same

phenotype of seeds collected from female sterility line grown under lower temperature condition has proved our hypothesis about this mutant line. Furthermore, crossing with other normal rice and analysis of segregation pattern in F2 population revealed that the female sterility was controlled by a recessive nuclear gene which we named *TFS1*. Based on the studies on Arabidopsis, tobacco, petunia and rice, researchers have identified many genes that controlled the formation of placenta, ovule primordium and the differentiation of ovule apical-basal axis and dorsal-ventral axis, the initiation of integument. Their results revealed that these biological processes were controlled by a complicated signal network, and the MADS-box genes played an important role in this network (Nayar et al., 2014). Such as *OsMADS3*, antisense RNA expressing transgenic plants had phenotype that pistil changed to sterile floret, and stamen became pistil. Mutant of gene *OsMADS16*, also shows stamen changed to carpel. All these results indicated that a cytological study on this mutant will be helpful for better understanding the control mechanism of temperature affects fertility of *tfs1* mutant rice.

We have carried out different cytological techniques, including CLSM, paraffin section, as well as SEM to study the ovule development of *tfs1* mutant. Our preliminary results from cytological observation have shown that *tfs1* is totally sterile after pollinated by pollens from different cultivars, including the 4266 plants, although pollens from tfs1 can fertilize 4266 plant and produce normal seed. Pollen viability examinations have further support our crossing result that the male part of *tfs1* is normal. To find out the reason that results in female sterility, we used CLSM and paraffin section technologies to study the development of embryo sac in *tfs1*. However, our result turns out that embryo sac in *tfs1* is complete too, which also has a typical 7 celled and 8 nucleated structure (Fig. 2-8; Fig.2-9). Fertility of rice is not only controlled by male and female gametophyte developments, but also by the interactions between these two kinds of gametophytes (Higashiyama and Yang, 2017; Li et al., 2011; Wang et al., 2012a). In *tfs1*, the normal embryo sac implies that the interaction between embryo sac and pollen or the afterward process, embryo generation, are defective in *tfs1*. To further explore the sterility mechanism, the CLSM and DIC microscope technologies, which have been frequently used in studies of embryo sac development and embryo development, were used again to observe the development of embryo in *tfs1* after pollination. Obviously, the embryo development in *tfs1* is totally defective, the egg cell is still observed in

ovules of *tfs1* 3 days after pollination (Fig. 2-11), suggesting the female sterility of *tfs1* is because of the failure of interaction between embryo sac and pollen, the double fertilization process.

Double fertilization is an unique process in angiosperms, which is start form pollen germination on stigma, and end with the fusion of two sperms with egg and polar nuclei, respectively(Steer and Steer, 1989; Taylor and Hepler, 1997). The first step of double fertilization is pollen guidance, which starts from pollen germination on stigma and pollen tube extending through the style then to the micropyle. Many mutants have been used to illustrate this step. It has been well-documented that after entering the embryo sac, pollen tube will pass through the filiform apparatus and reach to one of the two synergid cells, resulting in the degeneration of this synergid cell. In maize, a small molecular protein, *ZmEA1* which express in filiform apparatus, decrease the expression of this gene may lead to pollen tube guidance failure. In Arabidopsis, a gene *MYB98*, a member of R2R3-MYB family, is a key regulator of transcriptional events in synergids. It also has been reported that an Arabidopsis mutant, named *feronia*, has problem of synergid cell degeneration after interacting with pollen tube, which makes the pollen tube continue growing in the embryo sac, thus lead to the failure of double fertilization. In the present study, we have pollinated the *tfs1* and 4266 pistils, aniline blue staining result demonstrated that both pollens from 4266 and *tfs1* could germinate and grow through the style tissue. It was also clear that pollen tube on both pistils can find the micropyle, indicating the growing of pollen tube is unblocked on *tfs1* compared with 4266 plant (Fig. 2-10). Our preliminary result seems to approve that pollen guidance in *tfs1* is normal, further experiments are still needed to explore cytological sterility mechanism of *tfs1*.However in rice, since the integument and somatic cell layers are much thicker than that in Arabidopsis, the transparent method is not suitable for pollen tube observation, which makes it more difficult in rice for double fertilization investigation. However, when we investigated the embryo development of *tfs1* and 4266 plant after pollination by CLSM and DIC micro technologies, we found that not only the egg cell was still found even at 5DAP, but also the synergid cells were observed in *tfs1*'s embryo sac at 3DAP. There was no embryo generate totally. On the contrary, embryo development in 4266 plant started from 1DAP to 5DAP when the first leaf primordium arrested (Fig. 2-11). The failure of synergid cell degeneration in *tfs1* pistil

suggested that pollen guidance in this step was abolished, which leads to the failure of sperm release and thus the double fertilization. We used SEM to study the development of floret, including the formation of lemma and palea. Our results revealed that the defective hull phenotype was determined at the best early stage of floret development, an abnormal growth of lemma and palea starts from the primordium stage (Fig. 2-12). These results suggest that the defective hull phenotypes, together with female sterility phenotype, are determined before the heading of panicle.

Map-based cloning is an effective forward genetic technique to identify genes underlying phenotypic variations. Map-based cloning which relies on the construction of segregation population and the molecular markers for linkage analysis has been widely applied to different species and many functional genes have been reported in the past decades. In rice, many important traits of great agronomic application are identified by this technique. As mentioned above, the cloning of photo/thermo-sensitive male sterility gene is a great breakthrough in hybrid rice. Recently, Wang et al (2012) has ingeniously applied this technique in cloning a gene*, OsSPL16*, that control of grain size, shape and quality in rice by using a NIL line (Wang et al., 2015; Wang et al., 2012b). After the identification of *OsSPL16*, the author further cloned a new gene, *GW7* also by using map-base cloning, which acts with *OsSPL16* as a regulatory module to determine grain shape and simultaneously improves rice yield and grain quality (Wang et al, 2015). Their results were highly praised by Professor McCouch from Connell University owing to the simultaneously improves yield and grain quality, which is a pair of conflit trait that makes the improvement of yield and grain quality at the same time a very difficult problem. More and more important gene for plant breeding are cloned by map-base cloning indicating it a effective forward genetic technique in identifying new genes, although it has the disadvantages of time-consuming and labor demanded. Moreover, a dense genetic marker developed for the nutation locus will be of great help for cloning of the mutant gene, thus the molecular marker has been greatly developed and applied to map-base cloning, such as SSR, SNP and Indel (Hou et al., 2010).

In this study, 81 SSR markers well-distributed in rice genome were selected to the primary cloning. A small F2 population with 40 individuals was used to screen these SSR markers and result shown that one of the markers from chromosome 3 was linked

to the mutant phenotype, indicating the mutant locus was in chromosome 3. More SSRs around this marker was developed to primarily locate the mutant locus, and two markers, 15236 and 15353, were found to link with mutant locus. Unfortunately, the centromere is included inside the region between these two markers, which make our map base cloning more difficult owing to the extremely low recombination rate at centromere region (Hou et al., 2010). Thus, for the fine mapping, we selected another 132 SSRs between these two markers to examine a large scale F2 population. Results shown that we can only locate the mutant locus to a 2.17M DNA region between R45 and R47 even though we have enlarged the F2 population to 20000 individuals (Fig. 3-8), further prove that the recombination rate in centromere is extremely low.

Next generation sequencing has been widely used in gene cloning a few years ago, which has greatly accelerated the cloning progress and improved the accuracy. Different strategies have been developed for various mutant phenotypes. Among those frequently used sequencing techniques, MutMap is a recently developed sequencing technique, which is also efficient and affordable compared with traditional methods. Base on whole genome sequencing (WGS), the MutMap has been applied in many species for loci studies. It is an efficient cloning technique without the construction of mapping population, not depending on genetic linkage map and the development of huge number of molecular markers. By using SNP analysis, Abe et al (2012) first time used MutMap to locate 7 mutant genes from 7mutnat lines which were crossed their parent line and create three DNA pool with around 30 individual DNA according to the mutant phenotype in F2 population. In their experiment, SNP-index has been applied, in which all reads that cover a specific DNA region containing the same SNP which is different with the reference genome was set as 1, whereas only half reads containing this SNP will be set as 0.5. With this SNP-index, the analysis result can be plotted and that the SNP near the mutant location are usually higher than others. The MutMap has been successfully used in many species and cloned many important genes, such as some important agronomic trait genes, rice male sterility gene, salt-tolerant gene and dwarf gene.

In our study, we have located the mutant position into a 2.17M DNA region in chromosome 3. Due to the low cross over rate in centromere region, we could not make any progress in reducing the distance. Then the MutMap was used to clone the target

gene, result shown that those SNP-indexs of 1 were found only in part of the 12 chromosomes (Fig.3-9, Fig.3-10 and Fig.3-11), which was also listed in table 3-3. As we can see although the chromosome 11 has the most candidate of SNP, two SNP with index of 1 were found in near the centromere region, however, only one of which was contained in the DNA region of our map-base cloning result (Fig 3-8). By combination map-base cloning with MutMap cloning, we have accelerated our mapping progress and cloned a candidate gene for *TFS1*, which is an Argonaute gene. Since it is only a SNP between *tfs1* mutant and 4266 plant in this candidate gene, we constructed two kinds of transgenic plants to verify the candidate, one is complement line with whole genome DNA of target gene, including the gene intron domain and promoter domain, the other is over-expression line with cDNA of candidate gene, both recombined vectors were transfer into *tfs1* mutant rice. Unfortunately, none of the OE lines has restored the sterility phenotype of *tfs1* plants. The expression of candidate gene has been examined by qRT-PCR, result shown that almost all of the OE lines had a low expression level of *TFS1* gene when compared with 4266 plant (Fig. 3-12), which may explain why those OE lines didn't display a restored phenotype. In transgenic plant, over expression often result in a co-suppression effect (Jorgensen, 1995). The same phenotype these OE plants with *tfs1* mutant might be resulted by co-suppression. However, later we have got several complement lines with the whole candidate gene fragment from 4266 plant. As we expected, they have become back to the phenotype of 4266 plants, which prove that the candidate gene is *TFS1* gene, the mutant gene is cloned.

The Argonaute protein has a complex structure which has been well documented, it is mainly contain six domains. Previous study revealed that and many mutations in these domains have severe phenotypes. To date there is no crystal structure of plant Argonaute proteins, since the structure of Argonaute is conserved, we can find some clues from the reported Argonaute models. But the function is unique for each Argonaute, they can load different kinds of guide RNA and pair with different target strands, which means special mechanism is remain unclear. Researchers usually focus on the function of domains when investigate the function of a protein, our mutant, which has a single SNP mutation, may show a new clue that the intermediate zoneis also as important as domains.

AGO proteins are typically composed of 700~1200 amino acid residues, having a molecular weight of ~100kDa, which are conserved and expressed in a wide range of species and tissues. As the core component of RISC, Argonaute proteins can affect the growth and development as well as the responses to abiotic and biotic stresses. Further studies reveal that AGO proteins not only regulate gene expression, but also participate in miRNA biogenesis(Mi et al., 2008). In plant, it has been reported that high degree of sequence complementary to their target mRNAs for miRNAs is required, which directs the cleavage of target mRNA. Endogenous classes of sRNA trigger TGS and PTGS depends on their length, 21- and 22-nucleotide in length including miRNAs, ta-siRNAs and nat-siRNAs can direct PTGS. While the 24-nucleotide sRNA, including p4- and p5-siRNAs, can direct DNA methylation and TGS. mall RNAs are a short double-stranded RNA of about 20-24nt in length and widely existed in plant and animals which are produced form long hairpin RNA or long double strand RNA by DICER proteins. After transported from nucleus to cytoplasm, pre-miRNA will be and further processed by DICER and its partner proteins which form a complex with AGO protein. Further researches revealed that DICER directly interact with AGOs in determining which strand in the duplex will be the guide RNA and finally form RISC leading to the strand with less stable paired 5' end is preferentially loaded into AGO proteins. These two kinds of function for AGOs indicate its important role in regulating RNA's turnover during plant growth and development. Thus in the present study, we firstly examined the mRNA profiles in different phenotype plants, the 4266 plant, *tfs1* mutant and both plants grown under lower temperature.

When *tfs1* mutant is grown under lower temperature, part of the floret is restored from the mutant phenotype, which was sampled for RNA-seq analysis. As shown in the RNA-seq results, the DEGs numbers from group of 4266LT-uf vs MR-uf and 4266LT vs MR is significantly lower than that from the other two groups, respectively (Fig. 4-5), especially in plants before fertilization. MR means those florets having same phenotype with 4266 plant, indicating gene expression pattern in MR and 4266 should be much more similar. Thus our result in RNA-seq is reliable and the following analysis will be of great help in searching the target gene for *TFS1*. To study which biological processes of these DEGs are involved, we have performed GO analysis. Our result revealed that the response to endogenous stimulus is highly GO enriched, some genes

are related to ABA pathway and some are related to ethylene pathway, which have been reported that play important roles in PCD regulation during endosperm development in maize. Transcription is also highly GO enriched, which is closely relative to the phenotype of *tfs1* mutant (Yu et al., 2012). Such as the MADS-box family, MYB, HD-Zip family and ARF family, which are reported in many species that play important roles in regulating this process (Hawker and Bowman, 2004). Most of members of these families have significant DE between these plants. This result is consistent with previous studies showing that these TFs were regulated by AGO protein. We also have performed the small RNA sequencing to study the mutant gene's effect on the sRNA. Consistent with RNA-seq result, the miRNA profiles in 4266 and MR florets are similar but different with that from MLT both in unfertilized and fertilized floret from these plants. Furthermore, the target gene for these DE miRNAs again points to the TFs that are evolved in flower development (Fig. 4-11), suggesting the mutant gene playing a key role in this process. To further prove this, the correlative analysis between results from RNA-seq and small RNA-seq has been done to discover the relationship between miRNA, siRNA and gene expression in *tfs1* mutant. Our sequencing result have provided some primary result for the regulation mechanism of TFS1 protein, to further explore the molecular function of TFS1, more experiments by reverse genetic techniques will be carried out in the future.

In conclusion, we have identified a female sterility line, which is proved to be controlled by a single recessive nuclear gene by crossing experiments. Cytological study by CLSM, paraffin section, SEM and so on have shown that this mutant has a complete embryo sac but a defective hull, the pollen germination grown on stigma of *tfs1* mutant plant were also normal when compared with 4266 plant. However, after fertilization, the synergy cell and egg cell in embryo sac of *tfs1* did not disappear, which results in failure of double fertilization. The map base cloning and MutMap were used to clone *TFS1* gene, by combining these two cloning techniques, we successfully located this mutant gene to an AGO protein. Our complementary line which restored the mutant phenotype has further approved our cloning result. Then RNA-seq and small RNA-seq were used to explore the regulation mechanism of *TFS1* gene. We found some DEGs have correlation with miRNAs and siRNAs, and the function of DEGs is consistent with the highly enriched GO terms from transcriptome analysis, which means *TFS1*

regulated the expression of both miRNA, siRNA and genes. This result is consistent with previous study and further suggest that mutation of *TFS1* has great influence on transcription and protein modification relative genes.

References

Achard, P., Cheng, H., De Grauwe, L., Decat, J., Schoutteten, H., Moritz, T., Van Der Straeten, D., Peng, J., and Harberd, N.P. (2006). Integration of plant responses to environmentally activated phytohormonal signals. Science *311*, 91-94.

Arikit, S., Xia, R., Kakrana, A., Huang, K., Zhai, J., Yan, Z., Valdes-Lopez, O., Prince, S., Musket, T.A., Nguyen, H.T.*, et al.* (2014). An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. Plant Cell *26*, 4584-4601.

Axtell, M.J., Jan, C., Rajagopalan, R., and Bartel, D.P. (2006). A two-hit trigger for siRNA biogenesis in plants. Cell *127*, 565-577.

Baumberger, N., and Baulcombe, D.C. (2005). Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proceedings of the National Academy of Sciences of the United States of America *102*, 11928-11933.

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

Berger, F. (1999). Endosperm development. Current opinion in plant biology *2*, 28-32.

Berger, F., Hamamura, Y., Ingouff, M., and Higashiyama, T. (2008). Double fertilization–caught in the act. Trends in plant science *13*, 437-443.

Blackmore, S., Wortley, A.H., Skvarla, J.J., and Rowley, J.R. (2007). Pollen wall development in flowering plants. New Phytologist *174*, 483-498.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (2012). The ABC model of flower development: then and now. Development *139*, 4095-4098.

Breakfield, N.W., Corcoran, D.L., Petricka, J.J., Shen, J., Sae-Seaw, J., Rubio-Somoza, I., Weigel, D., Ohler, U., and Benfey, P.N. (2012). High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis. Genome Res *22*, 163-176.

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

Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S., and Peacock, W.J. (1997). Fertilization-independent seed development in Arabidopsis thaliana. Proceedings of the National Academy of Sciences *94*, 4223-4228.

Chen, L., and Liu, Y.-G. (2014). Male sterility and fertility restoration in crops. Annual review of plant biology *65*, 579-606.

Chen, T.-H., Lam, L., and Chen, S.-C. (1985). Somatic embryogenesis and plant regeneration from cultured young inflorescences of Oryza sativa L.(rice). Plant cell, tissue and organ culture *4*, 51-54.

Chen, Y.-H., Li, H.-J., Shi, D.-Q., Yuan, L., Liu, J., Sreenivasan, R., Baskar, R., Grossniklaus, U., and Yang, W.-C. (2007). The central cell plays a critical role in pollen tube guidance in Arabidopsis. The Plant cell *19*, 3563-3577.

Chen, Z., Yan, W., Wang, N., Zhang, W., Xie, G., Lu, J., Jian, Z., Liu, D., and Tang, X. (2014). Cloning of a rice male sterility gene by a modified MutMap method. Yi chuan= Hereditas *36*, 85-93.

Cheung, A.Y., Palanivelu, R., Tang, W.-H., Xue, H.-W., and Yang, W.-C. (2013). Pollen and plant reproduction biology: blooming from East to West. Molecular plant *6*, 995-997.

Chhun, T., Aya, K., Asano, K., Yamamoto, E., Morinaka, Y., Watanabe, M., Kitano, H., Ashikari, M., Matsuoka, M., and Ueguchi-Tanaka, M. (2007). Gibberellin regulates pollen viability and pollen tube growth in rice. The Plant cell *19*, 3876-3888.

Chitwood, D.H., and Timmermans, M.C. (2010). Small RNAs are on the move. Nature *467*, 415.

Curtis, M.D., and Grossniklaus, U. (2008). Molecular control of autonomous embryo and endosperm development. Sexual Plant Reproduction *21*, 79-88.

Cutter, A.D. (2012). The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. Trends in ecology & evolution *27*, 209-218.

Dasheng, Z., Hexin, T., and Dabing, Z. (2009). Understanding the molecular mechanism of rice pollen development. The responsibility for this publication rests with the International Rice Research Insti-tute Copyright International Rice Research Institute 2009 This publication is copyrighted by the International Rice Research Institute (IRRI) and is licensed for use under a Creative Commons Attribution-NonCommercial-ShareAlike 30, 305.

Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., Xu, C., Li, X., Xiao, J., and Zhang, Q. (2012). A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. Proceedings of the National Academy of Sciences *109*, 2654-2659.

Douglas, R.N., Wiley, D., Sarkar, A., Springer, N., Timmermans, M.C., and Scanlon, M.J. (2010). ragged seedling2 Encodes an ARGONAUTE7-like protein required for mediolateral expansion, but not dorsiventrality, of maize leaves. The Plant cell *22*, 1441-1451.

Dreni, L., Jacchia, S., Fornara, F., Fornari, M., Ouwerkerk, P.B., An, G., Colombo, L., and Kater, M.M. (2007). The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. The Plant Journal *52*, 690-699.

Duckett, C., Grierson, C., Linstead, P., Schneider, K., Lawson, E., Dean, C., Poethig, S., and Roberts, K. (1994). Clonal relationships and cell patterning in the root epidermis of Arabidopsis. Development *120*, 2465-2474.

Eckardt, N.A. (2007). Elucidating the function of synergid cells: a regulatory role for MYB98 (Am Soc Plant Biol).

Fan, L.M., Wang, Y.F., Wang, H., and Wu, W.H. (2001). In vitro Arabidopsis pollen germination and characterization of the inward potassium currents in Arabidopsis pollen grain protoplasts. Journal of experimental botany *52*, 1603-1614.

Fan, Y., Yang, J., Mathioni, S.M., Yu, J., Shen, J., Yang, X., Wang, L., Zhang, Q., Cai, Z., Xu, C.*, et al.* (2016). PMS1T, producing phased small-interfering RNAs, regulates photoperiod-sensitive male sterility in rice. Proceedings of the National Academy of Sciences of the United States of America *113*, 15144-15149.

Fang, X., and Qi, Y. (2016). RNAi in Plants: An Argonaute-Centered View. The Plant cell *28*, 272-285.

Feng, J., Lu, Y., Liu, X., and Xu, X. (2000). Pollen development and its stages in rice (Oryza sativa L.). Zhongguo shuidao kexue *15*, 21-28.

Fierst, J.L., and Hansen, T.F. (2010). Genetic architecture and postzygotic reproductive isolation: evolution of Bateson–Dobzhansky–Muller incompatibilities in a polygenic model. Evolution *64*, 675-693.

Fladung, M., Bossinger, G., Roeb, G., and Salamini, F. (1991). Correlated alterations in leaf and flower morphology and rate of leaf photosynthesis in a midribless (mbl) mutant of Panicum maximum Jacq. Planta *184*, 356-361.

Fujii, S., and Toriyama, K. (2005). Molecular mapping of the fertility restorer gene for ms-CW-type cytoplasmic male sterility of rice. Theoretical and applied genetics *111*, 696-701.

German, M.A., Luo, S., Schroth, G., Meyers, B.C., and Green, P.J. (2009). Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nat Protoc *4*, 356-362.

Grini, P.E., Jürgens, G., and Hülskamp, M. (2002). Embryo and endosperm development is disrupted in the female gametophytic capulet mutants of Arabidopsis. Genetics *162*, 1911-1925.

Harper, R.M., Stowe-Evans, E.L., Luesse, D.R., Muto, H., Tatematsu, K., Watahiki, M.K., Yamamoto, K., and Liscum, E. (2000). The NPH4 locus encodes the auxin response factor ARF7, a conditional regulator of differential growth in aerial Arabidopsis tissue. The Plant Cell *12*, 757-770.

Haughn, G.W., and Somerville, C.R. (1988). Genetic control of morphogenesis in Arabidopsis. genesis *9*, 73-89.

Hawker, N.P., and Bowman, J.L. (2004). Roles for Class III HD-Zip and KANADI genes in Arabidopsis root development. Plant Physiology *135*, 2261-2270.

Higashiyama, T. (2002). The synergid cell: attractor and acceptor of the pollen tube for double fertilization. Journal of plant research *115*, 0149-0160.

Higashiyama, T., and Yang, W.-c. (2017). Gametophytic pollen tube guidance: attractant peptides, gametic controls, and receptors. Plant physiology *173*, 112-121.

Hong-Qi, Z., and Croes, A. (1982). A new medium for pollen germination in vitro. Acta botanica neerlandica *31*, 113-119.

Hou, X., Li, L., Peng, Z., Wei, B., Tang, S., Ding, M., Liu, J., Zhang, F., Zhao, Y., and Gu, H. (2010). A platform of high - density INDEL/CAPS markers for map - based cloning in Arabidopsis. The Plant Journal *63*, 880-888.

Hu, C., Zeng, Y., Lu, Y., Li, J., and Liu, X. (2009). High embryo sac fertility and diversity of abnormal embryo sacs detected in autotetraploid indica/japonica hybrids in rice by whole - mount eosin B - staining confocal laser scanning microscopy. Plant breeding *128*, 187-192.

Huang, B.-Q., and Sheridan, W.F. (1996). Embryo sac development in the maize indeterminate gametophyte1 mutant: abnormal nuclear behavior and defective microtubule organization. The Plant cell *8*, 1391-1407.

Huck, N. (2003). The Arabidopsis mutant feronia disrupts the female gametophytic control of pollen tube reception. Development *130*, 2149-2159.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nature reviews Molecular cell biology *9*, 22-32.

International Rice Genome Sequencing, P. (2005). The map-based sequence of the rice genome. Nature *436*, 793-800.

Inui, M., Martello, G., and Piccolo, S. (2010). MicroRNA control of signal transduction. Nature reviews Molecular cell biology *11*, 252.

Itoh, J.-I., Nonomura, K.-I., Ikeda, K., Yamaki, S., Inukai, Y., Yamagishi, H., Kitano, H., and Nagato, Y. (2005). Rice plant development: from zygote to spikelet. Plant and Cell Physiology *46*, 23-47.

Ji, L., Liu, X., Yan, J., Wang, W., Yumul, R.E., Kim, Y.J., Dinh, T.T., Liu, J., Cui, X., and Zheng, B. (2011). ARGONAUTE10 and ARGONAUTE1 regulate the termination of floral stem cells through two microRNAs in Arabidopsis. PLoS genetics *7*, e1001358.

Jones-Rhoades, M.W., and Bartel, D.P. (2004). Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. Molecular Cell *14*, 787-799.

Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A., and Timmermans, M.C. (2004). microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. Nature *428*, 84.

Kaplan, J. (2008). The end of the adaptive landscape metaphor? Biology & Philosophy *23*, 625-638.

Kasahara, R.D., Portereiko, M.F., Sandaklie-Nikolova, L., Rabiger, D.S., and Drews, G.N. (2005). MYB98 is required for pollen tube guidance and synergid cell differentiation in Arabidopsis. The Plant cell *17*, 2981-2992.

Kerim, T., Imin, N., Weinman, J.J., and Rolfe, B.G. (2003). Proteome analysis of male gametophyte development in rice anthers. Proteomics *3*, 738-751.

Kerstetter, R.A., Laudencia-Chingcuanco, D., Smith, L.G., and Hake, S. (1997). Loss-of-function mutations in the maize homeobox gene, knotted1, are defective in shoot meristem maintenance. Development *124*, 3045-3054.

Kidner, C.A., and Martienssen, R.A. (2005). The role of ARGONAUTE1 (AGO1) in meristem formation and identity. Developmental biology *280*, 504-517.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature methods *12*, 357-360.

Knox, R. (1984). The pollen grain. In Embryology of angiosperms (Springer), pp. 197-271.

Kranz, E., and Kumlehn, J. (1999). Angiosperm fertilisation, embryo and endosperm development in vitro. Plant Science *142*, 183-197.

Kubo, T. (2013). Genetic mechanisms of postzygotic reproductive isolation: An epistatic network in rice. Breeding science *63*, 359-366.

Kubo, T., and Yoshimura, A. (2005). Epistasis underlying female sterility detected in hybrid breakdown in a Japonica–Indica cross of rice (Oryza sativa L.). Theoretical and applied genetics *110*, 346-355.

Kumar, M., Basha, P.O., Puri, A., Rajpurohit, D., Randhawa, G.S., Sharma, T.R., and Dhaliwal, H.S. (2010). A candidate gene OsAPC6 of anaphase-promoting complex of rice identified through T-DNA insertion. Functional & integrative genomics *10*, 349-358.

Kurata, N., Miyoshi, K., Nonomura, K.-I., Yamazaki, Y., and Ito, Y. (2005). Rice mutants and genes related to organ development, morphogenesis and physiological traits. Plant and cell physiology *46*, 48-62.

Kwee, H.S., and Sundaresan, V. (2003). The NOMEGA gene required for female gametophyte development encodes the putative APC6/CDC16 component of the Anaphase Promoting Complex in Arabidopsis. The Plant Journal *36*, 853-866.

La Rota, C., Chopard, J., Das, P., Paindavoine, S., Rozier, F., Farcot, E., Godin, C., Traas, J., and Moneger, F. (2011). A data-driven integrative model of sepal primordium polarity in Arabidopsis. The Plant cell *23*, 4318-4333.

Lark, K.G., Chase, K., Adler, F., Mansur, L.M., and Orf, J.H. (1995). Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proceedings of the National Academy of Sciences *92*, 4656-4660.

Li, D., Chen, L., Jiang, L., Zhu, S., Zhao, Z., Liu, S., Su, N., Zhai, H., Ikehashi, H., and Wan, J. (2007). Fine mapping of S32 (t), a new gene causing hybrid embryo sac sterility in a Chinese landrace rice (Oryza sativa L.). Theoretical and Applied Genetics *114*, 515-524.

Li, F., Liu, W., Tang, J., Chen, J., Tong, H., Hu, B., Li, C., Fang, J., Chen, M., and Chu, C. (2010). Rice DENSE AND ERECT PANICLE 2 is essential for determining panicle outgrowth and elongation. Cell research *20*, 838.

Li, H.-J., Xue, Y., Jia, D.-J., Wang, T., Liu, J., Cui, F., Xie, Q., Ye, D., and Yang, W.-C. (2011). POD1 regulates pollen tube guidance in response to micropylar female signaling and acts in early embryo patterning in Arabidopsis. The Plant cell *23*, 3288-3302.

Li, S., Qian, Q., Fu, Z., Zeng, D., Meng, X., Kyozuka, J., Maekawa, M., Zhu, X., Zhang, J., and Li, J. (2009). Short panicle1 encodes a putative PTR family transporter and determines rice panicle size. The Plant Journal *58*, 592-605.

Li, Z., Pinson, S.R., Paterson, A.H., Park, W.D., and Stansel, J.W. (1997). Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (Oryza sativa L.) population. Genetics *145*, 1139-1148.

Litt, A., and Kramer, E.M. (2010). The ABC model and the diversification of floral organ identity. Paper presented at: Seminars in cell & developmental biology (Elsevier).

Liu, W., Kohlen, W., Lillo, A., Op den Camp, R., Ivanov, S., Hartog, M., Limpens, E., Jamil, M., Smaczniak, C., Kaufmann, K.*, et al.* (2011). Strigolactone biosynthesis in Medicago truncatula and rice requires the symbiotic GRAS-type transcription factors NSP1 and NSP2. The Plant cell *23*, 3853-3865.

Long, Y., Zhao, L., Niu, B., Su, J., Wu, H., Chen, Y., Zhang, Q., Guo, J., Zhuang, C., and Mei, M. (2008). Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. Proceedings of the National Academy of Sciences *105*, 18871-18876.

Lopez-Dee, Z.P., Wittich, P., Pe, M.E., Rigola, D., Del Buono, I., Gorla, M.S., Kater, M.M., and Colombo, L. (1999). OsMADS13, a novel rice MADS-box gene expressed during ovule development. Developmental genetics *25*, 237-244.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lu, C., Jeong, D.-H., Kulkarni, K., Pillay, M., Nobuta, K., German, R., Thatcher, S.R., Maher, C., Zhang, L., and Ware, D. (2008a). Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proceedings of the National Academy of Sciences *105*, 4951-4956.

Lu, S., Sun, Y.H., and Chiang, V.L. (2008b). Stress-responsive microRNAs in Populus. Plant J *55*, 131-151.

Lukowitz, W., Gillmor, C.S., and Scheible, W.-R. (2000). Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you. Plant physiology *123*, 795-806.

Luo, D., Xu, H., Liu, Z., Guo, J., Li, H., Chen, L., Fang, C., Zhang, Q., Bai, M., and Yao, N. (2013). A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nature genetics *45*, 573-577.

Luo, M., Bilodeau, P., Dennis, E.S., Peacock, W.J., and Chaudhury, A. (2000). Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. Proceedings of the National Academy of Sciences *97*, 10637-10642.

Mallory, A., and Vaucheret, H. (2010). Form, function, and regulation of ARGONAUTE proteins. The Plant cell *22*, 3879-3889.

Martin, F.W. (1959). Staining and observing pollen tubes in the style by means of fluorescence. Stain technology *34*, 125-128.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., and Long, C. (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5′ terminal nucleotide. Cell *133*, 116-127.

Middleton, A.M., Ubeda-Tomas, S., Griffiths, J., Holman, T., Hedden, P., Thomas, S.G., Phillips, A.L., Holdsworth, M.J., Bennett, M.J., King, J.R.*, et al.* (2012). Mathematical modeling elucidates the role of transcriptional feedback in gibberellin signaling. Proceedings of the National Academy of Sciences of the United States of America *109*, 7571-7576.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell *133*, 128-141.

Moore, M.J., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., and Darnell, R.B. (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nature protocols *9*, 263-293.

Morin, R.D., Aksay, G., Dolgosheina, E., Ebhardt, H.A., Magrini, V., Mardis, E.R., Sahinalp, S.C., and Unrau, P.J. (2008). Comparative analysis of the small RNA transcriptomes of Pinus contorta and Oryza sativa. Genome Res *18*, 571-584.

Nagasaki, H., Itoh, J.-i., Hayashi, K., Hibara, K.-i., Satoh-Nagasawa, N., Nosaka, M., Mukouhata, M., Ashikari, M., Kitano, H., and Matsuoka, M. (2007). The small interfering RNA production pathway is required for shoot meristem initiation in rice. Proceedings of the National Academy of Sciences *104*, 14867-14871.

Nayar, S., Kapoor, M., and Kapoor, S. (2014). Post-translational regulation of rice MADS29 function: homodimerization or binary interactions with other seed-expressed MADS proteins modulate its translocation into the nucleus. Journal of experimental botany *65*, 5339-5350.

Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., and Laufs, P. (2006). The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. The Plant Cell Online *18*, 2929-2945.

Nowack, M.K., Grini, P.E., Jakoby, M.J., Lafos, M., Koncz, C., and Schnittger, A. (2006). A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis. Nature genetics *38*, 63.

Olmedo-Monfil, V., Durán-Figueroa, N., Arteaga-Vandázquez, M., Demesa-Arévalo, E., Autran, D., Grimanelli, D., Slotkin, K., Martienssen, R.A., and Vielle-Calzada, J.-P. (2010). Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature *464*, 628.

Pagnussat, G.C., Alandete-Saez, M., Bowman, J.L., and Sundaresan, V. (2009). Auxin-dependent patterning and gamete specification in the Arabidopsis female gametophyte. Science *324*, 1684-1689.

Pagnussat, G.C., Yu, H.-J., Ngo, Q.A., Rajani, S., Mayalagu, S., Johnson, C.S., Capron, A., Xie, L.-F., Ye, D., and Sundaresan, V. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. Development *132*, 603-614.

Palmer, R.G., Winger, C.L., and Albertsen, M.C. (1978). Four Independent Mutations at the ms1, Locus in Soybeans. Crop Science *18*, 727-729.

Patel, P., Ramachandruni, S.D., Kakrana, A., Nakano, M., and Meyers, B.C. (2015). miTRATA: a web-based tool for mi croRNA Tr uncation a nd T ailing A nalysis. Bioinformatics *32*, 450-452.

Patel, R., Tsan, A., Tam, R., Desai, R., Schoenbrunner, N., Myers, T.W., Bauer, K., Smith, E., and Raja, R. (2012). Mutation scanning using MUT-MAP, a high-throughput, microfluidic chip-based, multi-analyte panel. PloS one *7*, e51153.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol *33*, 290-295.

Prasad, K., Sriram, P., Kumar, S.C., Kushalappa, K., and Vijayraghavan, U. (2001). Ectopic expression of rice OsMADS1 reveals a role in specifying the lemma and palea, grass floral organs analogous to sepals. Development genes and evolution *211*, 281-290.

Raghavan, V. (1986). Embryogenesis in angiosperms: a developmental and experimental study, Vol 17 (CUP Archive).

Raghavan, V. (1988). Anther and pollen development in rice (Oryza sativa). American Journal of Botany, 183-196.

Raghavan, V. (2005). Role of non-zygotic parental genes in embryogenesis and endosperm development in flowering plants. Acta Biologica Cracoviensia *47*, 31-36.

Reflinur, Chin, J.H., Jang, S.M., Kim, B., Lee, J., and Koh, H.-J. (2012). QTLs for hybrid fertility and their association with female and male sterility in rice. Genes & Genomics *34*, 355-365.

Reiser, L., and Fischer, R.L. (1993). The ovule and the embryo sac. The Plant cell *5*, 1291.

Rick, C.M. (1966). Abortion of male and female gametes in the tomato determined by allelic interaction. Genetics *53*, 85.

Robinson-Beers, K., Pruitt, R.E., and Gasser, C.S. (1992). Ovule development in wild-type Arabidopsis and two female-sterile mutants. The Plant Cell *4*, 1237-1249.

Rotman, N., Rozier, F., Boavida, L., Dumas, C., Berger, F., and Faure, J.-E. (2003). Female control of male gamete delivery during fertilization in Arabidopsis thaliana. Current Biology *13*, 432-436.

Rubio-Somoza, I., Cuperus, J.T., Weigel, D., and Carrington, J.C. (2009). Regulation and functional specialization of small RNA–target nodes during plant development. Current opinion in plant biology *12*, 622-627.

Russell, S.D. (1992). Double fertilization. International Review of Cytology *140*, 357-388.

Sang, X., Li, Y., Luo, Z., Ren, D., Fang, L., Wang, N., Zhao, F., Ling, Y., Yang, Z., and Liu, Y. (2012). CHIMERIC FLORAL ORGANS1, encoding a monocot-specific

MADS box protein, regulates floral organ identity in rice. Plant physiology *160*, 788-807.

Sentoku, N., Sato, Y., Kurata, N., Ito, Y., Kitano, H., and Matsuoka, M. (1999). Regional expression of the rice KN1-type homeobox gene family during embryo, shoot, and flower development. The Plant Cell *11*, 1651-1663.

Sheoran, I.S., and Saini, H.S. (1996). Drought-induced male sterility in rice: changes in carbohydrate levels and enzyme activities associated with the inhibition of starch accumulation in pollen. Sexual Plant Reproduction *9*, 161-169.

Song, X., Qiu, S., Xu, C., Li, X., and Zhang, Q. (2005). Genetic dissection of embryo sac fertility, pollen fertility, and their contributions to spikelet fertility of intersubspecific hybrids in rice. Theoretical and applied genetics *110*, 205-211.

Stanley, R.G., and Linskens, H.F. (2012). Pollen: biology biochemistry management (Springer Science & Business Media).

Steer, M.W., and Steer, J.M. (1989). Pollen tube tip growth. New Phytologist *111*, 323-358.

t Hoen, P.A., Ariyurek, Y., Thygesen, H.H., Vreugdenhil, E., Vossen, R.H., de Menezes, R.X., Boer, J.M., van Ommen, G.J., and den Dunnen, J.T. (2008). Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res *36*, e141.

Takagi, H., Tamiru, M., Abe, A., Yoshida, K., Uemura, A., Yaegashi, H., Obara, T., Oikawa, K., Utsushi, H., and Kanzaki, E. (2015). MutMap accelerates breeding of a salt-tolerant rice cultivar. Nature biotechnology *33*, 445-449.

Tang, C., Xie, Y., and Yan, W. (2017). AASRA: An Anchor Alignment-Based Small RNA Annotation Pipeline. bioRxiv.

Taylor, L.P., and Hepler, P.K. (1997). Pollen germination and tube growth. Annual review of plant biology *48*, 461-491.

Tiwari, S.B., Hagen, G., and Guilfoyle, T. (2003). The roles of auxin response factor domains in auxin-responsive transcription. The Plant Cell *15*, 533-543.

Tsukamoto, T., Qin, Y., Huang, Y., Dunatunga, D., and Palanivelu, R. (2010). A role for LORELEI, a putative glycosylphosphatidylinositol - anchored protein, in Arabidopsis thaliana double fertilization and early seed development. The Plant Journal *62*, 571-588.

Twell, D. (2002). The developmental biology of pollen. Annual Plant Reviews *6*, 86-153.

Twell, D., Klein, T.M., Fromm, M.E., and McCormick, S. (1989). Transient expression of chimeric genes delivered into pollen by microprojectile bombardment. Plant physiology *91*, 1270-1274.

Vaucheret, H., Mallory, A.C., and Bartel, D.P. (2006). AGO1 homeostasis entails coexpression of MIR168 and AGO1 and preferential stabilization of miR168 by AGO1. Molecular cell *22*, 129-136.

Wan, J., Yamaguchi, Y., Kato, H., and Ikehashi, H. (1996). Two new loci for hybrid sterility in cultivated rice (Oryza sativa L.). TAG Theoretical and Applied Genetics *92*, 183-190.

Wang, H., Zhang, X., Liu, J., Kiba, T., Woo, J., Ojo, T., Hafner, M., Tuschl, T., Chua, N.H., and Wang, X.J. (2011). Deep sequencing of small RNAs specifically associated with Arabidopsis AGO1 and AGO4 uncovers new AGO functions. THe plant journal *67*, 292-304.

Wang, S.-Q., Shi, D.-Q., Long, Y.-P., Liu, J., and Yang, W.-C. (2012a). GAMETOPHYTE DEFECTIVE 1, a putative subunit of RNases P/MRP, is essential for female gametogenesis and male competence in Arabidopsis. PloS one *7*, e33595.

Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., Wang, Y., Chen, X., Zhang, Y., and Gao, C. (2015). The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. Nature genetics *47*, 949-954.

Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., Zeng, R., Zhu, H., Dong, G., and Qian, Q. (2012b). Control of grain size, shape and quality by OsSPL16 in rice. Nature genetics *44*, 950.

Wang, Y.P., and Li, K.B. (2009). Correlation of expression profiles between microRNAs and mRNA targets using NCI-60 data. BMC Genomics *10*, 218.

Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., Su, D., Chen, Y., Guo, J., and Luo, D. (2006). Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. The Plant Cell *18*, 676-687.

Wilson, Z.A., and Zhang, D.-B. (2009). From Arabidopsis to rice: pathways in pollen development. Journal of experimental botany *60*, 1479-1492.

Wu, J., Yang, Z., Wang, Y., Zheng, L., Ye, R., Ji, Y., Zhao, S., Ji, S., Liu, R., and Xu, L. (2015). Viral-inducible Argonaute18 confers broad-spectrum virus resistance in rice by sequestering a host microRNA. Elife *4*, e05733.

Wu, L., Zhang, Q., Zhou, H., Ni, F., Wu, X., and Qi, Y. (2009). Rice MicroRNA effector complexes and targets. The Plant cell *21*, 3421-3435.

Xia, R., Meyers, B.C., Liu, Z., Beers, E.P., Ye, S., and Liu, Z. (2013). MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA Biogenesis in Eudicots. Plant Cell *25*, 1555-1572.

Xiao, B., Liu, X., Lu, Y., Li, J., and Zhao, X. (2004). Structure and fertilization of embryo sac in intersubspecific hybrids of autotetraploid rice. Zuo wu xue bao *31*, 1150-1156.

Yadegari, R., and Drews, G.N. (2004). Female gametophyte development. The Plant cell *16*, S133-S141.

Yang, S.-L., Jiang, L., San Puah, C., Xie, L.-F., Zhang, X.-Q., Chen, L.-Q., Yang, W.-C., and Ye, D. (2005). Overexpression of TAPETUM DETERMINANT1 alters the cell fates in the Arabidopsis carpel and tapetum via genetic interaction with excess microsporocytes1/extra sporogenous cells. Plant Physiology *139*, 186-191.

Yang, W.-C., Shi, D.-Q., and Chen, Y.-H. (2010). Female gametophyte development in flowering plants. Annual review of plant biology *61*, 89-108.

Yang, W.-C., and Sundaresan, V. (2000). Genetics of gametophyte biogenesis in Arabidopsis. Current opinion in plant biology *3*, 53-57.

Yang, Y., Zhong, J., Ouyang, Y.-d., and Yao, J. (2013). The integrative expression and co-expression analysis of the AGO gene family in rice. Gene *528*, 221-235.

Ye, Q., Zhu, W., Li, L., Zhang, S., Yin, Y., Ma, H., and Wang, X. (2010). Brassinosteroids control male fertility by regulating the expression of key genes involved in Arabidopsis anther and pollen development. Proceedings of the National Academy of Sciences *107*, 6100-6105.

Yoshida, H., and Nagato, Y. (2011). Flower development in rice. Journal of experimental botany *62*, 4719-4730.

Yu, S., Galvao, V.C., Zhang, Y.C., Horrer, D., Zhang, T.Q., Hao, Y.H., Feng, Y.Q., Wang, S., Schmid, M., and Wang, J.W. (2012). Gibberellin regulates the Arabidopsis floral transition through miR156-targeted SQUAMOSA promoter binding-like transcription factors. Plant Cell *24*, 3320-3332.

Zeng, Y., Hu, C., Lu, Y., Li, J., and Liu, X. (2007). Diversity of abnormal embryo sacs in indica/japonica hybrids in rice demonstrated by confocal microscopy of ovaries. Plant Breeding *126*, 574-580.

Zeng, Y.X., Hu, C.Y., Lu, Y.G., Li, J.Q., and Liu, X.D. (2009). Abnormalities occurring during female gametophyte development result in the diversity of abnormal embryo sacs and leads to abnormal fertilization in indica/japonica hybrids in rice. Journal of integrative plant biology *51*, 3-12.

Zhai, L., Wang, L., Teng, F., Zhou, L., Zhang, W., Xiao, J., Liu, Y., and Deng, W. (2016). Argonaute and Argonaute-Bound Small RNAs in Stem Cells. International journal of molecular sciences *17*, 208.

Zhang, D., and Yang, L. (2014). Specification of tapetum and microsporocyte cells within the anther. Current opinion in plant biology *17*, 49-55.

Zhang, Q., Shen, B., Dai, X., Mei, M., Maroof, M.S., and Li, Z. (1994). Using bulked extremes and recessive class to map genes for photoperiod-sensitive genic male sterility in rice. Proceedings of the National Academy of Sciences *91*, 8675-8679.

Zhang, X., Niu, D., Carbonell, A., Wang, A., Lee, A., Tun, V., Wang, Z., Carrington, J.C., Chang, C.-e.A., and Jin, H. (2014). ARGONAUTE PIWI domain and microRNA duplex structure regulate small RNA sorting in Arabidopsis. Nature communications *5*, 5468.

Zhao, Z., Jiang, L., Zhang, W., Yu, C., Zhu, S., Xie, K., Tian, H., Liu, L., Ikehashi, H., and Wan, J. (2007). Fine mapping of S31, a gene responsible for hybrid embryo-sac abortion in rice (Oryza sativa L.). Planta *226*, 1087-1096.

Zhao, Z., Wang, C., Jiang, L., Zhu, S., Ikehashi, H., and Wan, J. (2006). Identification of a new hybrid sterility gene in rice (bi Oryza sativa L.). Euphytica *151*, 331-337.

Zhou, H., Liu, Q., Li, J., Jiang, D., Zhou, L., Wu, P., Lu, S., Li, F., Zhu, L., and Liu, Z. (2012). Photoperiod-and thermo-sensitive genic male sterility in rice are caused by a point mutation in a novel noncoding RNA that produces a small RNA. Cell research *22*, 649.

Zhu, D., and Deng, X.W. (2012). A non-coding RNA locus mediates environment-conditioned male sterility in rice. Cell research *22*, 791.

Zhu, Q.H., Spriggs, A., Matthew, L., Fan, L., Kennedy, G., Gubler, F., and Helliwell, C. (2008). A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res *18*, 1456-1465.