Running head: Gene Expression Profiling of Pollen Mother Cells in Rice

Corresponding author: Wei-Hua Tang
Address: National Key Laboratory of Plant Molecular Genetics, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China.
Phone: 86-21-54924072
Email: whtang@sibs.ac.cn

Research area: System biology, Molecular Biology and Gene Regulation
Global Gene Profiling of Laser-Captured Pollen Mother Cells Indicates Molecular Pathways and Gene Subfamilies Involved in Rice Meiosis

Xiang Tang1,2,5, Zhi-Yong Zhang1,5, Wen-Juan Zhang3, Xing-Ming Zhao3, Xuan Li4, Dong Zhang1, Qiao-Quan Liu2 and Wei-Hua Tang1*

1National Key Laboratory of Plant Molecular Genetics, 4Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

2Key Laboratory of Plant Functional Genomics of Ministry of Education, Yangzhou University, Yangzhou 225009, China

3Institute of Systems Biology, Shanghai University, Shanghai 200444, China

5Equal contribution

Footnote

This work was supported by the Ministry of Science and Technology of China (grant nos. 2007CB108700, 2011CB944600, 2006AA10A102 and 2009ZX08009-136B), and the National Natural Science Foundation of China (grant no. 30770196). X.-M. Zhao was partly supported by Shanghai Rising-Star Program (10QA1402700).

*Corresponding author, e-mail whtang@sibs.ac.cn.
Abstract

Pollen mother cells (PMCs) represent a critical early stage in plant sexual reproduction in which the stage is set for male gamete formation. Understanding the global molecular genetics of this early meiotic stage has so far been limited to whole stamen or floret transcriptome studies, but since PMCs are a discrete population of cells in developmental synchrony, they provide the potential for precise transcriptome analysis and for enhancing our understanding of the transition to meiosis. As a step towards identifying the pre-meiotic transcriptome, we performed microarray analysis on a homogenous population of rice PMCs isolated by laser microdissection, and compared them to those of tri-cellular pollen (TCP) and seedlings. Known meiotic genes, including OsSPO11-1, PAIR1, PAIR2, PAIR3, OsDMC1, OsMEL1, OsRAD21-4, OsSDS and ZEP1 all showed preferential expression in PMCs. The KEGG pathways significantly enriched in PMC-preferential genes are DNA replication and repair pathways. Our genome-wide survey showed that, in the build-up to meiosis, PMCs accumulate the molecular machinery for meiosis at the mRNA level. We identified 1158 PMC-preferential genes, and suggested candidate genes and pathways involved in meiotic recombination and meiotic cell cycle control. Regarding the developmental context for meiosis, the DEF-like, AGL2-like and AGL6-like subclades of MADS box transcription factors are PMC-preferentially expressed, the trans-zeatin type of cytokinin might be preferentially synthesized and gibberellin signaling pathway is likely active in PMCs. Ubiquitin-mediated proteolysis pathway is enriched in the 127 genes that are expressed in PMCs but not in TCP or seedlings.

Keywords: Agilent microarray, DNA replication and repair, Oryza sativa, pre-meiotic transcriptome, ubiquitin-mediated protein degradation
Introduction

In flowering plants, male reproductive cells develop in anthers. After differentiation from progenitor cells in the anther primordium, the sporogenous cells divide to generate a population of pollen mother cells (PMCs). More or less synchronously, PMCs enter meiosis (then called meiocytes) to produce four connected haploid microspores (tetrad), which then separate as unicellular microspores. Each microspore undergoes an asymmetric mitosis to produce the bicellular pollen, which comprises a vegetative cell and an embedded generative cell. The generative cell divides to form two sperm, and the resulting male gametophyte is called the tricellular pollen (McCormick 1993). The male reproductive cells are surrounded by several different layers of sporophytic cells of the anther throughout their development.

Complex gene networks in both gametophytic and sporophytic tissues of the anther regulate male reproductive development. Genome-wide gene expression data in a cell type- and stage-specific manner should provide a detailed framework for modeling cell fate and male reproductive development. Previous transcriptomic studies using stamens or florets, which are composed of mixed types of cells, may identify some relevant genes based on the expressional changes among different anther development stages. For example, Lu et al., (2006) dissected the rice stamens at various stages for microarray analysis, reported 26 genes to be preferentially up-regulated during early stamen development and therefore been suggested to participate in regulation of rice stamen development at pre-meiotic stages. Ma et al., (2008) performed microarray analysis on maize anthers of seven developmental stages (from pre-meiotic to mature pollen), reported 234 genes as pre-meiosis-related and 674 genes as persistent through meiosis. However, such studies cannot provide genome-wide gene expression information of a single cell type.

Laser microdissection combined with microarray enables cell-type specific expression profiling. With this tool, Suwabe et al., (2008) reported the separated transcriptomes of the male gametophyte and their surrounding tapetum during and after meiosis, but expression data from pre-meiotic PMCs are still lacking.

PMCs represent a key stage for transcriptome research, since only PMCs and megaspore mother cells will enter meiosis, while all other cell divisions are mitotic. A complete set of RNA transcripts in a homogenous population of cells poised to undergo meiosis holds tremendous
information on how a cell prepares for meiosis.

The process of meiosis is highly conserved in all eukaryotes, but comparison of meiosis-related pathways in yeast and mouse indicates significant diversity exists in the signal cascades that control the process of meiosis (Chu et al., 1998; Marston and Amon, 2004; Baltus et al., 2006). In maize, the requirement of a plant-specific protein Ameiotic 1 (Am1), for meiosis initiation (Pawlowski et al., 2009) also supports this idea. Thus obtaining the pre-meiotic transcriptome in plants is likely to uncover novel genes and pathways. In addition, rice is an important crop, and crop traits are reassorted during meiosis. Directly studying molecular mechanisms of meiosis in rice can reduce the distance from biological knowledge to the application in rice breeding.

With this goal in mind, we used laser capture microdissection of rice (Oryza sativa L. ssp. japonica) stamens to isolate PMCs and their transcripts, followed by transcriptome analysis using microarray hybridization. Using two color probe hybridization with Agilent 60-mer oligo microarrays, PMC transcripts were compared to transcripts from two tissues, tricellular pollen (TCP) that comprises three non-dividing cells, and seedling, which contains many mitotic dividing cells. Known meiotic genes show significant higher expression in PMCs than in TCP or seedlings. Based on the selective expression patterns and protein sequence homology, we sort out genes encoding core proteins involved in meiotic double-strand DNA breaks (DSB) formation and repairs, genes controlling of meiotic cell cycle progression, as well as molecular context for meiosis.

Results and Analysis

Laser microdissection and microarray analysis of rice PMCs

Stamens in immature panicles of rice cv Nipponbare harbor PMCs just before entering meiosis (Fig. 1, A and B). Chromosomal staining, callose staining and morphology of surrounding cell layers in the anther helped to identify this specific developmental stage (Fig. 1, C-E; Supplemental Fig. 1). Using a tissue preparation method optimized for preservation of both mRNA and cytological features (Tang et al., 2006), we isolated a homogenous population of PMCs by laser capture microdissection (Fig. 1, G-I) from cross-sections of these panicles. Total RNAs of high quality were extracted from two independent biological samples of PMCs (~800 per isolation) (Supplemental Fig. 2A). The polyA+ mRNA was then linearly amplified (Supplemental Fig. 2B)
for microarray hybridization. We used the Agilent 44k rice genome microarray (GPL6864) to profile PMCs, using TCP (Fig. 1F) and one week-old seedlings for comparisons. The microarray covers about 95% of the currently annotated rice genes/loci. A total of 29,008 distinct rice genes have representative probes (60-mers) on the array, including 28,840 putative protein-coding genes and 168 miRNA-coding genes. The microarray data were processed based on sufficient considerations for cell-type specific transcriptome analysis to avoid data distortion (details in methods).

Verification of microarray data

The variation range of mRNA copy numbers of individual genes within a cell was estimated as five orders of magnitude (Patanjali et al., 1991). For 60-mer oligo microarray hybridization, the mRNA concentration (which represents expression level) linearly correlates with signal intensity spanning more than five orders of magnitude until intensity values approach saturation (Hughes et al., 2001; Shi et al., 2006). The detection depth allows us to further group gene expression levels as low, medium or high based on hybridization intensities. In PMCs, there are about 25% expressed genes whose Log₂ intensity values are below 11, and another 25% whose Log₂ intensity value are above 15. Therefore we grouped the expression levels as low, medium and high for log₂ intensity value ranges 9-11, 11-15 and above 15.

We performed quantitative RT-PCR to test the expression of 13 genes, using as a reference Os03g0268000, which encodes a putative protein phosphatase that is highly expressed in PMCs, TCP and seedling. Expression levels were consistent with microarray data (Fig. 2) for medium to highly expressed genes, but some differences were seen for weakly expressed genes.

We also performed in situ hybridization for three PMC-expressed genes including a SKP1-like (Os07g0409500), Am1-like (Os03g0650400) and SPL-like (Os01g0212500). The results confirmed their expression in PMCs (Fig. 3). SKP1-like (Os07g0409500) is specifically expressed in PMCs, while Am1-like and SPL-like are expressed in PMCs as well as in surrounding somatic cells.

26 genes that preferentially up-regulated during early stamen development were suggested to participate in regulation of rice stamen development at pre-meiotic stages (Lu et al., 2006). Our microarray data (Supplemental Table 1) show that all 26 genes are expressed in PMCs, and many
of them (such as U78891 and AK070863) are expressed higher in PMCs than in TCP. The expression patterns of these genes from our microarray are consistent with Lu et al. (2006) and their suggested role in meiosis.

In order to estimate the absolute copy number of mRNA per cell for individual genes, we used laser microdissection to capture various numbers of PMCs (7, 16, 36 and 41) in separate pools. Then using plasmid DNA containing fragments of two rice genes, Os09g0480300 and Am1-like, as a reference, we performed real-time PCR to measure the mRNA copy number per captured PMC. The average mRNA numbers were approximately 600 and 1000, respectively. Correspondingly, the microarray intensity values for PMCs are 79,423 and 114,355 for these respective genes. Given the intensities sum of all PMC-expressed genes on the microarray was 455,220,997 and the linear correlation between 60mer probe intensities and mRNA concentration, we estimated the total mRNA copy number in a single PMC is about 3,650,000 (Supplemental Fig. 3).

Expression patterns of PMCs are more similar to that of seedlings than to that of TCP

60mer oligo microarrays can resolve gene expressional differences over five orders of magnitude (signal distribution ranges from 47 to 661,955 in our case, corresponding to 5.57 to 19.34 after log2 transformation in Fig. 4A). The density plots of all the hybridization signals from two biological replicates of PMCs as well as the signal distribution of negative control probes show that the hybridization intensities of expressed genes and non-expressed genes resolved to two peaks (Fig. 4, A and B; methods section). At least 59% (17,196 of 29,008) of rice genes are expressed in PMCs. This percentage is higher than that in TCP (41%, 11,867 genes), but lower than that in seedling (73%, 21,310 genes). The number of expressed rice genes in TCP is slightly higher than that of expressed maize genes in mature pollen (10,539, Ma et al., 2008).

In addition, PMCs share more expressed genes with seedling (16,128 of 17,196, 94%) than with TCP (10,202 of 17,196, 59%) (Fig. 4C). Combining transcriptomic data from the meiosis-microspore, tetrad, unicellular pollen, bicellular pollen, TCP and tapetum cells at meiosis, tetrad, and unicellular stages (Hobo et al., 2008; Suwabe et al., 2008) with our microarray data, our principal component analysis shows that the transcriptome of PMCs is more similar to that of seedling than to that of pollen/microspores after meiosis (Fig. 5A). The hierarchical clustering
dendrogram (Fig. 5B) shows that the seedling groups together with the pre-meiotic PMC, the male meiocytes and the finishing stage of meiosis, tetrads, while the later stages of uni-cellular, bi-cellular and tri-cellular pollen group together. These results further suggest that the cellular transcriptome is significantly changed at the end of meiosis.

**Expression of DNA replication and repair pathways are dominant in PMCs**

Because PMCs are a relatively uniform cell type, it is reasonable to consider that our transcript data represent pathways operating within the same cell. KEGG pathway enrichment assays can help interpreting microarray data (Kanehisa and Goto, 2000). Among all genes present on the microarray, 1618 has been annotated to 133 KEGG pathways. 91 pathways have 10 or more genes on this microarray, a threshold we selected as containing enough pathway steps to be meaningful. While the significantly enriched pathways in PMC-expressed genes and in TCP-expressed genes are scattered among different categories, and in contrast to that the enriched pathways in TCP- and seedling-preferentially expressed genes belong to metabolism, the five enriched pathways in PMC-preferentially expressed genes all belong to DNA replication and repair in the major category of genetic information processing (Table 1; Supplemental Table 2). It is a reasonable conjecture that PMC would activate genes directly responsible for DNA replication and recombination in order to prepare for meiosis.

**Meiotic genes show preferential expression in PMCs**

Then we checked the gene expression for all the twelve rice meiotic genes. Figure 6 left panel shows their involved steps and their expression data from our microarray. Supplemental Table 3 provides the description of these genes and their expression fold changes. All, except for two, of the twelve rice meiotic genes showed higher expression in PMCs than in TCP and seedlings. For example, OsSPO11-1 is thought to catalyze DSB which initiates homologous chromosome recombination (Yu *et al*., 2010). *OsSPO11-1* showed more than 10 fold higher expression in PMCs than in TCP or seedlings. PAIR1, a coiled-coiled protein has been reported to perform unknown function during DSB formation (Nonomura *et al*., 2004). *PAIR1* only expressed in PMCs, but not in TCP or seedlings. From a first impression, OsRAD51 (*Rajanikant et al*., 2008) and OsMER3/OsRCK (*Wang et al*., 2009; *Chang et al*., 2009) were exceptions. But RAD51 is a
recombinase that function in both meiosis and mitosis in budding yeast (Masson and West 2001), therefore the high expression of OsRAD51 in both PMCs and seedlings is consistent with its role. OsMER3/OsRCK functions in the formation of interfering crossovers. OsMER3/OsRCK has been reported to be preferentially expressed in young flowers by RT-PCR (Chang et al., 2009), while our microarray data based on Os02g0617500 probe showed no expression in any tested samples. It turned out that the 60mer probe of Os02g0617500 was not capable to represent OsMER3/OsRCK because it was designed based on previous incorrect annotation (see supplemental table 3 for detail). The PMC preferential expression patterns of the ten genes are consistent with their expected role in meiosis.

At the mRNA level, the molecular machinery for meiosis presents prior to the cellular behavior of meiosis

We define PMCs as sporogenous cells just before meiosis. The callose deposition, which blocks cytoplasmic connections between PMCs, could be considered as a landmark for meiosis initiation (McCormick, 1993). The PMCs we isolated for microarray analysis had not initiated meiosis based on their nuclear morphology (Fig. 1) and lack of callose deposition (Supplemental Fig. 1). However, DNA replication and repair pathways are enriched in genes that preferentially expressed in PMCs, and known meiotic genes showed selective expression in PMCs. This result nicely demonstrated a long-held theory that the molecular machinery for cell behavior is put in place before the cell actually carries out the behavior.

Meiosis can be considered as a modification of mitosis with meiotic-specific additions. Many cells in growing seedlings are undergoing mitosis, while TCP comprise cells that have finished cell division and probably have shut down cell division machinery and reprogrammed for cell fusion in fertilization. From the meiotic/mitotic point of view, it is reasonable that PMCs had more in common with seedlings than with TCP at the transtriptome level (Fig. 4C; Fig. 5), although morphologically PMCs are less similar to seedlings than to TCP.

Candidate genes involved in DSB formation and repairs

The conservation of meiosis among eukaryotes allows identification of rice meiotic genes to be initiated from candidate genes with sequence homology to characterized meiotic genes in other
eukaryotes (yeast, mammals, Arabidopsis and maize). However, sequence homology does not guarantee functional conservation, and in the case that multiple rice homologs can be found for one meiotic gene, it might be difficult to pinpoint the ortholog. The expression pattern resulted from our microarray analysis can provide another piece of information in functional characterization. The right panel of Figure 6 provides 34 candidate meiotic genes identified based on their sequence homology and PMC-preferential expression pattern from our microarray data.

For example, in the process of DSB formation and repair, maize POOR HOMOLOGOUS SYNAPSIS 1 (PHS1) functions in loading the recombination machinery onto chromosomes (Pawlowski et al., 2004). The putative OsPHS1 showed more than 30 fold higher expression in PMCs than in TCP or in seedlings, which suggesting its functional conservation with maize PHS1.

For another example, the heterotrimeric complex of replication protein A (RPA) is highly conserved, and is required for multiple processes including DNA replication, repair, and homologous recombination. Rice has three paralogs for RPA1 subunit, three for RPA2 subunit, one for RPA3 subunit, and two other RPA1-like proteins. Besides the characterized RPA1a (Chang et al., 2009a), one of the RPA2 paralogs (Os06g0693300) and another RPA1-like protein (Os06g0103400) are more likely to function in meiotic recombination based on their PMC-preferential expression pattern.

The base excision repair (BER) pathway is enriched in PMC-preferential genes. BER is thought to be responsible for repairing damaged DNA with single-strand breaks. Single-strand breaks of DNA can be processed by either short-patch (where a single nucleotide is replaced) or long-patch (where 2-10 new nucleotides are synthesized) BER. Figure 6 (right) and Supplemental Figure 4 show that of the seven genes that are significantly elevated in PMCs over TCP or seedling, six are of the long-patch type exclusively and one is common to both short- and long-patch BER. This suggests the long-patch BER might share common molecules with meiotic recombination, such as DNA glycosylases, methyl-CpG binding domain protein and exodeoxyribonuclease III for DSB end processing, and poly(ADP)-ribose polymerase and DNA ligase I for DSB repair.

**Candidate genes function in control of meiotic cell cycle**

Maize Am1 has been reported to control the initiation of meiosis (Pawlowski et al., 2009). However, Am1 has been found to be expressed highly in many tissues at mRNA level. Given the
specific function in meiosis only, one may wonder whether Am1 transcripts are really equally abundant in all cell types no matter meiotic or non-meiotic. Os03g0650400 is a rice homologue of Am1 (overall amino acids identity 62%, Supplemental Fig. 5A). Figure 6 (right) showed that Os03g0650400 was expressed about eight-fold higher in PMCs than in TCP or seedling. In-situ hybridization results (Fig. 3D) also show that Os03g0650400 transcripts are more abundant in PMCs than in surrounding cells at anther stage 5. The expression pattern for this gene is consistent with its pre-meiotic function.

Cyclins appear to play a major role regulating meiosis progression (Hamant et al., 2006). There are 33 cyclin-like genes present on the microarray, and our data suggest that, besides OsSDS which has been reported (Chang et al., 2009), only Os03g0203800 (OsCycD2;3) showed significantly higher expression in PMCs than in both TCP and seedlings (Supplemental Table 4).

The S-phase kinase-associated protein 1 (SKP1) is an adaptor protein of the SCF complex in the ubiquitin-mediated proteolysis pathway. Arabidopsis SKP1-like1 (ASK1) functions in chiasmata resolution and sister chromatid cohesion release (Yang et al., 1999) by promoting degradation of cyclins with F-box (Bai et al., 1996). There are 17 homologs of Arabidopsis SKP1-like genes in rice. Our expression data show that several of them preferentially expressed in PMCs (Fig. 6; Supplemental Table 4), and in situ hybridization results (Figs. 3A and B) show that one SKP1-like (Os07g0409500) was specifically expressed in PMCs, not in surrounding somatic cells. Another SKP1-like gene (Os07g0624900) is also among the 127 genes that we identified as only expressed in PMCs, but not in TCP or seedlings (Supplemental Table 5). Furthermore, F-box-like proteins (18) are the most significantly enriched group within the 127 genes. Ten of these F-box-like proteins were included in the list of rice F-box protein superfamily (Xu et al., 2009), and 4 of the 10 contain a FBD domain (pfam:cl11661) in addition to the N-terminal conserved F-box. Another F-box protein contains a DUF295 domain, while two more DUF295-containing proteins are expressed in PMCs, but not in TCP or seedling. F-box proteins are substrate-recognition components of the Skp1-Rbx1-Cul1-F-box protein (SCF) ubiquitin ligases.

The 127 gene list also includes two other components in ubiquitin-mediated proteolysis pathway, Os05g0352700 (a RING domain containing protein) and Os10g0141400 (a putative RPN10, 26S proteasome non-ATPase regulatory subunit, responsible for polyubiquitin chain
binding). The deviation of a conserved polyubiquitin chain binding motif (LAM/LALRL/V) (Fu et al., 1998) in Os10g0141400 (LAETFRLA) might suggest different substrates or binding mechanism in meiosis. The enrichment of ubiquitin-mediated protein degradation pathway genes in the 127 gene list suggests that proteolysis might also be a key mechanism that drives the events of meiosis, as it is of mitosis.

Specific subfamilies of transcription factors that preferentially expressed in PMCs might help providing developmental context for meiosis

There are 2384 transcription factors in 63 subfamilies in japonica rice according to DRTF database (Gao et al., 2006); 1541 (65%) of these are present on the microarray, and 908 (59%) are expressed in PMCs (Supplemental Table 7). Among PMC-expressed transcription factors, there are more highly expressed than weakly expressed ones (Supplemental Fig. 7).

The developmental context is required for successful completion of meiosis. For example, MADS box transcription factors, including ABC class proteins, are important in setting up the identity of floral organs, including anther. Two transcription factors, OsMADS3 and OsMADS58, might function similar to AGAMOUS (Yamaguchi et al., 2006). Supplemental Figure 6 shows detailed expression data of MADS transcription factors from a phylogenetic perspective. Among the 75 rice MADS proteins, 41 are present on our microarray. In the MIKC clade, members of the DEF-like, AGL2-like and AGL6-like subclades all show higher expression in PMCs than in TCP and seedling (Fig. 6). Within MIKC clade, OsMADS2, a member of GLO-like subclade, and OsMADS56, a member of TM3-like subclade also showed PMC-preferential expression. OsMADS72, the only member of Mα present on microarray, is expressed in PMCs, but not in TCP or seedling.

The Polycomb-group (Pc-G) transcription regulator family, the PHD family, the Alfin family and C3H family transcription factors are significantly enriched in genes highly expressed in PMCs, not enriched in genes highly expressed in TCP or seedlings (Supplemental Fig. 7). All 29 Pc-G proteins on the microarray are expressed in PMCs, a uniquely high level of representation among the 38 transcription factor sub-families having 10 or more members. PMC-highly expressed Pc-G include Os06g0275500 (homolog of maize enhancer of zeste 1) and Os12g0613200 (trithorax-like). Interestingly the two antagonistic groups of Pc-G (repressors and activators) both
are highly expressed in PMCs. Pc-Gs contain SET domain, the catalytic domain of protein methyltransferase. All the 34 genes encoding SET domain containing proteins present on the microarray (Supplemental Table 6) are expressed in PMCs, six of them are preferentially expressed in PMCs. Five out of the six belong to A subclade.

Protein kinases are common components of signaling pathways that function in anther development. Among the 1108 protein kinases on microarray (representing 77% of the 1429 annotated in the rice genome, based on Dardick et al., 2007), 572 (52%) are expressed in PMC (including 205 putative receptor kinases), 27 (including 13 putative receptor kinases) are expressed more than four-fold higher in PMCs than in TCP or seedling (Supplemental Table 14). A putative MAPK, Os02g0135200, expressed more than 20-fold higher in PMCs than in TCP and seedlings. From a phylogenetic standpoint, the expression pattern for protein kinases varies widely among members within subclades.

**Micro RNA level control of gene expression might be active in PMCs**

OsMEL1, a germ cell–specific member of the ARGONAUTE family has been reported be essential for meiosis progression. Our microarray results show that major components of micro RNA biosynthesis pathways including Argonautes, DICER-like and RDRP, are expressed in PMCs, suggesting that miRNA-related control of gene expression might be active in PMCs (Supplemental Fig. 8A). Some 60mer probes in our microarray (such as miR167h in Supplemental Fig. 8D) were designed to bind the stem-loop region of microRNA encoding genes. The linear amplification procedure in our RNA preparation amplifies exclusively polyA+ mRNA, so our microarray can also provide expression data for primary miRNA (pri-miRNA), but not for pre-miRNA or miRNA (which lack polyA tails). Pri-miRNA levels are thought to correlate well with miRNA level. Supplemental Figure 8B lists the nine microRNAs might be expressed in PMCs, among which miR167h is expressed in PMCs, exclusive of TCP or seedling.

**A global snapshot of plant hormone biosynthesis pathways in PMCs**

Hormone control is a key for male germ cell differentiation in mammals (Berruti 2006). However, while plant hormones are known to regulate almost every aspects of plant development including anther/pollen development, our knowledge on hormone regulation of meiosis is limited.
Seven types of plant hormone have been implicated to function in anther/pollen development (see supplemental table 16 for review). Because hormones can function as both local (paracrine) and long-distance signals (Faiss et al., 1997; Blakeslee et al., 2005), it is difficult to distinguish local signal from long-distance signal in whole organ studies. Recently, Hirano et al., (2008) reported that cell type-specific expression profiles of phytohormone biosynthesis and signaling genes in tapetum and microspore at meiosis and later stages. This enabled analyzing endogenous phytohormone biosynthesis and phytohormone response in a given cell type independently of other cells. Their results indicated that the sets of genes required for synthesis of auxin (IAA) and gibberellin (GA$_4$) highly expressed in post-meiotic microspore/pollen, while genes for gibberellin signaling were preferentially expressed in meiotic microspore and tapetum. Our microarray data allowed us to survey expression profiles of phytohormone related genes at pre-meiotic PMCs.

Because biosynthesis of the seven phytohormones are interconnected at metabolism level within a cell, we show global expression pattern of genes in seven hormones biosynthetic pathways in PMCs and TCP (Fig. 7 and Supplemental Fig. 9), based on the KEGG pathway map. The arrow colors indicate expression levels of the genes encoding enzymes responsible for this reaction. Among the eight hormones, the pathway genes of cytokinin show dominant expression pattern in PMCs. The genes encoding enzymes responsible for each step for biosynthesis from the TCA cycle to cytokinins (mainly trans-zeatin type) are all expressed in PMCs. Furthermore, genes encoding proteins in the pathway that leads to synthesis of trans-zeatin are mostly expressed higher in PMCs than the other two samples, indicating that PMCs might synthesize trans-zeatin. Cytokinins are capable of promoting mitotic cell division. Our results suggest a localized biosynthesis of trans-zeatin during meiotic cell division.

Combining microarray data from the microspore and the surrounding tapetum cells after PMC stages (Suwabe et al., 2008) with our data, we observed that the expression of genes responsible for the last two steps of trans-zeatin synthesis (Os05g0551700, Os07g0693500) were significantly reduced in later stages and in tapetum cells (Supplemental Fig. 9). Particularly, Os05g0551700 showed no expression or low expression in any cell types other than PMC, according to Suwabe et al. (2008) and Jiao et al. (2009). Os07g0693500 showed gradually reduced expression in rice anthers from stage 2 to stage 7, according to Lu et al. (2006). These data suggest that trans-zeatin biosynthesis might be reduced in later stages and in surrounding tapetum.
The green arrows in Figure 7 indicate genes responsible for this reaction that are not expressed in PMCs. Expression data show that the genes encoding enzymes catalyzing the final steps in biosynthesis of ethylene and brassinosteroids (such as OsACOs and OsDWARF/CYP85A1, D2) are not expressed in PMCs, suggesting that ethylene and brassinosteroids may not be synthesized in PMCs. However, genes encoding enzymes deactivating brassinosteroids (such as OsBASIL1, OsBASIL2) are expressed in PMCs, suggesting PMCs might inactivate the brassinosteroids produced in surrounding tissues. In addition, the brassinosteroid receptors OsBRI1 and OsBAK1 as well as ethylene signaling genes (OsETR1, OsEIN4, OsCTR1, etc.) are highly expressed in PMCs (Supplemental Table 8), indicating that PMCs might be able to respond to environmental brassinosteroids and ethylene. While in TCP, ethylene signaling genes (OsETR1, OsEIN4, OsCTR1, etc.) are no longer expressed, suggesting that TCP may not respond to ethylene.

Gibberellin signaling pathway might be active in PMCs

Figure 8 shows that genes encoding enzymes for the final cytosolic steps in gibberellin biosynthesis (OsGA20ox and OsGA3ox) are expressed in PMCs, and some of the genes for earlier synthesis pathways (such as KAO in ER steps, CPS in plastid steps), as well as the genes encoding GA deactivating enzymes are not expressed in PMCs. Thus GA might be synthesized in PMCs from intermediate products. Figure 8 also shows that all the genes in the gibberellin signaling pathway are expressed at medium to high levels in PMCs, including SLR1, encoding the only DELLA protein in rice, GID1, encoding the gibberellin receptor, GID2, which encodes an F-box subunit of the SCF E3 complex that can specifically interact with phosphorylated SLR1, and EL1, encoding the casein kinase I that phosphorylates SLR1 (Dai and Xue, 2010). But in TCP, SLR1 and GID1 are no longer expressed. These expression patterns suggest that turnover of SLR1 might be active in PMCs but not active in TCP. A SLR-like gene, SLRL2 (Os05g0574900), expressed in PMCs but not in tapetum (Hirano et al., 2008), TCP or seedlings (Supplemental Table 8). SLR1 belongs to GRAS transcription factor family. Another GRAS family transcription factor, Os11g0139600, was also expressed in PMCs but not in TCP or seedling (Supplemental Table 5). In addition, another gene that can suppress SLR1, OsSPY, is also highly expressed in PMCs. Therefore suppression of SLR1 to OsGAMYB is likely to be released, in consistent with the high expression of OsGAMYB and two OsGAMYB-like genes in PMCs. GAMYB has been reported
to regulate rice anther development (Aya et al., 2009). Combining data from Aya et al., (2009) and our microarray data, we also provide a list of 29 PMC-expressed genes (Supplemental Table 9) that contain putative MYB transcription factor recognition sites in their promoter region. These genes might be downstream genes regulated by OsGAMYBs. In conclusion, our microarray analysis suggests that gibberellin signaling may function in PMCs, consistent with a previous report that a gid1 mutant showed abnormal PMCs and failed to complete meiosis to form tetrads (Aya et al., 2009).

Combining microarray data from the microspore and the surrounding tapetum cells after PMC stages (Suwabe et al., 2008; Hirano et al., 2008), we can see that SLRI is always expressed in tapetum, and is highly to medium expressed in PMCs and tetrad, but is very low or not expressed in unicellular microspore, bi-cellular microspore or TCP. We have shown that transcriptomes of PMCs group together with tetrads, but that uni-cellular pollen to TCP form a distinct group (Fig. 5B). Since DELLA proteins have been considered as integrators of response to multiple growth-regulatory signals (Alvey and Harberd, 2005), it might be interesting to determine whether the reduction of rice DELLA expression after meiosis can be responsible for the transcriptome shift at the end of meiosis.

The 1158 PMC-preferential genes and 127 genes in PMCs but not in TCP or seedlings comprise candidates for meiosis regulators

The cell is the basic unit of life, not only structurally but also functionally. The entire transcriptome of a cell is a key determinant of identity and behavior of the cell. Here we reported the high-resolution transcriptome of an important type of cell, PMC, moving forwards to comprehensively illustrate the molecular identity of PMCs and explain the cellular behavior of PMCs at molecular level.

Our experiment provides three very different sets of microarray data, one for a cell just about to enter meiosis (PMC), one for an organism comprised of three cells (a vegetative cell and two sperm) that all have stopped cell division in preparing for double fertilization (TCP), and one for a mixed tissue containing many types of cells in various stages of mitotic cell division cycle (seedling). Although an overstatement, comparing PMCs to seedling might be considered as meiosis versus mitosis, comparing PMCs and seedlings to TCP might be considered as dividing...
cells versus non dividing cells. The meiotic-related genes should be enriched in the genes that
expressed significantly higher in PMCs than in seedlings and TCP. Supplemental Table 13
provides the whole set of genes (1158) at least four-fold higher expression in PMCs than in either
TCP or seedlings, which we defined as PMC-preferential core transcripts. The PMC preferential
core transcripts should also contain genes that have not been implicated in meiosis, which can be
candidates for further functional research.

Among the 1158 core genes, 127 genes that expressed in PMCs, but not in seedling or TCP
(Supplemental Table 5), could be more specifically meiotic related. Besides the 21 genes in
ubiquitin-mediated proteolysis pathway we already mentioned, there are 5 pentatricopeptide
repeat (PPR) motif containing proteins (3.9%), while this family only comprises 0.88% of the rice
genome. A putative histone deacetylase (HDAC, Os02g0214900) is worth mention. It is also
known as HDA703, a class I RPD3/HDA1 family HDAC, whose expression was not detected in
the leaves, stems, or roots of 7-day-old seedlings (Fu et al., 2007). Down-regulation of HDA703
by amiRNA reduced rice fertility (Hu et al., 2009). In mammals, HDAC inhibition downregulates
homologous recombination DNA repair pathways (Kachhap et al., 2010).

Our microarray data cannot tell whether these genes are expressed in tapetum cells or other cell
types during vegetative growth. For 59 out of 127 genes, the expression patterns in tapetum cells
(Suwabe et al., 2008) and in 40 vegetative cell types (Jiao et al., 2009) were reported. Supplemental Table 5 also shows that nine of them are not expressed or expressed at low level in
tapetum cells or the 40 vegetative cell types.

We also randomly selected 5 of 127 genes to check their expression pattern in rice immature
panicles at anther stages 3, 5, 7 and panicles at TCP stage, along with other organs (leaf, stem and
root) by a conventional RT-PCR assay. Our results show that none of them were expressed in
young leaf, root or stem (Supplemental Fig. 10). The expression of Os05g0484000 (a cyclin-like
F-box family gene) can only be detected in panicles at anther stage 5 (harboring PMCs). The
expression of Os08g0164000 and Os10g0484800 were not detected in immature panicles at anther
stage 3 or in mature panicle at TCP stage.

In sexual reproduction, pre-meiotic PMCs in plants are roughly equivalent to differentiating
(Kit-positive) spermatagonia in mammals. Probably due to the asynchrony in meiosis entry of
spermatagonia (Pellegrini et al., 2010), the global transcriptome data of a single type of male germ
cells at the specific stage just before entering meiosis in mammals are also lacking (Chalmel et al., 2007). The process of meiosis is highly conserved in all eukaryotes. Thus obtaining the pre-meiotic transcriptome in plants might also provide hints for mammal studies.

Materials and Methods

Plant materials and tissue collection

Rice (Oryza stativa L. ssp. japonica cv. Nipponbare) plants were grown in the greenhouse at 28±1°C and 13 h day/11 h night cycle.

According to the description in Lu et al. (2006) and Chen et al. (2005), and our own observations, we collected immature panicles at developmental stage 5 for isolation of PMCs. Features used to distinguish this stage include: 1) rice was grown for about two months, not flowering yet; 2) the distance between the last two leaf collars is within 2 cm; 3) the immature panicle length is about 2.5 cm; 4) the floret length is about 2 mm. The stage of rice anthers was further confirmed by microscopic examination of the cross sections. The cross section of stamens at this stage start to show a “butterfly” shape, and four layers of cells are visible. DAPI staining (4’, 6-diamino-2-phenylindole solution at 1µg/ml concentration with 0.2% (v/v) Tween-20) of PMCs showed one nucleus per cell without visible chromosomes (Fig. 1, A and B), while in a slightly later stage (judged by floret length 2.5-3.5mm) these cells enter meiosis, with chromosomes visible (Fig. 1, C and D). Cell walls of PMCs have little visible callose, while meiocytes deposit callose in cell walls (Chen et al., 2007). We stained stamen sections selected with decolorized aniline blue (0.1%) and found that the majority of developing microspores had little visible callose, only a small proportion showing callose between meiocytes (Supplemental Fig. 1). With all the above, we concluded that our target cells were in the pre-meiotic stage.

To isolate TCP, we collected anthers from flowering panicles and released pollen by incubation and shaking in 40% (w/v) sucrose. The TCP were then precipitated after centrifuging (300 g, 4°C).

Whole seedlings were collected one week after planting in water, and grown at 26±1°C with a 14 h day/10 h night cycle.

Sample preparation for laser microdissection

Immature panicles harboring PMCs were prepared by microwave-accelerated acetone-fixation
and paraffin-embedding method as described in Tang et al. (2006). Cross sections (8-10μm thickness) of florets were then cut on a rotary microtome (Leica Microtome, Germany). A paraffin-tape transfer system (Instrumedics, Hackensack, NJ, USA) was used to mount paraffin sections onto slides according to the manufactory protocol. The slides were de-paraffinized twice for 5 min each in pure Histoclear II and air dried, and they were ready for laser microdissection within a day.

**Laser capture microdissection (LCM)**

Homogenous PMCs were isolated using a Veritas Microdissection System (Arcturus/Molecular Devices). PMCs can be distinguished from their surrounding cells in cross sections of rice florets at the selected stage. After labeling the target cells on the monitor, a fine UV-laser beam was used to cut around the target PMCs to disconnect from surrounding cells, and then an infrared (IR) laser beam was targeted to the specific area of the thermoplastic polymer film of the transfer cap that just above the PMCs, activating the film to fuse with the target cells. By this means the target cells were captured by the cap and could then be transferred to RNA extraction buffer. Two biological replicates of PMCs, composed of around 800 cells each, were captured for RNA isolation.

**RNA extraction and amplification**

Total RNA from LCM-derived PMCs and collected TCP were extracted with PicoPure RNA isolation kit (Arcturus/Molecular Devices). The integrity of total RNA sample was evaluated on an Agilent 2100 Bioanalyzer using RNA-6000 Pico LabChips (Agilent Technologies). Only those RNA samples with a 28S:18S ribosomal RNA ratio >1 were used for further amplification (Supplemental Fig. 2A). Total RNA of seedlings was extracted directly by Trizol reagent (Chomczynski and Sacchi, 1987).

All total RNA samples were linearly amplified using a TargetAmp two-round aminoallyl-aRNA amplification kit (Epicentre Biotechnologies, Madison, WI, U.S.A.) with Superscript III and SuperScript II reverse transcriptases (Invitrogen, Carlsbad, CA, U.S.A.). For each amplification, approximately 1 ng of total RNA was used as starting material and about 8 μg amino-allyl cRNA was recovered. The quality of amplified RNA was also assessed using an Agilent 2100 Bioanalyzer, and only those with a “bell-shape” curve with peak size above 300
nucleotides (Supplemental Fig. 2B) were used for probe labeling.

**Amino-allyl cRNA labeling and microarray hybridization**

Amino-allyl cRNA were coupled with Cy3 or Cy5 mono-reactive NHS esters for 30 min in the dark according to Ambion dye-coupling protocol (Amino- Allyl MessageAmp II Kit, Ambion). After the reactions were quenched with 4 M hydroxylamine HCl, RNeasy columns (Qiagen, Valencia, CA, U.S.A.) were used to separate labeled RNAs from uncoupled dye molecules. Samples were eluted with water and quantitated on a spectrophotometer to measure sample concentration and dye incorporation. Cy3- and Cy5-labeled cRNA (1μg each) were combined, and hybridized for 17 h at 65°C in a rotating hybridization oven (60-mer Oligo Microarray Processing Protocol v 4.1, Agilent Technologies). The slides were washed at room temperature and scanned on Agilent Microarray Scanner, with scan resolution 5 μm, and PMT 100%, 10%.

**Microarray and experimental design**

The rice 4x44K oligo microarray (chip code 25152411447) (Agilent Technologies, Palo Alto, CA, USA) consists of 45,152 spots, including 1283 representing negative controls and 43,734 oligonucleotides synthesized based on the nucleotide sequence and full-length cDNA data of the Rice Annotation Project (RAP). The same microarray chip was used by Shimono et al. (2007) and Suwabe et al. (2008). A total 43,734 60mer probes represent 29,008 distinct rice genes. Among the 28,840 protein-coding genes, 22,532 have single sequence probes, 4619 have 2 sequence probes each, 1235 have 3 sequences probes each, 330 have 4 sequences each, 92 have 5 sequences each, 16 have 6 sequences each, 11 have 7 sequences each, 3 have 8 sequences each, and 2 have 9 sequences each.

Four microarrays of two colors were hybridized by ShanghaiBio Corp. Two independent biological replicates for each cell type, PMC and TCP were labeled with Cy3. A mixture of two biological samples of seedlings was labeled with Cy5 and used as a common reference in all the hybridizations. The common reference was prepared once and aliquoted identically for all the experiments. An additional independent biological sample of seedling was also labeled with Cy5 and hybridized with the same array.
Microarray feature extraction and pre-processing

TIFF Images were inspected visually to ensure the number of saturated spots on each array range from 2-10, and probe features were extracted with Agilent Feature Extraction Software (Feature Extraction software 9.5.3) using configuration suggested by Agilent (Protocol for use with Agilent Gene Expression oligo microarrays Version 5.7, March 2008).

The quality of the microarray data was ensured by the following criteria: 1) biological replicates must have a correlation r≥0.9; (The correlation r between TCP1 and TCP2 is 0.9698; between PMC1 and PMC2 is 0.9037; between additional biological rep of seedling and the common reference seedling is 0.9443) and 2) the irregular spots were further removed when their median intensities were below the local background or replicated probes having a large variation, i.e. median intensities between replicated probes having >5 fold difference.

Among different probes representing the same gene, the hybridization intensities of probes whose target sequence positions are within 0 to 300 nucleotides downstream of the end of the open reading frame were similar (less than two fold difference), while intensities of probes targeted to other regions were significantly reduced (Supplemental Fig. 11). The intensity reduction is probably because mRNA fragments far upstream from the polyA tail were not preserved as well after RNA amplification with oligo-dT primers. For genes with multiple sequence probes, we manually chose the probe with targeted sequence position closest to 200 nts downstream of the translation termination codon to represent the gene. We surveyed genes with single probes; roughly 90% of them are targeted with 0-300 nts downstream of translation termination codon. Because we chose a 4-fold difference of intensity as a threshold for significantly different expression in addition to statistical tests, the intensity difference caused by probe position difference (< 2-fold) should not affect our results.

Data normalization and statistical analysis

Data normalization is performed without a background subtraction step as well as with background subtraction. In general, the results are similar either with or without background subtraction. We chose to present the results without background in the tables and figures, and present the results with background subtraction in supplemental tables. Following the recommendation in Zahurak et al., (2007), we slightly prefer not to subtract local background for
two reasons: 1) we observed a slightly larger noise to signal ratio for low intensity probes when local background is subtracted; 2) the correlation $r$ between biological replicates was decreased after background subtraction.

Although loess normalization performs well in many cases using two-color microarray, it is not appropriate for processing microarray data from laser-microdissected cells. Intensity-based normalization is more suitable (‘t Hoen et al., 2004), because we found data distortion after using loess normalization. Jiao et al. (2009) also didn’t use Loess normalization in analyzing the cell-type specific transcriptome data for 40 cell types. Loess normalization requires the assumption that most of the genes are constantly expressed over the range of intensities or that the numbers of up and down regulated genes over the intensity range are equal (Yang et al., 2002; Zahurak et al., 2007). Because only a small proportion of genes are expressed at similar levels between PMC, TCP and seedling, and a significant number of genes are differentially expressed between them, an alternative normalization is performed in our study. We here give rationale to a simple method which basically using MS (absolute intensity value) to represent individual gene expressional level given baseline of microarray (provided by more than 1000 negative control probes’ intensity value) and the maximum distribution of intensity ratio close to 1:1 for microarrays using identical amount of cRNA for hybridization. The maximum distribution of intensity ratio close to 1:1 indicates although significant difference of gene expression among LCM samples, there are still the most abundant distribution of gene expression which are not changed genes (Supplemental Fig. 12).

To meet the needs for profiling differential expressed gene among the tissues, we perform a two-step between array normalization: 1, baseline transformation based on the negative control probes; 2, scale normalization based on the positively "house-keeping" genes (Supplemental Table 10) that constantly expressed among the tissues. The baselines for PMC, TCP and seedling are similar (Supplemental Fig. 13). The saturated intensities and the number of probes at saturated level for PMC, TCP and seedling are similar (Supplemental Table 11). The routine "house-keeping" genes such as actin, tubulin or histone are different between PMC, TCP, and seeding, but the total intensities for the 101 genes encoding enzymes in the oxidative phosphorylation pathway (energy metabolism) are similar (Supplemental Table 15). This supports that the core energy supply is maintained at similar level among different types of living cells. To
assess positively expressed gene, we defined the 95th percentile of the intensities from the negative controls as the threshold for significant gene expression. This cut-off is a result from balancing the number of false positive versus missing out truly expressed genes. Gene replicates with at least one above the threshold was examined with Student's T-test, and a p<0.001 was considered significantly expressed. Among significantly expressed genes, the top 25 percentile are considered highly expressed, the middle 50 percentile considered intermedianly expressed, and the bottom 25 lowly expressed.

We performed a second analysis using the method similar to that in Nakazono et al., 2003. The results are included in the supplemental materials. We found the second method produced almost identical results to our original method.

GeneSpring GX9.0 (Agilent) was used for microarray analysis. Microsoft excel 2003 and software R was used for statistical tests. Online genomic analysis tools of Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database was used for finding overlapping genes and enrichment analysis were done by chi square test. Our data were deposit in GEO under accession GSE17098.

**Validation of microarray data by quantitative RT-PCR, In situ hybridization and a conventional RT-PCR**

We performed quantitative real-time PCR reaction of reverse transcribed RNA with SYBR Green I detection on an iCycler (Bio-Rad) as described by Tang et al. (2006).

To measure the absolute mRNA copy number in each laser captured PMCs, we inserted the genomic DNA fragments of Os09g0480300 and Os03g0650400 (AM1-like) into pTG-19 vector separately.

For in situ hybridization, a 563 bp AM1-like (Os03g0650400) DNA fragment, a 494 bp SKP1-like (Os07g0409500) DNA fragment and a 561 bp SPL-like (Os01g0212500) DNA fragment were separately inserted into pBluescript SK+ vector for RNA probe synthesis. The antisense and sense RNA probes were synthesized by in vitro transcription using T7 and T3 RNA polymerase, respectively, using DIG RNA Labeling Mixture (Roche). We used microwave tissue processing technique for the fixation and embedding of stage 5 anthers (Schichnes et al., 1999). In situ
hybridization experiments using 10 μm sections were carried out as described (Cox and Goldberg, 1988; Langdale, 1993).

All the primers are listed in Supplemental Table 12.

Acknowledgements

We thank Drs. Jon Duvick, Sheila McCormick, Hong Ma, Zhu-Kuan Cheng, Ying-Xiang Wang and Shunong Bai for advice on manuscript preparation, Drs. Dabing Zhang and Wanqi Liang for advice on PMC staging, Drs. Min-Ping Qian, Lin Ye, Yu Wang and Xin-Guang Zhu for help with statistical analysis of microarray data.

References


Dai C, Xue HW (2010) Rice early flowering1, a CKI, phosphorylates DELLA protein SLR1 to negatively regulate gibberellin signaling. EMBO J 29: 1916-1927


Prasad K, Parameswaran S, Vijayraghavan U (2005) OsMADS1, a rice MADS-box factor, controls differentiation of specific cell types in the lemma and palea and is an early-acting regulator of inner floral organs. Plant J 43: 915-928


**Figure Legend**

Figure 1. Laser microdissection of rice pollen mother cells (PMCs). A-F, DAPI-stained anthers and microspore/pollen of rice at various stages. A and B, Cross-sections of anthers at stage 5 collected from ~2 mm florets, used as sample for laser microdissection isolation; C and D, Microspore during meiosis released from anthers collected from 3-4 mm florets; E, Microspore at tetrad stage released from anthers collected from ~5 mm florets; F, Tricellular pollen (TCP) from ~6 mm florets. G-I, Isolation of PMCs from cross-sections of anthers at stage 5 by a Veritas Laser Microdissection System. G and H, A cross-section of an anther on the slide before (G) and after (H) laser capture, orange arrow points the targeted cell for capture; I. PMCs captured in the cap from the area showed in G. Scale bar=50μm.

Figure 2. Validation of microarray results by real-time PCR. The values indicate the logarithm of expression ratio to the reference gene (Os03g0268000). Each value represents the mean of 3 biological replicates. Error bars represent standard deviation. Stars signify a significant difference
between real-time PCR result and microarray data at the level of P<0.05. The two lines divide the gene expression levels into three groups: high, median and low level.

Figure 3. In situ hybridization of *SKP1-like* (Os07g0409500), *Am1-like* (Os03g0650400) and *SPL-like* (Os01g0212500) in stage 5 anthers. A. Longitudinal section of rice florets hybridized with *SKP1-like* antisense RNA probe; B. Cross section of rice anthers hybridized with *SKP1-like* antisense RNA probe; C. Longitudinal section of rice florets hybridized with *SKP1-like* sense RNA probe; D. A cross section hybridized with *AM1-like* antisense RNA probe; E. A cross-section hybridized with *AM1-like* sense RNA probe; F. A cross section hybridized with *SPL-like* antisense RNA probe; G. A cross section hybridized with *SPL-like* sense RNA probe. Scale bar=50μm.

Figure 4. Summary of global gene expression levels. A. Density plot of the expression of 29008 distinct genes in pollen mother cells. The broken lines roughly indicate the intensity (log transformed) ranges of the genes not expressed (below 7.5), no or very low expressed (7.5 to 9), low expressed (9 to 11), medium expressed (11-15) and high expressed (above 15). See Methods and Material section for detail statistical analysis. B. Percentages of genes expressed or not expressed in pollen mother cells (PMCs), tricellular pollen (TCP) and seedlings mixed tissues (seedling). C. Venn diagram illustrating overlap in gene expression among PMCs, TCP and seedlings.

Figure 5. Principal component analysis (PCA) for gene expression data of 9 different cell types and corresponding hierarchical clustering for these cell types. A. Spatial distribution of the first three principal components. B. Hierarchical clustering dendrogram, using Pearson correlation coefficients and Single Linkage method.

Figure 6. A framework of meiotic recombination pathways genes in rice with expression data in PMCs, TCP and seedlings. The microarray data for all the listed genes were organized the same as the top gene in the left panel. Left panel lists the genes have been functionally characterized in rice. Right panel lists the putative meiosis-related genes supported/suggested by our microarray data.
Figure 7. Plant hormone biosynthesis pathways in rice from KEGG website with colored arrows indicating the gene expression in PMCs of responsible enzymes.

Figure 8. Expression of genes in the gibberellin biosynthetic and signaling pathways. The biosynthetic pathway is modified from Yamaguchi, 2008. The signaling pathway is modified from Ueguchi-Tanaka et al., 2007. The intensities are directly labeled besides the RAP ID in the order as labeled for GID2. The color scale for intensities as indicated at left-bottom in the figure.

Table title
Table 1. KEGG pathways enriched in PMC-, TCP-, seedling- expressed/preferentially expressed genes

Supplemental Figures and Tables:
Supplemental Figure 1. Aniline blue staining of rice anthers at the stage for PMCs isolation.
Supplemental Figure 2. RNA quality assessment.
Supplemental Figure 3. Calculation of mRNA copy number.
Supplemental Figure 4. Rice genes in base excision repair pathways with expression data.
Supplemental Figure 5. Protein sequence alignments for Am1 and SPL.
Supplemental Figure 6. Expression data of rice MADS transcription factors with a phylogenic view.
Supplemental Figure 7. Percentage of highly expressed transcription factors.
Supplemental Figure 8. MicroRNA pathways in PMCs.
Supplemental Figure 9. Overview of plant hormone biosynthesis pathways in rice TCP (A), seedling (B) and tapetum (C).
Supplemental Figure 10. RT-PCR results for 5 genes in rice panicles at different developmental stages and other organs (young rice leaf, stem and root).
Supplemental Figure 11. Probe position effects on hybridization Intensities.
Supplemental Figure 12. Density plot of Log2 intensity ratios of seedling vs. PMC and TCP vs. PMC.
Supplemental Figure 13. Density plot of Log2 intensity for all the negative probes as the baseline.
Supplemental Table 1. Microarray data for the 26 genes involved in rice early stamen development (Lu et al., 2006).

Supplemental Table 2. Detailed results of KEGG pathway enrichment analysis.

Supplemental Table 3. Description of twelve known meiotic genes.

Supplemental Table 4. Microarray data for the genes similar to meiotic genes.

Supplemental Table 5. Microarray data of 127 genes that expressed in PMCs, but not in TCP or seedling.

Supplemental Table 6. Microarray data for 34 rice SET domain-containing proteins.

Supplemental Table 7. Transcription factors and regulators expression summary.

Supplemental Table 8. Microarray data for 288 genes in plant hormone biosynthesis and signaling pathways.

Supplemental Table 9. 29 PMC-expressed genes containing putative MYB binding motif in the promoter region.

Supplemental Table 10. 20 housekeeping genes that used for normalization.

Supplemental Table 11. Saturated intensity values and number of probes at saturated level.

Supplemental Table 12. Primers sequences.

Supplemental Table 13. Microarray data for 1158 PMC-preferential genes.

Supplemental Table 14. Microarray data for 1108 rice kinase genes (Genes that expressed >4-fold higher in PMC than in TCP and seedling are indicated by * after RAP ID).

Supplemental Table 15. Microarray data for genes in oxidative phosphorylation pathway.

Supplemental Table 16. Literatures related to plant hormones function in anther or pollen development.
Table 1. KEGG pathways enriched in PMC-, TCP-, seedling-expressed/preferentially expressed genes

<table>
<thead>
<tr>
<th>KEGG pathways that have 10 or more genes on this microarray</th>
<th>PMC expressed</th>
<th>TCP expressed</th>
<th>Seedling expressed</th>
<th>PMC preferentially expressed</th>
<th>TCP preferentially expressed</th>
<th>Seedling preferentially expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of genes</td>
<td>p^a</td>
<td># of genes</td>
<td>p^a</td>
<td># of genes</td>
<td>p^a</td>
</tr>
<tr>
<td>1. Metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1 Carbohydrate Metabolism (19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Starch and sucrose metabolism</td>
<td>68</td>
<td>47</td>
<td>37</td>
<td>55</td>
<td>2</td>
<td>11&lt;0.0001</td>
</tr>
<tr>
<td>Pentose &amp; glucuronate conversions</td>
<td>14</td>
<td>11</td>
<td>9</td>
<td>1.00</td>
<td>13</td>
<td>0.98</td>
</tr>
<tr>
<td>Citrate cycle (TCA cycle)</td>
<td>34</td>
<td>31</td>
<td>0.40</td>
<td>30 0.0035</td>
<td>32</td>
<td>0.75</td>
</tr>
<tr>
<td>Glycolylate &amp; dicarboxylate metabolism</td>
<td>31</td>
<td>24</td>
<td>0.68</td>
<td>27</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.2 Energy Metabolism (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carbon fixation</td>
<td>64</td>
<td>52</td>
<td>0.20</td>
<td>60</td>
<td>0.75</td>
<td>2 0.92</td>
</tr>
<tr>
<td>Methane metabolism</td>
<td>28</td>
<td>13</td>
<td>10</td>
<td>23</td>
<td>2 0.19</td>
<td>1 0.80</td>
</tr>
<tr>
<td>Photosynthesis</td>
<td>32</td>
<td>26</td>
<td>10</td>
<td>31</td>
<td>0.36</td>
<td>0 0.0001</td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td>90</td>
<td>83</td>
<td>0.10</td>
<td>1.76&lt;0.0001</td>
<td>83</td>
<td>2 0.40</td>
</tr>
<tr>
<td>1.3 Lipid Metabolism (8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alpha-Linolenic acid metabolism</td>
<td>26</td>
<td>11</td>
<td>10</td>
<td>24</td>
<td>0</td>
<td>0 0.0002</td>
</tr>
<tr>
<td>Linoleic acid metabolism</td>
<td>14</td>
<td>2</td>
<td>5</td>
<td>11</td>
<td>0</td>
<td>6 0.0054</td>
</tr>
<tr>
<td>Ether lipid metabolism</td>
<td>13</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>0</td>
<td>2 0.0057</td>
</tr>
<tr>
<td>1.4 Nucleotide Metabolism (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5 Amino Acid Metabolism (16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylalanine metabolism</td>
<td>28</td>
<td>13</td>
<td>8</td>
<td>24</td>
<td>2 1.91</td>
<td>1 0.81</td>
</tr>
<tr>
<td>1.6 Metabolism of Other Amino Acids (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxyoanmino acid metabolism</td>
<td>15</td>
<td>12</td>
<td>7</td>
<td>15</td>
<td>0 0.28</td>
<td>0 0.0001</td>
</tr>
<tr>
<td>1.7 Glycan Biosynthesis and Metabolism (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.8 Metabolism of cofactors and vitamins (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Porphyrin and chlorophyll metabolism</td>
<td>26</td>
<td>25</td>
<td>0.14</td>
<td>14</td>
<td>25 0.50</td>
<td>0 0.0001</td>
</tr>
<tr>
<td>1.10 Biosynthesis of secondary metabolites (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carotenoid biosynthesis</td>
<td>17</td>
<td>15</td>
<td>0.80</td>
<td>8</td>
<td>16 0.82</td>
<td>2 0.031</td>
</tr>
<tr>
<td>Phenylpropanoid biosynthesis</td>
<td>40</td>
<td>15</td>
<td>5</td>
<td>32</td>
<td>0 0.45</td>
<td></td>
</tr>
<tr>
<td>1.11 Xenobiotics Biodegradation (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Genetic Information Processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1 Transcription (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basal transcription factors</td>
<td>21</td>
<td>21</td>
<td>0.066</td>
<td>20 0.0030</td>
<td>21 0.20</td>
<td>2 0.08</td>
</tr>
<tr>
<td>2.2 Translation (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribosome</td>
<td>189</td>
<td>184 0.0001</td>
<td>153&lt;0.0001</td>
<td>183 0.029</td>
<td>2 0</td>
<td>0 20</td>
</tr>
<tr>
<td>2.3 Folding, sorting and degradation (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteasome</td>
<td>46</td>
<td>45</td>
<td>0.022</td>
<td>41 0.0004</td>
<td>45</td>
<td>0.18</td>
</tr>
<tr>
<td>Ubiquitin mediated proteolysis</td>
<td>46</td>
<td>44</td>
<td>0.062</td>
<td>40 0.0013</td>
<td>45</td>
<td>0.18</td>
</tr>
<tr>
<td>2.4 Replication and Repair (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA replication</td>
<td>37</td>
<td>36</td>
<td>0.050</td>
<td>25 0.67</td>
<td>36 0.28</td>
<td>4 0.0049</td>
</tr>
<tr>
<td>Homologous recombination</td>
<td>23</td>
<td>23</td>
<td>0.055</td>
<td>14</td>
<td>22 0.59</td>
<td>3 0.0044</td>
</tr>
<tr>
<td>Base excision repair</td>
<td>34</td>
<td>31</td>
<td>0.40</td>
<td>18</td>
<td>31</td>
<td>6&lt;0.0001</td>
</tr>
<tr>
<td>Mismatch repair</td>
<td>26</td>
<td>25</td>
<td>0.14</td>
<td>18 0.59</td>
<td>26 0.15</td>
<td>4 0.0002</td>
</tr>
<tr>
<td>Nucleotide excision repair</td>
<td>38</td>
<td>36</td>
<td>0.13</td>
<td>31 0.029</td>
<td>37 0.27</td>
<td>4 0.0060</td>
</tr>
<tr>
<td>3. Environmental Information Processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1 Membrane Transport (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2 Signal transduction (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: a. “Preferentially expressed” is defined as the intensities in the target sample is significantly higher than in anyone of the other samples (p<0.001) and the fold change is greater than 4; b. the p values of Chi square test for enriched pathways are listed, depleted ones are omitted. p<0.01 is bold, considered as significantly enriched; C. The number in parentheses is the number of pathways included in this category, while names of pathway provided only if there is at least enriched in one case. See supplemental table 2 for a full list.
Figure 1. Laser microdissection of rice pollen mother cells (PMCs). A-F, DAPI-stained anthers and microspore/pollen of rice at various stages. A and B, Cross-sections of anthers at stage 5 collected from ~2 mm florets, used as sample for laser microdissection isolation; C and D, Microspore during meiosis released from anthers collected from 3-4 mm florets; E, Microspore at tetrad stage released from anthers collected from ~5 mm florets; F, Tricellular pollen (TCP) from ~6 mm florets, used as TCP sample for microarray. G-I, Isolation of PMCs from cross-sections of anthers at stage 5 by a Veritas Laser Microdissection System. G and H, A cross-section of an anther on the slide before (G) and after (H) laser capture, orange arrow points the targeted cell for capture; I. PMCs captured in the cap from the area showed in G. Scale bar=50 μm.
Figure 2. Validation of microarray results by real-time PCR. The values indicate the logarithm of expression ratio to the reference gene (Os03g0268000). Each value represents the mean of 3 biological replicates. Error bars represent standard deviation. Stars signify a significant difference between real-time PCR result and microarray data at the level of P<0.05. The two lines divide the gene expression levels into three groups: high, median and low level.
Figure 4. Summary of global gene expression levels. A. Density plot of the expression of 29008 distinct genes in pollen mother cells. The broken lines roughly indicate the intensity (log transformed) ranges of the genes not expressed (below 7.5), no or very low expressed (7.5 to 9), low expressed (9 to 11), medium expressed (11-15) and high expressed (above 15). See Methods and Material section for detail statistical analysis. B. Percentages of genes expressed or not expressed in pollen mother cells (PMC), tricellular pollen (TCP) and seedling mixed tissues (seedling). C. Venn diagram illustrating overlap in gene expression among PMC, TCP and seedling.
Figure 5. Principal component analysis (PCA) for gene expression data of 9 different cell types and corresponding hierarchical clustering for these cell types. A. Spatial distribution of the first three principal components. B. Hierarchical clustering dendrogram, using Pearson correlation coefficients and Single Linkage method.
<table>
<thead>
<tr>
<th>Known meiotic genes in rice</th>
<th>Putative meiotic genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Meiotic recombination (DSB formation and repair)</td>
<td>Long-patch base-excision repair (BER)</td>
</tr>
<tr>
<td>Double-strand break formation</td>
<td></td>
</tr>
<tr>
<td>End processing</td>
<td></td>
</tr>
<tr>
<td>Loading the recombination machinery</td>
<td>Nucleotide-excision repair (NER)</td>
</tr>
<tr>
<td>Strand invasion</td>
<td>DNA replication</td>
</tr>
<tr>
<td>Interfering crossovers</td>
<td>DNA ligase I common for BER, NER and DNA replication</td>
</tr>
<tr>
<td>Non-interfering crossovers</td>
<td></td>
</tr>
</tbody>
</table>

2. Meiotic chromosome structure control
Sister chromatid cohesion
Homologous chromosome pairing & synapsis

3. Meiotic cell cycle control

4. Anther development regulation (context for meiosis)
Transcription factors
AGL6-like clade
AGL2-like clade
Receptor kinases
Ligand for receptor kinases

Figure 6. A framework of meiotic recombination pathways genes in rice with expression data in PMCs, TCP and seedlings. The microarray data for all the listed genes were organized the same as the top gene in the left panel. Left panel lists the genes have been functionally characterized in rice. Right panel lists the putative meiosis-related genes supported/suggested by our microarray data.

Downloaded from on December 30, 2017 - Published by www.plantphysiol.org
Copyright © 2010 American Society of Plant Biologists. All rights reserved.
Figure 7. Plant hormone biosynthesis pathways in rice from KEGG website with colored arrows indicating the gene expression in PMCs of responsible enzymes.
Figure 8. Expression of genes in the gibberellin biosynthetic and signaling pathways. The biosynthetic pathway is modified from Yamaguchi, 2008. The signaling pathway is modified from Ueguchi-Tanaka et al., 2007. The intensities are directly labeled besides the RAP ID in the order as labeled for GID2. The color scale for intensities as indicated at left-bottom in the figure.