The Third International Symposium on Optimization and Systems Biology (OSB'09)
Zhangjiajie, China, September 20–22, 2009

Analysis of Transcriptome of Rice Anther

Wenjuan Zhang[1] Wei-Hua Tang[2] Ji-Guang Wang[3]
Zhi-Ping Liu[1] Xing-Ming Zhao[2,4,*] Luonan Chen[1,4]

[1] Department of Electrical Engineering and Electronics, Osaka Sangyo University, Osaka 574-8530, Japan
[2] Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China
[3] Academy of Mathematics and Systems Science, CAS, Beijing 100080, China
[4] Institute of Systems Biology, Shanghai University, Shanghai 200444, China

Abstract Important phenomena related to rice, such as Cytoplasmic male sterility (CMS), self-incompatibility and sex determination, are involved in sexual reproduction. For rice, the development process of male gametophyte and pollen contains several stages, including pre-meiotic anther (PMA), meiosis (MEI), tetrad (TET), uninuclear (UN) microspore, bicellular (BC) pollen, tricellular (TC) pollen, MEI, TET, and UN tapetum. With the 44K rice microarray chip and RNAs derived from LM-separated microspore/pollen and tapetum cells, in this paper, we identified the expression patterns involved in different stages. Furthermore, we also detected the enriched promoter motifs and enriched functions for each cluster, which are expected to provide biological insights into developmental procedure of male gametophyte and pollen.

Abbreviations CMS, Cytoplasmic male sterility; PMA, pre-meiotic anther; MEI, meiosis; TET, tetrad; UN, uninuclear; BC, bicellular; TC, tricellular; LM, Laser microdissection; MCL , Markov Cluster; std, standard deviation; GO, Gene Ontology; BiNGO, Biological Networks Gene Ontology tool.

1 Background

Rice is one of the most important crops in the world. Important phenomena related to rice, such as Cytoplasmic male sterility (CMS), self-incompatibility and sex determination, are observed in sexual reproduction, which includes a key phase for producing male and female gametes[1]. The male gametophyte, i.e. the pollen, develops in the anther of flowering plants [2]. The microspore/pollen, endothecium and outer tapetum are differentiated from L2 layer in the anther tissues, which consist of three layers, i.e. L1, L2 and L3. However, gene functions in pollen and tapetum are distinct [3]. Pollen mother cells undergo meiosis to release microspores in the anther locule. These microspores then mature into pollen grains through cell division and

* Corresponding author: Xing-Ming Zhao, xm_zhao@shu.edu.cn

formation of complex pollen wall. Through an asymmetric mitosis, the uninuclear microspore divides into a larger vegetative cell and a smaller generative cell that consist of bicellular pollen. Sequentially, the generative cell undergoes a second mitosis to form two sperm cells in rice. Tapetum cells function as nurse cells during pollen maturation, where the tapetum makes contribution to pollen development by secreting nutrients and other secondary metabolites. The tapetum provides materials for pollen wall formation, and then degrades in the later stages of pollen development [2,3].

The important biological events described above involve complex gene regulation in both gametophytic and sporophytic tissues during anther development. Therefore, it is necessary to analyze the transcriptomes of microspore/pollen and tapetum, which can help to understand the development process of anther [2]. Laser microdissection (LM) is a powerful tool for isolating specific cells from complicated tissues, and can be applied to extract purified transcriptomes. In this study, with the 44K rice microarray chip and RNAs derived from LM-separated microspore/pollen and tapetum cells, the expression patterns involved in different stages were identified by utilizing transcriptomes of microspore/pollen and tapetum. Furthermore, enriched promoter motifs were predicted and enriched functions were detected, which can provide biological insights into developmental mechanisms of the male gametophyte and pollen.

2 Data and methods

2.1 Data Source

LM was applied for isolating specific cells from cross-section of the developing rice anthers of various stages. The 44K-microarray chip and amplified labeled RNAs isolated from microspore/pollen and tapetum cells were used for Microarray experiments. The stages studied in this work include PMA, MEI, TET, UN microspore, BC pollen, TC pollen, and MEI, TET, and UN tapetum. Microarray experiments for each stage of the last eight stages were replicated at least three times. We downloaded a total of 31 samples of microarray data from the supplementary files of research works[1-3].

2.2 Cluster analysis

To identify the co-expression genes for further analysis, Markov Cluster (MCL) Algorithm was performed with MCL software (http://www.micans.org/mcl) [4]. MCL is a fast and scalable unsupervised cluster algorithm for graphs based on simulation of (stochastic) flow in graphs. The number of clusters is not dealt with in an arbitrary manner, but rather by strong internal logic. MCL breakthroughs the limitation of hierarchical cluster analysis, such as arbitrariness in the choice of threshold to cut the tree, and has been shown to outperform hierarchical cluster analysis [4,5]. The six power of the Pearson's correlation coefficient was used as similarity score, and the threshold was set at 0.9.

2.3 GO function analysis

Gene Ontology (GO, http://www.geneontology.org) is a widely used gene function categorization system. GO terms that are significantly overrepresented in each cluster were determined by the Biological Networks Gene Ontology tool (BiNGO) [6]. Bingo 2.3 is implemented as a plug-in for Cytoscape. The Oryza sativa annotation file (gene_association.gramene_oryza, Version 1.53. Date: 03/05/2009) was downloaded from http://www.geneontology.org. Hypergeometric distribution was used to detect over-represented GO categories in each cluster comparing to the whole genome. Terms with Hypergeometric test p-values less than 0.05 were taken as significant ones.

2.4 Promoter sequences

Promoter DNA sequences were got from ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.0/. Each promoter sequence was taken as 1000 nt upstream from the transcription start site (TSS).

2.5 Enriched motifs analysis

To identify cis-elements of the differentially co-expressed genes, AlignACE [7] was used to identify motifs in the promoter DNA sequences of each cluster. The length of promoter motifs was set at 6 and 7 nt. For the motifs discovered by AlignACE, we performed STAMP to see whether there are matched motifs in PLACE (http://www.dna.affrc.go.jp/PLACE/), and it is taken as a match if the alignment score E-values is less than 1×10^{-5} and there is no more than one base mismatched and it is significant (p-value<0.05) compared to the enrichment of the motifs in all genes [8,9].

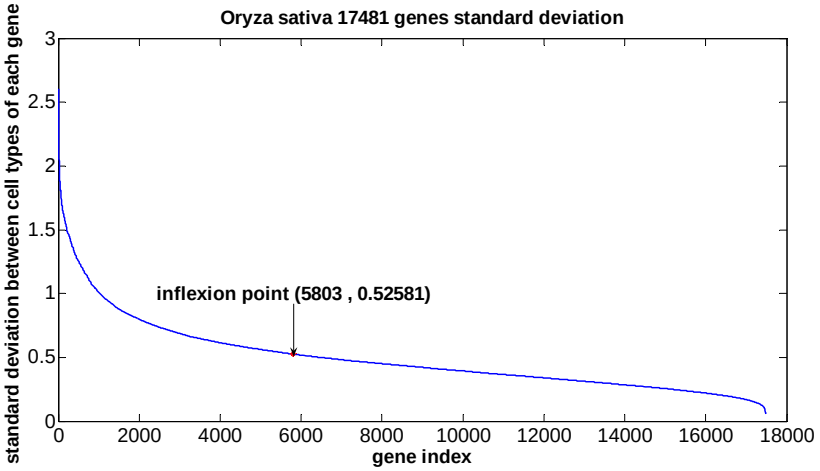

Figure 1: The distribution of expression standard deviation for 17481 genes from rice

3 Results and discussion

3.1 Identifying expression patterns during rice sexual reproduction development

There are 27789 probes and 31 samples in the Microarray data in total. The expression values were log-transformed with base 2 first. The expression value was replaced by the mean value if there are replicated microarray experiments for one sample. The expression values were normalized by subtracting the mean and then divided by the standard deviation. If there are replicated probes for the same gene, the largest expression value is regarded as the gene's expression value. Consequently, 19902 genes were retained in total, where 17481 genes have annotations of upstream 1k sequences, which were treated as the promoter regions of corresponding genes.

Based on the hypothesis that differentially expressed genes through sexual reproduction play a very important role in rice reproductive development, the differentially expressed genes were identified. The 17481 genes were ranked with standard deviation (std) of gene expression data during the nine sexual reproduction stages in descending order, and a std curve was drawn and the inflexion point of the line was identified as shown in Fig. 1. The first 5803 genes with the higher standard deviation before the inflexion point were selected. The differentially co-expressed genes were clustered into different groups with Markov Cluster Algorithm (MCL). As a result, 780 clusters were generated and 140 clusters with size larger than 4 were chosen for further analysis. In addition, we investigated whether the genes within the same group are functionally related with function terms from Gene Ontology. Table 1 lists several clusters and their corresponding enriched functional terms. It can be seen that the genes belonging to the same cluster indeed share similar functions.

Table 1: Enriched GO terms for some clusters.

Cluster ID	Number of genes	Enriched functional terms	P-value
463	184	cytoplasmic membrane-bounded vesicle	1.25E-11
84	5	membrane	3.40E-02
155	14		4.30E-02
507	7		3.40E-02
608	10		4.66E-02
125	37	plastid	5.66E-03
364	7		1.06E-04
429	19		4.94E-02

Right now, there are many rice genes that have been annotated. The relationship between expression patterns and gene functions was investigated. We focused on ten function categories that are likely related to anther development, i.e., (a) transcription, (b) translation, (c) gibberellin metabolic process, (d) calcium ion binding, (e) boron

transport, (f) membrane, (g) cytoplasmic membrane-bounded vesicle, (h) plastid, (i) tapetal layer development, and (j) programmed cell death. The expression profiles of genes annotated with above functional terms from different clusters were shown in Fig. 2.

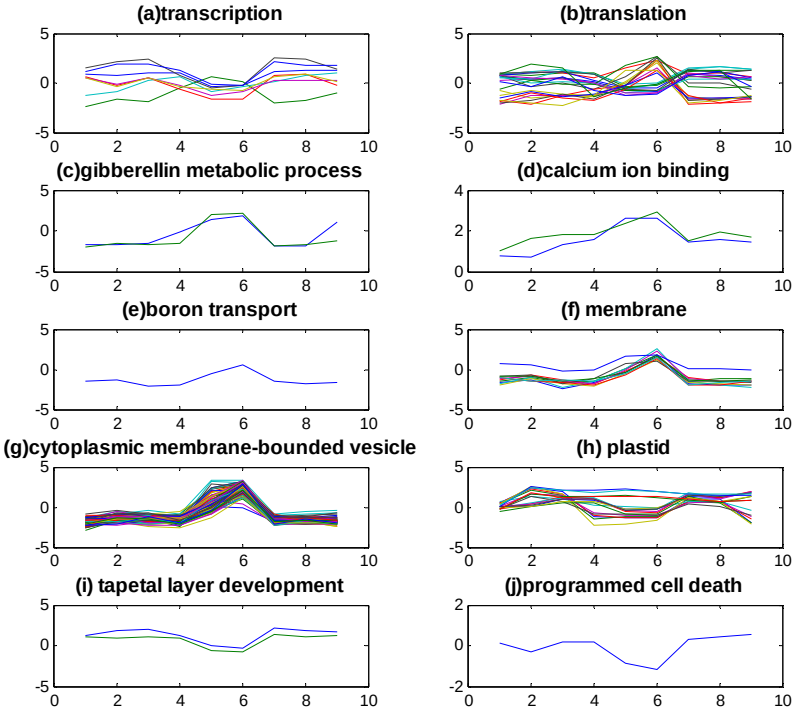

Figure 2: Differential expression patterns of significant GO terms in rice sexual reproduction development. X-axis: nine stages in rice sexual reproduction development, involving PMA, MEI, TET, UN microspore, BC pollen, TC pollen, MEI, TET, and UN tapetum in order. Y-axis: log-transformed and normalized expression data value. Ten GO terms: (a) transcription, (b) translation, (c) gibberellin metabolic process, (d) calcium ion binding, (e) boron transport, (f) membrane, (g) cytoplasmic membrane-bounded vesicle, (h) plastid, (i) tapetal layer development, (j) programmed cell death.

Transcription-related genes showed high expression in microspore and tapetum at MEI to UN stages. And there were two characteristics of the expression profiles in translation-related genes including synchronous MEI to UN expression in microspore and tapetum, and dominant expression in BC and TC pollen. This indicates that regulation of transcription-related genes might function at three stages

of MEI, TET, and UN. There are a part of translation-related genes acting at the same stages as transcription-related genes, and these gene products are later used for microsporogenesis. The translation-related genes acting at BC and TC may be used for the genesis of germocyte and prepare for the ensuing pollination and fertilization.

Genes related to gibberellin metabolic process, calcium ion binding and boron transport were notably expressed in BC and TC stages. During the maturation of anther tissues, important biological events for the male gamete such as pollen germination and pollen tube elongation are related with the pollen-specific genes whose transcripts are present in mature pollen grains and are translated subsequently [3]. This is consistent with the expression patterns found in our study. Genes related to membrane and cytoplasmic membrane-bounded vesicle were dominantly expressed at BC and TC stages, and these genes are used for pollen tube elongation. Plastid-related genes were characterized by high expression in MEI to TET microspores and tapetum, which is consistent with the phenomena observed in literature that cellular organelles such as endoplasmic reticulum, small vacuoles, plastid, and mitochondria are increasing gradually during pollen mother cell undergoing meiosis to tetrad [10]. Genes related to tapetal layer development were highly expressed in MEI tapetum, indicating that tapetum may begin to differentiate from the MEI stage. Programmed cell death-related genes have notable expression in UN tapetum. These genes might be transcribed in UN tapetum but translated in ensuing BC pollen and lead to tapetum degeneration.

Through analysis of gene expression profiles, it can be found that there are indeed specific expression patterns for distinct development stages in anther. Furthermore, gene function enrichment analysis demonstrates that these expression patterns are really related to the corresponding development stages.

3.2 Promoter motif prediction for clusters related to anther development

In order to understand the gene regulatory mechanisms underlying rice sexual reproduction, 6 and 7 nt length motifs in the promoter regions of genes were predicted by using AlignACE. The enriched motifs were respectively identified for the clusters related to anther development. For motifs discovered by AlignACE, we further performed STAMP to query best-matching motifs from PLACE, a database of plant cis-acting regulatory DNA elements. The motifs were treated as enriched if they satisfied three conditions: (1) match score E-value less than 1×10^{-5} ; (2) no more than one base mismatched; (3) it is significant (p-value<0.05) compared to the enrichment of the motifs in all genes. The over-represented motifs for GO category enriched clusters can be found in Table 2.

Although there may be false positives in the predicted promoter motifs, the identified motifs provide insights into regulatory mechanisms underlying rice sexual reproduction. In the further study, we will detect the genes containing the motifs and further construct a gene regulatory network to understand the regulation relationship during rice sexual reproduction more deeply.

Table 2: Over-represented motifs for GO enriched clusters

GO ID	Description	AlignACE Motif	Motif sequence in PLACE	Factor or site name	p-value
a	transcription	TTTTTCT	GAAAAA	GT1GMSCAM4	2.30E-02
		TTTTTT	TTTTTTCC	PYRIMIDINEBOXHVEPB1	2.50E-02
		TATATATA	TGTATATAT	SORLREP3AT	2.10E-02
b	translation	GCGGCK	AGCCGCC	AGCBOXNPGLB	3.00E-03
		TTTCCC	GCGGGAAA	E2F1OSPCNA	0
		CYCYCTCTCTCTCTC	GAGAGAGAGAGAGAGA	GAGA8HVBKN3	0
		CGATCGA	AGATCGACG	NONAMERATH4	4.00E-03
		CCCAAAA	TACACTTTTGG	OBP1ATGST6	1.60E-02
		CCACCA	TCCACCATA	POLLEN2LELAT52	3.00E-03
		TGCATG	CATGCA	RYREPEATBNNAPA	1.20E-02
		TTCTTC	CTGAAGAAGAA	TL1ATSAR	9.00E-03
		GTCAAAM	TTTGACY	WBBOXPCWRKY1	2.00E-03
c	gibberellin metabolic process	CACACAC	CAAACAC	2SSEEDPROTBANAPA	1.20E-02
		CGATCGA	AGATCGACG	NONAMERATH4	4.00E-03
		GGCCGG	ACCGGCCCACTT	PREMOTIFNPCABE	1.50E-02
		AACTTTC	CAACTTTCATAT	RSEPVGRP1	3.00E-03
		TGCATGC	CATGCA	RYREPEATBNNAPA	1.20E-02
d	calcium ion binding	-------	-------	-------	-------
e	boron transport	RAAACAT	AAATAGATAAATAAAAACATT	3AF1BOXPSRBCS3	6.00E-03
f	membrane	GATCGA	AGATCGACG	NONAMERATH4	4.00E-03
		GATCGATC	GATCATCGATC	RNFG1OS	6.00E-03
g	cytoplasmic membrane-bounded vesicle	CGATCGA	AGATCGACG	NONAMERATH4	4.00E-03
		GGCCGG	ACCGGCCCACTT	PREMOTIFNPCABE	1.50E-02
		TGCATGC	CATGCA	RYREPEATBNNAPA	1.20E-02
h	plastid	-------	-------	-------	-------
i	reproductive structure development (tapetal layer development)	CATGCA	CATGCA	RYREPEATBNNAPA	1.20E-02
		TATATATA	TGTATATAT	SORLREP3AT	2.10E-02
j	programmed cell death	-------	-------	-------	-------

4 Conclusions

In this work, we analyzed the transcriptome data of rice anther. Utilizing Markov Clustering, expression patterns were identified for different development stages. The function enrichment analysis shows that the identified expression patterns are indeed related to rice sexual reproduction. Furthermore, promoter motifs were predicted for the clusters, where the enriched motifs are expected to provide biological insights into regulatory mechanisms underlying rice sexual reproduction.

Acknowledgments

This work was partly supported by Innovation Funding of Shanghai University, Knowledge Innovation Program of the Chinese Academy of Sciences (Grant No. KSCX2-YW-N-005), Open Funding of National Key Laboratory of Plant Molecular Genetics, SRF for ROCS, SEM, Innovation Program of Shanghai Municipal Education Commission (09ZZ93), and JSPS-NSFC collaborative project.

References

[1] Watanabe M. Towards a Comprehensive Understanding of Molecular Mechanisms of Sexual Reproduction in Higher Plants. Plant Cell Physiol. 49(10): 1404–1406 (2008)
[2] Suwabe K, Suzuk G, Takahashi H. Separated Transcriptomes of Male Gametophyte and Tapetum in Rice: Validity of a Laser Microdissection (LM) Microarray. Plant Cell Physiol. 49(10): 1407–1416 (2008)
[3] Hobo T, Suwabe K, Aya K. Various Spatiotemporal Expression Profiles of Anther-Expressed Genes in Rice. Plant Cell Physiol. 49(10): 1417–1428 (2008)
[4] Van Dongen S. Graph Clustering Via a Discrete Uncoupling Process. SIAM J Matrix Anal Appl. 30-1:121-141(2008)
[5] Watson M. CoXpress: differential co-expression in gene expression data. BMC Bioinformatics. 7:509-520(2006)
[6] Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape Plugin to Assess Overrepresentation of Gene Ontology Categories in Biological Networks. Bioinformatics. 21(16):3448-3449(2005)
[7] Hughes JD, Estep PW, Tavazoie S et al. Computational Identification of Cis-regulatory Elements Associated with Groups of Functionally Related Genes in Saccharomyces Cerevisiae, J Mol Biol, 296: 1205-1214(2000)
[8] Mahony S, Benos PV. STAMP: a Web Tool for Exploring DNA-binding Motif Similarities. Nucleic Acids Res. 35(Web Server issue):W253-W258(2007)
[9] Sieglaffa DH, Augustine Dunna W, Xie XS et al. Comparative Genomics Allows the Discovery of Cis-regulatory Elements in Mosquitoes. PNAS. 106(9): 3053-3058(2009)
[10] Liu HF, Liao JP, Wu GJ. Ultrastructure of Microspore Development in Jatropha curcas L. J Trop Subtrop Bot. 15(2):107-114(2007)