Plant Biotechnology 26, 485–493 (2009)

Original Paper

Special Issue

Characterizing gene coexpression modules in *Oryza sativa* based on a graph-clustering approach

Atsushi Fukushima[1], Shigehiko Kanaya[2], Masanori Arita[1,3,4,*]

[1] RIKEN Plant Science Center, Yokohama, Kanagawa 230-0045, Japan; [2] Graduate School of Information Science, Nara Institute of Science and Technology, Nara 630-0101, Japan; [3] The University of Tokyo, Chiba 277-8561, Japan; [4] Keio University, Yamagata 997-0052, Japan
* E-mail: arita@k.u-tokyo.ac.jp Tel: +81-4-7136-3988 Fax: +81-4-7136-4074

Received September 4, 2009; accepted October 13, 2009 (Edited by H. Takahashi)

Abstract Recent advances in genome research have yielded a vast amount of large-scale data (e.g. DNA microarray) and have begun to deepen our understanding of plant cellular systems. Meta-analysis such as gene coexpression across publicly available microarrays has demonstrated that this approach is useful for investigating transcriptome organization and for predicting unknown gene functions in biological processes ranging from yeast to humans. However, no overall coexpression-network module in rice has been examined in detail. Here we present the coexpression clusters of rice genes based on unbiased graph clustering of the coexpression network of 4,495 genes. The coexpression network was constructed by using over 230 microarrays; it manifested several properties of a typical complex network (e.g. scale-free degree distribution). Using the DPClus algorithm that can extract densely connected clusters we detected 1,220 clusters. We evaluated these clusters using gene ontology enrichment analysis. We conclude that this approach is important for generating experimentally testable hypotheses for uncharacterized gene functions in rice and we posit that meta-analysis across publicly available microarrays will become increasingly important in crop science.

Key words: Coexpression, graph clustering, meta-analysis, regulon, rice.

Recent advances in genome research have produced a vast amount of genome-scale data such as transcriptome data and have begun to deepen our understanding of plant cellular systems. Oligonucleotide microarrays facilitate high-throughput and the simultaneous measurement of gene expressions. The model plant *Arabidopsis thaliana* has been subjected to thousands of microarray experiments and the results were deposited in online databases such as The Arabidopsis Information Resource (Swarbreck et al. 2008). Large-scale data-analysis across publicly available microarrays in Arabidopsis demonstrated that gene expressions are nonrandom, showing coordinated expression patterns (Williams and Bowles 2004; Ren et al. 2007; Krom and Ramakrishna 2008). As some details on the organization of the transcriptome of the Arabidopsis genome are now known (Fukushima et al. 2008; Mentzen and Wurtele 2008) it is possible to perform comparative gene-expression analysis of other plant species in relation to the Arabidopsis transcriptome. So-called "*meta-analysis*" provides opportunities for rice researchers to perform functional prediction based on gene coexpression such as ATTED-II (Obayashi et al. 2009). See also the review by Saito and colleagues (Saito et al. 2008).

Biological networks have been reported to be characterized by power-law degree distribution and modularity (Ravasz et al. 2002; Barabasi and Oltvai 2004). The gene coexpression- or association network reported by Stuart and colleagues (Stuart et al. 2003) also exhibited the characteristics typical for complex networks. A coexpression network is one of the threshold graphs and the thresholding procedure is considered an important mechanism underlying the generation of the power-law degree distribution. In the post-genomic era, the coexpression network approach (Aoki et al. 2007; Saito et al. 2008; Ogata and Shibata 2009) has been used to identify unknown gene functions, especially in pathways associated with glucosinolates, flavonoids, and phenylpropanoids (Gachon et al. 2005; Hirai et al. 2007; Yonekura-Sakakibara et al. 2008). Studies on the transcriptional coordination of enzyme genes in biochemical pathways have also been reported (Wei et al. 2006).

From a global perspective, Stuart and colleagues (Stuart et al. 2003) reported over 20,000 pairs of coexpressed genes across more than 3,000 microarrays from yeast, worms, and flies to humans. They

Abbreviations: n.s., not significant; GO, gene ontology; BP, biological process; CC, cellular component; MF, molecular function
This article can be found at http://www.jspcmb.jp/

demonstrated biological functions of conserved relationships associated with the cell cycle, ribosomal function, and metabolic pathways. Using the graphical Gaussian model, Ma and colleagues (Ma et al. 2007) presented a coexpression network in Arabidopsis; it allows the identification of metabolic processes as coherent coexpression modules. Biehl and colleagues (Biehl et al. 2005) identified 23 distinct clusters of coexpressed genes known as regulons. Mentzen and Wurtele (Mentzen and Wurtele 2008) who used Markov graph clustering estimated that the Arabidopsis transcriptome consists of approximately 1,000 regulons. They documented the coexpression of clusters involved in processes such as photosynthesis, protein synthesis, and mitosis. Integration of the coexpression approach, cis-regulatory elements, gene ontology, and orthology improves the prediction of protein–protein interactions (De Bodt et al. 2009) and alleviates inference by biological gene functions (Vandepoele et al. 2009). This type of systems approach can efficiently create a framework for generating testable hypotheses for further experiments.

Recently, rice (*Oryza sativa*) microarray analyses have been made available for a wide range of genetic modification perturbations, stresses, and chemical treatments. For example, >200 microarray data are deposited in NCBI GEO (Barrett et al. 2009). Despite the development of several coexpression databases such as RiceArrayNet (Lee et al. 2009), Oryza_Express (http://riceball.lab.nig.ac.jp/oryzaexpress/), and the Rice Coexpression Database (http://www.ricearray.org/rice_coexpression.shtml), the entire gene expression organization in the rice genome remains largely uncharacterized. With the accumulation of rice microarray data the meta-analysis approach can now be applied to this model crop plant. Using graph clustering of the coexpression network of 4,495 genes we investigated coexpression clusters (hereafter termed "regulons") of the rice genome. The resulting coexpression network exhibited several properties of a typical complex network such as scale-freeness or modularity. Over 1,000 clusters were detected by DPClus (Altaf-Ul-Amin et al. 2006) and assessed by Gene Ontology (GO) enrichment analysis.

Materials and methods

Microarray data

The Affymetrix GeneChip® Rice Genome Array includes 57,381 probe sets representing 51,279 transcripts (Affymetrix, USA). We obtained 234 GeneChips from NCBI GEO (Barrett et al. 2009) and PLEXdb (Wise et al. 2007). The data correspond to 17 experiments including analysis of the development time course, stress treatment, and genetic modifications (Figure 1, Supplemental Table 1, and Supplemental Figure 1).

Constructing gene coexpression networks

The raw CEL files within each dataset were pre-processed by the Robust Multichip Average (RMA) (Irizarry et al. 2003). The resulting $\log_2$ values were normalized to the same range by means-based scale normalization. Probe sets with the prefix "RPTR" or "AFFX" were eliminated. We calculated qualitative detection calls (present/absent) using the MAS5.0 algorithm. The obtained signal values were scaled so that the mean equaled 100. To identify highly expressed probe sets and to minimize noise in the microarray data, probe sets with low expression values (defined as probes whose expression in every GeneChip was lower than 200) were eliminated from further analysis (Figure 1B). A correlation matrix generated by Pearson's correlation coefficient (r) was calculated for the remaining 24,528 probes. We retained for analysis only 4,495 probes that were correlated above the threshold of $r=0.93$ ($p<8.1\text{e-}103$) with more than one other probe. We selected this threshold because of the limitations of the DPClus calculation. Finally, we extracted 32,544 pairs for the graph clustering. Annotation information for rice was obtained from the Affymetrix website. The data constructed can be downloaded and retrieved from our website http://prime.psc.riken.jp/rico/.

Topological network analysis

All network analyses were performed in R with "igraph" (Csardi and Nepusz 2006) and "netmodels" packages (http://cran.r-project.org/web/packages/netmodels/netmodels.pdf). The Pajek software program (Batagelj and Mrvar 1998) was used to visualize whole networks.

Graph clustering

To detect the coexpressed gene groups in rice we used DPClus (Altaf-Ul-Amin et al. 2006), a graph clustering software that can extract densely connected clusters. The algorithm is based on density- and periphery tracking of clusters. It is freely downloadable from http://kanaya.naist.jp/DPClus/. The parameter settings of cluster property c_p and density values were set to 0.5. The resulting clusters are listed in Supplemental Data 1. For comparisons with Arabidopsis we also constructed a coexpression network using 2364 ATH1 GeneChips (Fukushima et al. 2008) (see also Supplemental Figure 2) and classified them in the same manner. The 403 Arabidopsis clusters are shown in Supplemental Data 2 (see also Supplemental Table 2).

Gene ontology (GO) analysis

For the analysis of significantly overrepresented GO categories among clusters detected by graph clustering we used BiNGO. It retrieves the relevant annotations by statistically assessing the enrichment of GO terms in clusters (Maere et al. 2005). The hypergeometric test for each functional category was corrected by the Benjamini and Hochberg false-discovery rate for multiple testing. The rice GO annotation was based on Blast2GO Chip annotation data (Conesa and Gotz 2008).

Generation of random clusters and evaluation of clustering significance

To evaluate the significance of the resulting clusters we compared the overrepresented GO terms in the 14 largest

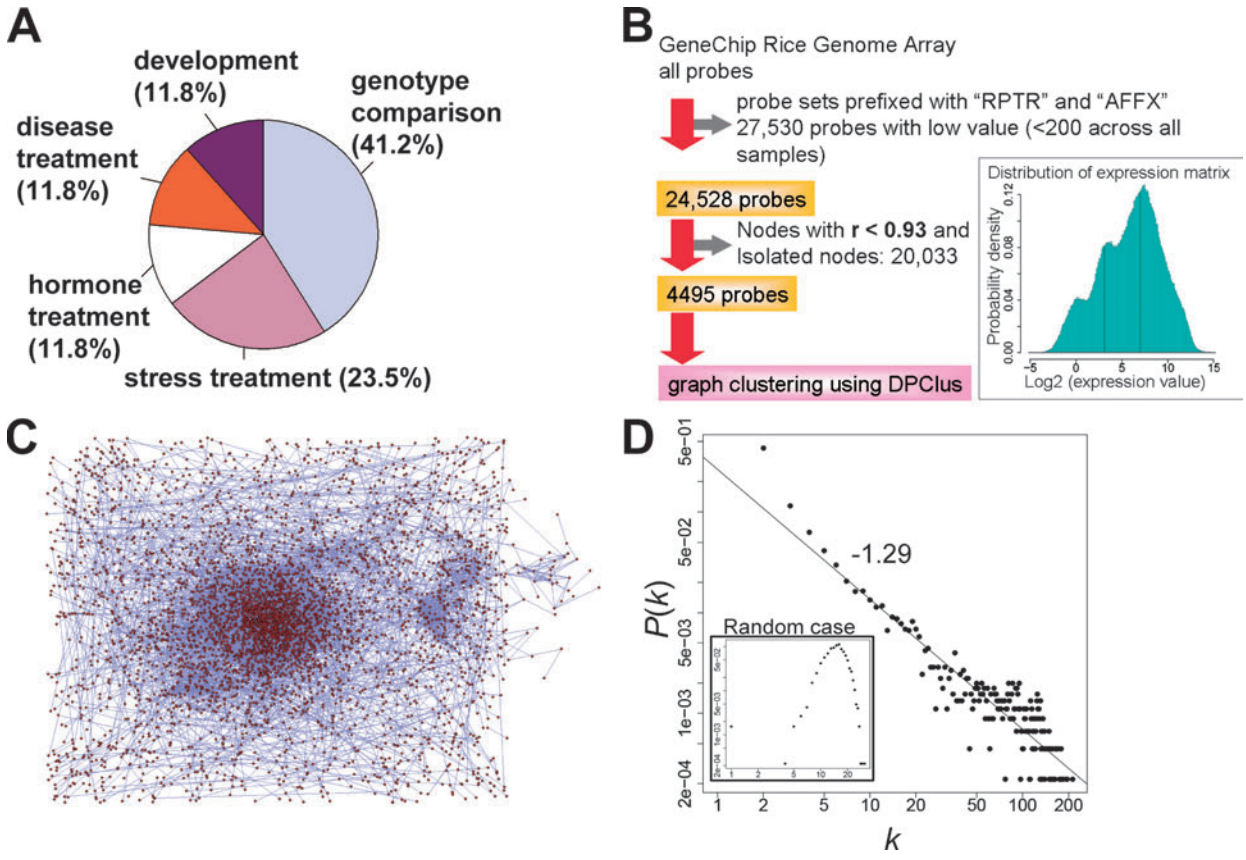

Figure 1. Topological properties of the gene coexpression network from rice data at a threshold of 0.93. (A) Pie chart of the experimental types of rice GeneChips used in this study. They were manually classified into 5 groups, i.e. genotype comparison, stress treatment, hormone treatment, disease treatment, and development. (B) Outline of the procedure for constructing the rice coexpression network using Affymetrix GeneChips. Distribution of the expression matrix is based on 4,495 genes calculated using the MAS5 algorithm. (C) An overview of the coexpression network with 4,495 genes and 32,544 edges (Pearson's correlation $r \geq 0.93$). The nodes represent genes, and the edges represent significant correlations between expression profiles. (D) The degree distribution of the coexpression network $P(k)$; k represents connectivity and $P(k)$ represents the connectivity distribution.

clusters in the rice dataset with the GO terms in 100 randomly generated sets of clusters. Each random cluster was generated by permutation of the gene identifiers without changing the cluster size. The evaluation was performed using the *S*-value (Mentzen and Wurtele 2008)

$$S_i = \frac{\sum_{j=1}^{n} p_{\min j}}{n},$$

where n represents the number of clusters ($n=14$) and i represents a cluster. This value is based on the best p-value p_{min} for overrepresentation of any GO term in each cluster. The best p-values were averaged over all clusters to provide the *S*-value.

Results

Construction of coexpression networks in rice

We manually classified 17 experimental datasets consisting of 234 rice GeneChips into 5 groups, i.e. genotype comparison, stress-, hormone-, and disease treatment, and development (Figure 1A). Of these, 41% and 24% were related to genotype comparison and stress treatment, respectively. The classification according to tissue type is shown in Supplemental Figure 1; it illustrates that 27%, 27%, and 11% were related to roots, leaves, and seedlings, respectively.

To identify the regulon organization of coexpressed rice genes we performed meta-analysis of the 234 rice GeneChips and constructed a transcriptome data matrix. To minimize noise and artifacts we calculated the signal intensity for all probe sets using MAS5 and RMA algorithms (see Materials and methods). Probes whose expression in all microarrays was lower than the two-fold mean value (=200) were removed (Figure 1B). To focus on the most notable coexpressed genes, only genes with a Pearson's correlation r higher than 0.93 with at least two other genes were extracted. The resulting 4,495 probe sets were used to construct the coexpression network (Figure 1C).

The structural properties of rice gene coexpression form a typical complex network

To assess whether the rice coexpression networks exhibited the common characteristics of a complex

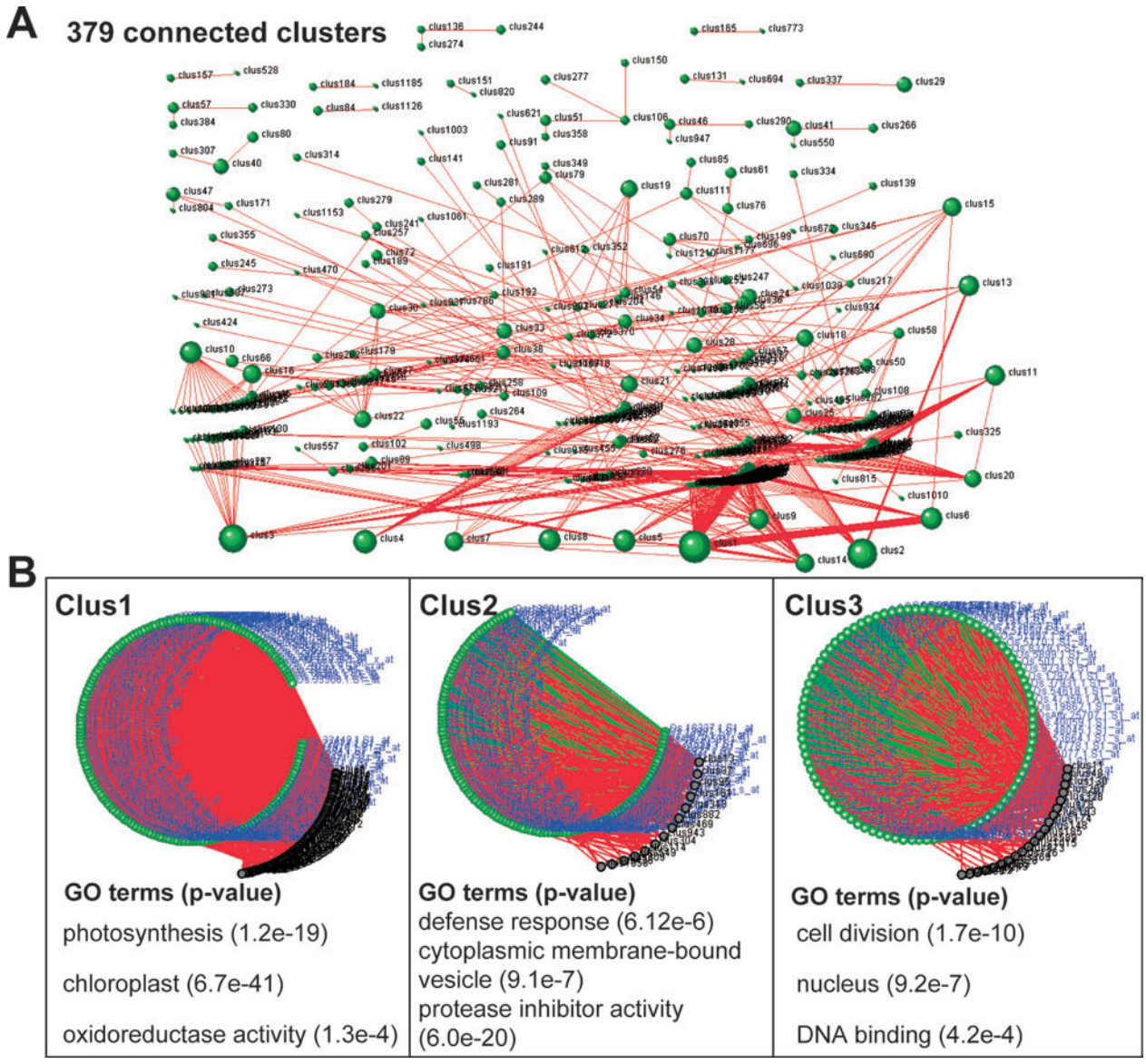

Figure 2. Graph clustering of the rice coexpression network. (A) Overview of a graph layout that represents dense subgraphs (cluster). A total of 379 connected clusters were detected at threshold $r \geq 0.93$. Both c_p and density values were set to 0.5. (B) Subgraphs of the top 3 clusters from the graph clustering results. To assess cluster fidelity, Gene Ontology (GO) term enrichment analyses were performed using BiNGO (Maere et al. 2005).

network (e.g. scale-free degree distribution) the topological properties of the network from rice microarray data were analyzed. Analysis of the connectivity of the network showed that the degree distribution followed a power-law with an exponent of 1.29 (Figure 1D). This is in accordance with earlier findings on microarray expression profiles (Bhan et al. 2002). The average path length and the average clustering coefficient of this network were 5.28 and 0.572, respectively, implying a modular structure. This global examination demonstrated that the properties of rice coexpression networks are consistent with the general network theory.

Graph clustering of coexpression networks reveals functional modules in the rice transcriptome

Using DPClus (Altaf-Ul-Amin et al. 2006) we identified 1,220 clusters in the rice coexpression network (4,495 nodes and 32,544 edges); they ranged in size from 2 to 165 genes. Figure 2A shows the 379 connected clusters; 841 clusters were independent with no links with other clusters or nodes. We detected 1,698 genes with at least 2 other clusters (Supplemental Table 3). Enriched GO terms in these groups were "plastid" and "intracellular membrane-bounded organelle". For example, genes involved in a maximum of 22 clusters were heat-shock protein-binding protein, putative (AK241571, OsAffx. 7755.1.S1_s_at) and rhodanese-like domain-containing

protein (AK120594, Os.11976.1.S1_a_at) (Supplemental Table 3), indicating the wide affinity of chaperon-like proteins.

The 14 largest clusters with at least 20 genes were analyzed with BiNGO (Maere et al. 2005) for their overrepresented GO categories (Table 1). Only the 3 largest clusters are presented in Figure 2B. To assess the significance of the clusters, we compared the overrepresented GO terms in the 14 largest clusters and randomized clusters (see Materials and methods). The averaged best *p*-value for overrepresented GO categories was better in real- than randomized data for the "Biological Process" category (Figure 3).

Table 1. Postulated physiological function of clusters

Cluster	# of genes	Predominant function in the biological process (*p*-value)
1	165	Photosynthesis (1.2e-19)
2	121	Defense response (6.2e-6)
3	90	Cell division (1.7e-10)
4	40	n.s.
5	30	Ribosome biogenesis and assembly (1.0e-16)
6	35	Photosynthesis (6.5e-28)
7	20	n.s.
8	31	n.s.
9	23	n.s.
10	30	Response to stress (4.4e-3)
11	25	Positive regulation of cell organization and biogenesis (2.6e-2)
12	20	Negative regulation of cell differentiation (2.7e-2)
13	27	n.s.
14	21	PSII-associated light-harvesting complex II catabolism (3.7e-2)

Clusters with 20 or more genes are shown. Annotations are postulated based on GO terms supplemented with information from the published literature (see Materials and methods). Abbreviation: n.s., not significant.

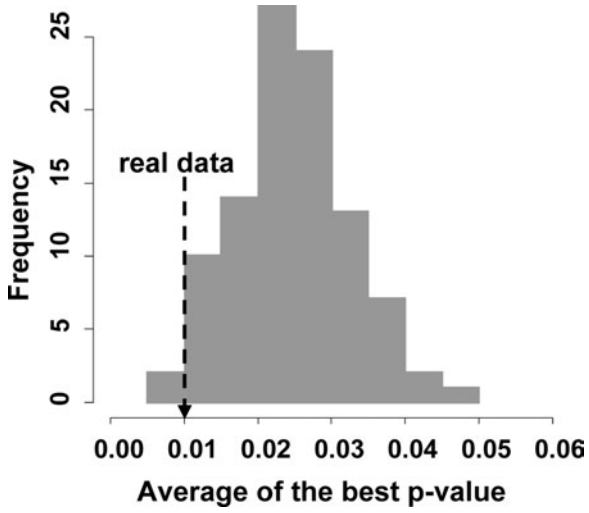

Figure 3. Assessment of statistical significance in the resulting clusters using DPClus. The best *p*-values for overrepresented GO terms within the domain Biological Process (BP), averaged over all clusters (*S* score) are compared to the corresponding values for 100 randomly generated clusters. The averaged *S* score (red arrow) was 0.0105; the *S* scores ranged from 6.5e-28 to 3.7e-2. This figure shows that the real dataset scored significantly better than the random dataset. GO term enrichment analysis was performed using BiNGO (Maere et al. 2005).

Predominant functions of clusters exhibit biological relevance

Photosynthesis

Six clusters (Clus1, Clus6, Clus14, Clus15, Clus26, and Clus28 shown in Table 1 and Supplemental Data 1) were involved in "photosynthesis" within the "Biological Process" domain (Figure 4). Clus14 and 28 were also related to photosynthesis function ("photosystem II associated light-harvesting complex II" and "photosynthetic electron transport," respectively). Members in clusters, Clus1, 6, and 26 were encoded predominantly by nuclear genes. Clus1 included genes encoding photosystem I proteins, thylakoid luminal protein-like, RuBisCO activase, thioredoxin M-type, and fructose-1,6-bisphosphatase. *OsSIG2* and *OsSIG6*, encoding potential plastid sigma factors of RNA polymerase (Kubota et al. 2007), were classified as members of Clus14 and 1, respectively. Kubota et al. (2007) reported that the expression of all 5 genes in rice (*OsSIG1*, *OsSIG2A*, *OsSIG2B*, *OsSIG5*, and *OsSIG6*) is induced by light.

In contrast, based on the functional annotations of the GeneChip probes, both Clus15 and 28 included plastid genes. These clusters were related to photosystem I/II proteins and a RuBisCO subunit. Taken together, graph clustering based on DPClus showed a good differentiation of clusters with nuclear and plastid genes. Therefore, it is suitable for describing certain biological processes from the perspective of the gene coexpression level.

Defense and stress responses

Clus2 contained 121 genes with the overrepresented GO term "defense response" (Supplemental Figure 3A), 7 were involved in seed allergenic proteins including RA5, RA16, RA17, RAG1, and RAG2 that belong to the alpha-amylase/trypsin inhibitor family. Another 11 genes in this cluster encoded glutelin, a major protein in rice seeds. Starch biosynthesis involved a putative gene encoding granule-bound starch synthase 1 (LOC_Os06g04200) in Clus2. In general, storage proteins and starch are predominant seed reserves and important factors in the quality and yield of rice. Our coexpression-based clustering approach may advance our understanding of the coordinated and regulated expression of these genes.

Clus10 included genes with the overrepresented GO term "response to stress" within the "Biological Process" domain. There were two genes encoding putative peroxidase and a gene encoding tonoplast intrinsic protein (OsTIP2;1), a member of the aquaporin family

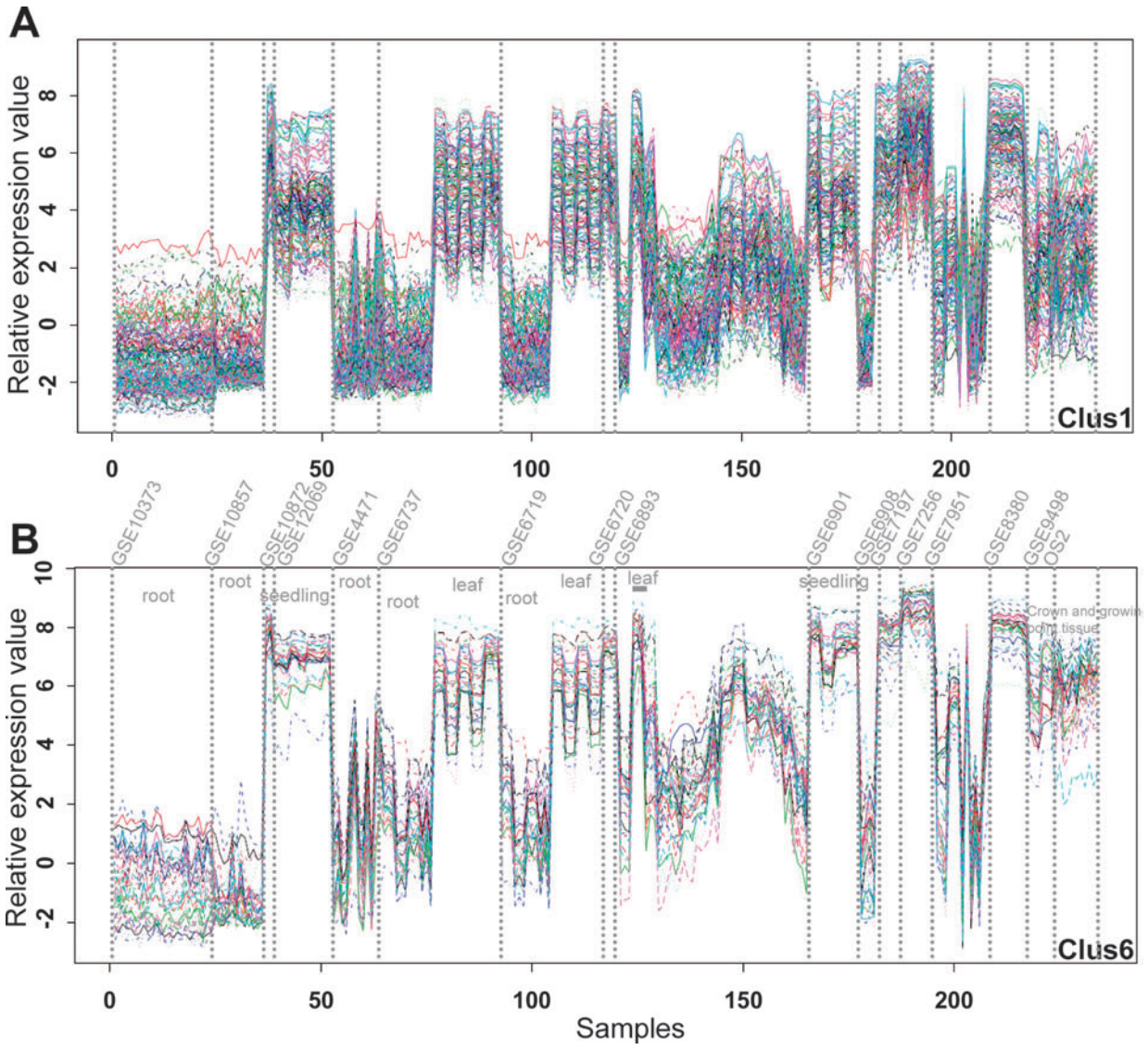

Figure 4. Clusters with the GO term "photosynthesis" and their gene expressions. Expression profiles of the genes in (A) Clus1 and (B) Clus6 across the 273 samples in the dataset. They were encoded predominantly by nuclear genes, indicating that these gene expressions are leaf-specific (or seedling-specific). Each gene expression is presented in a different color.

protein that acts as a water-transport channel. Aquaporin is involved in plant growth and water relations, as well as in the response to various types of stress (Sakurai et al. 2005). Other genes in this cluster encoded cysteine synthase (rcs4) (Nakamura et al. 1999) and putative glutathione S-transferase (GSTU6). Their coordination at the transcript level may allow plants to protect themselves against toxic oxygen intermediates.

Cell division

Clus3, the third-largest cluster, included 90 nuclear genes associated with the GO term "cell division." Most of the genes in this cluster were highly expressed in the shoot apical meristem (SAM) (Supplemental Figure 3B). Clus3 also contained genes encoding cyclin CYCB2;1 (LOC_Os04g47580) and cyclin-dependent kinase B2-1 (CDKB2;1, or cdc2Os3). This coincides with the observation that CYCB2;1 interacts with CDKB2;1 (Lee et al. 2003; Guo et al. 2007). In addition, this cluster included a gene encoding aurora kinase (LOC_Os01g09580) that belongs to a cell cycle-dependent serine/threonine protein kinase family (Kurihara et al. 2006), replication proteins A1 and A2, which are involved in DNA replication and recombination (van der Knaap et al. 1997), histone H3, and the structural maintenance of chromosomes (SMC4), playing an important role in both chromosome condensation and segregation (Cobbe and Heck 2004).

Ribosome biogenesis and assembly

Clus5, the fifth-largest cluster, included 30 genes involved in ribosome biogenesis and assembly; for

example, plastid ribosomal protein L11, ribosomal protein S9, and L12 (Supplemental Figure 3C). Many members of this cluster were genes encoding a putative or hypothetical ribosomal protein. Clus25 contained genes encoding 40S and 60S ribosomal proteins. Although the ribosomal proteins are well known and highly conserved among species, the unknown mechanisms regulating the coordination of ribosomal protein synthesis in plants are suggested.

Discussion

The aims of this study were to (1) construct a coexpression network of rice, (2) assess the relevance of rice coexpression clusters on the basis of current biological knowledge (e.g. gene ontology), and (3) compare clusters in rice and Arabidopsis. Using over 200 GeneChips, we constructed a gene coexpression network in rice. The resulting network exhibits typical properties including the power-law and modularity (Figure 1). The exponent, 1.29 in this case, is different from the exponent of a theoretical network (equal to 2) described by Masuda et al. (Masuda et al. 2004), implying a long-tailed distribution of gene expressions (see also Supplemental Figure 2C). Graph clustering based on the DPClus algorithm delineated 1,220 coexpression clusters. Our study of their biological relevance revealed that the coexpression modules of rice mainly represent photosynthesis, defense, cell division, and ribosomal protein complexes (Figure 2 and Table 1). In the cluster associated with photosynthesis we could clearly differentiate between clusters consisting of nuclear and plastid genes. This implies that the approach described here is useful for the computational prediction of the coordination of cellular compartments.

Our approach applied a stringent threshold for the correlation coefficients and the cut-off values for expression. It extracted conserved relationships associated with the cell cycle, ribosomal function, and metabolic pathways. For example, nuclear and plastid genes involved in photosynthesis were classified into different clusters (Supplemental Data 1) consistent with the results on Arabidopsis regulons obtained by Mentzen and Wurtele (Mentzen and Wurtele 2008). The clusters associated with cell division and ribosomal function, i.e. Clus3 and Clus5, contained many genes that have been experimentally characterized or are predicted to be involved in cell cycle processes and ribosomal biogenesis, respectively. We found that Clus10 included genes that participate in the response to stress and that gene OsTIP2;1 in the cluster is a plant aquaporin, which is distributed in a wide range of plant tissues. In Arabidopsis, for example, this protein can transport ammonia (Loque et al. 2005). Although there may be a difference in the diversity and number of microarrays used in our study of rice and in Arabidopsis studies, our findings strongly suggest that DPClus graph clustering can detect species-specific differences (see also Supplemental Data 2). Our approach could help to evaluate the function(s) of genes and of gene families in metabolic pathways.

We analyzed coexpression data statistically using Pearson's correlation coefficient to estimate the strength of the relationship between gene expression profiles. This coefficient is sensitive to outliers and assumes linearity between variables, therefore, if there is a nonlinear relationship between gene expression profiles it will yield poor estimates. However, we can defend our use of the Pearson correlation because the RMA normalization (see Materials and methods) is modeled as the sum of exponential and normal distributions for the signal and background, respectively. Although another similarity measure (e.g., mutual information) could be used to measure nonlinear "association" such an approach disclosed little difference in the highly correlated part of gene expression profiles (Daub et al. 2004).

The expression data used here had only 285 negative correlations at a significance level of $r<-0.93$, $p<8.1\text{e-}103$, indicating that there was an imbalance in the number of positive and negative correlations. The difficulty of capturing biologically meaningful negative correlations using DNA microarrays (e.g., analytical defects) may be one explanation. A study of the coexpression structure including negative correlations in the rice transcriptome remains to be undertaken.

In conclusion, our analysis facilitates the generation of new testable hypotheses for the unknown functions of genes. Unbiased graph clustering on a large-scale microarray dataset yields insights into the organization of the rice transcriptome. Unknown gene functions must be deciphered to gain a full understanding of the biological processes such as metabolic coordination, development, and stress responses. Meta-analysis across many experiments will become increasingly important in crop science.

Acknowledgements

We would like to thank Dr. Henning Redestig for fruitful comments on the manuscript, Drs. Kazuki Saito and Keiko Yonekura-Sakakibara for reading the manuscript, Dr. Miyako Kusano for discussions, Ursula Petralia for editing the manuscript, Masayoshi Wada for classifying Arabidopsis microarrays, and Tetsuya Sakurai and Kenji Akiyama for computational assistance.

References

Altaf-Ul-Amin M, Shinbo Y, Mihara K, Kurokawa K, Kanaya S (2006) Development and implementation of an algorithm for detection of protein complexes in large interaction networks.

BMC Bioinformatics 7: 207

Aoki K, Ogata Y, Shibata D (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol* 48: 381–390

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5: 101–113

Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, et al. (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucl Acids Res* 37: D885–890

Batagelj V, Mrvar A (1998) Pajek-Program for large network analysis. *Connections* 21: 47–57

Bhan A, Galas DJ, Dewey TG (2002) A duplication growth model of gene expression networks. *Bioinformatics* 18: 1486–1493

Biehl A, Richly E, Noutsos C, Salamini F, Leister D (2005) Analysis of 101 nuclear transcriptomes reveals 23 distinct regulons and their relationship to metabolism, chromosomal gene distribution and co-ordination of nuclear and plastid gene expression. *Gene* 344: 33–41

Cobbe N, Heck MM (2004) The evolution of SMC proteins: phylogenetic analysis and structural implications. *Mol Biol Evol* 21: 332–347

Conesa A, Gotz S (2008) Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *Int J Plant Genomics* 2008: 619832

Csardi G, Nepusz T (2006) The igraph software package for complex network research. *Int J Complex Syst* 1695 http://igraph.sf.net

Daub CO, Steuer R, Selbig J, Kloska S (2004) Estimating mutual information using B-spline functions—an improved similarity measure for analysing gene expression data. *BMC Bioinformatics* 5: 118

De Bodt S, Proost S, Vandepoele K, Rouze P, Van de Peer Y (2009) Predicting protein-protein interactions in *Arabidopsis thaliana* through integration of orthology, gene ontology and co-expression. *BMC Genomics* 10: 288

Fukushima A, Wada M, Kanaya S, Arita M (2008) SVD-based anatomy of gene expressions for correlation analysis in *Arabidopsis thaliana*. *DNA Res* 15: 367–374

Gachon CM, Langlois-Meurinne M, Henry Y, Saindrenan P (2005) Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis*: functional and evolutionary implications. *Plant Mol Biol* 58: 229–245

Guo J, Song J, Wang F, Zhang XS (2007) Genome-wide identification and expression analysis of rice cell cycle genes. *Plant Mol Biol* 64: 349–360

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, et al. (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc Natl Acad Sci USA* 104: 6478–6483

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249–264

Krom N, Ramakrishna W (2008) Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, *Arabidopsis*, and *populus*. *Plant Physiol* 147: 1763–1773

Kubota Y, Miyao A, Hirochika H, Tozawa Y, Yasuda H, et al. (2007) Two novel nuclear genes, OsSIG5 and OsSIG6, encoding potential plastid sigma factors of RNA polymerase in rice: tissue-specific and light-responsive gene expression. *Plant Cell Physiol* 48: 186–192

Kurihara D, Matsunaga S, Kawabe A, Fujimoto S, Noda M, Uchiyama S, Fukui K (2006) Aurora kinase is required for chromosome segregation in tobacco BY-2 cells. *Plant J* 48: 572–580

Lee J, Das A, Yamaguchi M, Hashimoto J, Tsutsumi N, Uchimiya H, Umeda M (2003) Cell cycle function of a rice B2-type cyclin interacting with a B-type cyclin-dependent kinase. *Plant J* 34: 417–425

Lee TH, Kim YK, Pham TT, Song SI, Kim JK, et al. (2009) RiceArrayNet: A database for correlating gene expression from transcriptome profiling, and its application to the analysis of coexpressed genes in rice. *Plant Physiol* in press.

Loque D, Ludewig U, Yuan L, von Wiren N (2005) Tonoplast intrinsic proteins AtTIP2;1 and AtTIP2;3 facilitate NH3 transport into the vacuole. *Plant Physiol* 137: 671–680

Ma S, Gong Q, Bohnert HJ (2007) An *Arabidopsis* gene network based on the graphical Gaussian model. *Genome Res* 17: 1614–1625

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448–3449

Masuda N, Miwa H, Konno N (2004) Analysis of scale-free networks based on a threshold graph with intrinsic vertex weights. *Phys Rev E Stat Nonlin Soft Matter Phys* 70: 036124

Mentzen WI, Wurtele ES (2008) Regulon organization of *Arabidopsis*. *BMC Plant Biol* 8: 99

Nakamura T, Yamaguchi Y, Sano H (1999) Four rice genes encoding cysteine synthase: isolation and differential responses to sulfur, nitrogen and light. *Gene* 229: 155–161

Obayashi T, Hayashi S, Saeki M, Ohta H, Kinoshita K (2009) ATTED-II provides coexpressed gene networks for *Arabidopsis*. *Nucl Acids Res* 37: D987–991

Ogata Y, Shibata D (2009) Practical network approaches and biologic interpretations of coexpression analyses in plants. *Plant Biotechnol* 26: 3–7

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551–1555

Ren XY, Stiekema WJ, Nap JP (2007) Local coexpression domains in the genome of rice show no microsynteny with *Arabidopsis* domains. *Plant Mol Biol* 65: 205–217

Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics—'majority report by precog'. *Trends Plant Sci* 13: 36–43

Sakurai J, Ishikawa F, Yamaguchi T, Uemura M, Maeshima M (2005) Identification of 33 rice aquaporin genes and analysis of their expression and function. *Plant Cell Physiol* 46: 1568–1577

Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302: 249–255

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, et al. (2008) The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucl Acids Res* 36: D1009–1014

van der Knaap E, Jagoueix S, Kende H (1997) Expression of an ortholog of replication protein A1 (RPA1) is induced by gibberellin in deepwater rice. *Proc Natl Acad Sci USA* 94: 9979–9983

Vandepoele K, Quimbaya M, Casneuf T, De Veylder L, Van de Peer Y (2009) Unraveling transcriptional control in *Arabidopsis* using cis-Regulatory elements and coexpression networks.

Plant Physiol 150: 535–546
Wei H, Persson S, Mehta T, Srinivasasainagendra V, Chen L, Page GP, Somerville C, Loraine A (2006) Transcriptional coordination of the metabolic network in *Arabidopsis*. *Plant Physiol* 142: 762–774
Williams EJ, Bowles DJ (2004) Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res* 14: 1060–1067
Wise RP, Caldo RA, Hong L, Shen L, Cannon E, Dickerson JA (2007) BarleyBase/PLEXdb. *Methods Mol Biol* 406: 347–363
Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. *Plant Cell* 20: 2160–2176