Global Landscape of a Co-Expressed Gene Network in Barley and its Application to Gene Discovery in Triticaceae Crops

Keiichi Mochida1,2,*, Yukiko Uehara-Yamaguchi1,2, Takuhiro Yoshida2, Tetsuya Sakurai2 and Kazuo Shinozaki1,2

1RIKEN Biomass Engineering Program, Yokohama, 230-0045 Japan
2RIKEN Plant Science Center, Yokohama, 230-0045 Japan

*Corresponding author: E-mail, mochida@psc.riken.jp; Fax, +81-45-503-9591

(Received February 14, 2011; Accepted March 16, 2011)

Accumulated transcriptome data can be used to investigate regulatory networks of genes involved in various biological systems. Co-expression analysis data sets generated from comprehensively collected transcriptome data sets now represent efficient resources that are capable of facilitating the discovery of genes with closely correlated expression patterns. In order to construct a co-expression network for barley, we analyzed 45 publicly available experimental series, which are composed of 1,347 sets of GeneChip data for barley. On the basis of a gene-to-gene weighted correlation coefficient, we constructed a global barley co-expression network and classified it into clusters of sub-network modules. The resulting clusters are candidates for functional regulatory modules in the barley transcriptome. To annotate each of the modules, we performed comparative annotation using genes in Arabidopsis and Brachypodium distachyon. On the basis of a comparative analysis between barley and two model species, we investigated functional properties from the representative distributions of the gene ontology (GO) terms. Modules putatively involved in drought stress response and cellulose biogenesis have been identified. These modules are discussed to demonstrate the effectiveness of the co-expression analysis. Furthermore, we applied the data set of co-expressed genes coupled with comparative analysis in attempts to discover potentially Triticaceae-specific network modules. These results demonstrate that analysis of the co-expression network of the barley transcriptome together with comparative analysis should promote the process of gene discovery in barley. Furthermore, the insights obtained should be transferable to investigations of Triticaceae plants. The associated data set generated in this analysis is publicly accessible at http://coexpression.psc.riken.jp/barley/.

Keywords: Barley • Co-expression analysis • Transcriptome.

Abbreviations: DHN, dehydrin; EST, expressed sequence tag; GO, gene ontology; PCC, Pearson correlation coefficient; QTL, quantitative trait locus; RG, rhamnogalacturonan; SNP, single nucleotide polymorphism; wPCC, weighted Pearson correlation coefficient.

Introduction

Comprehensive and high-throughput analysis of gene expression has become a significant approach for screening candidate genes, predicting gene function, discovery of cis-regulatory motifs and characterizing transcriptional regulatory networks (Goda et al. 2008, Brady and Provart 2009, Vandepoele et al. 2009). The relatively recent rapid accumulation of large-scale gene expression data sets has provided us with an efficient and valuable resource for many secondary uses such as co-expression analyses and comparative analyses (Barrett et al. 2009, Sasaki et al. 2010). Many genomic-scale data sets of plants have become accessible over the last few years. As a result of this, we have been able to construct detailed molecular maps for any multicellular organism. Networks and pathways have been reconstructed using omics data sets, such as transcriptomes, genome-wide transcription factor-binding networks, proteomes and metabolomes. Subsequently, these networks and pathways have been used to infer functional interactions among genes, proteins and metabolites (Moreno-Risueno et al. 2010).

Several attempts have been made to use co-expression analysis of a transcriptome. These investigations have employed various approaches in the discovery of genes and in prediction of the function of genes (Aoki et al. 2007). For example, co-expression data of the Arabidopsis transcriptome provided by the ATTED-II database have been applied in investigations of key genes involved in specific metabolic pathways and further used in characterizing the configuration of a metabolome coupled with mutant lines of the targeted genes (Obayashi et al. 2009). The ATTED-II database was used to identify novel genes involved in lipid metabolism. This effort led to the identification of a novel gene, UDP-glucose pyrophosphorylase3, that is required in the first step of sulfolipid biosynthesis.

© The Author 2011. Published by Oxford University Press on behalf of Japanese Society of Plant Physiologists.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
(Okazaki et al. 2009). Co-expression analysis was also used to identify all of the genes related to flavonoid biosynthesis. This effort led to a further detailed analysis of two flavonoid pathway genes, UGT78D3 and RHMT (Yonekura-Sakakibara et al. 2008).

In order to provide co-expression data sets, various databases of co-expression data of plant species have been constructed and made available to the public. The ATTED-II database is a representative database that provides an Arabidopsis co-expression data set with a web-accessible user interface (Obayashi et al. 2009). The RiceArrayNet and Oryza Express databases provide web-accessible co-expression data sets for rice (Lee et al. 2009, Hamada et al. 2011).

Among the cereals, barley (Hordeum vulgare), a diploid Triticeae plant, currently ranks fourth after maize (Zea mays), rice (Oryza sativa) and wheat (Triticum aestivum) in terms of total production (Schulte et al. 2009). Various genomic resources for barley have recently become available. Large-scale collections of cDNAs and expressed sequence tags (ESTs) have been obtained for various strains (Zhang et al. 2004, Sato et al. 2009). Using the assembly of ESTs, comprehensive efforts to discover single nucleotide polymorphisms (SNPs) were also carried out (Kota et al. 2008). Assembled EST data sets were used in the design of a SNP-typing chip for construction of a high-density genetic map (Close et al. 2009, Sato and Takeda 2009). The chip was also used to characterize genome-wide polymorphism patterns among genetic resources for barley (Rostoks et al. 2005, Rostoks et al. 2006). Information resources including full-length cDNA and genetic markers of barley were also developed and released (Mochida et al. 2008, Mochida et al. 2009).

Recently, several large-scale gene expression analyses of barley have been conducted using a GeneChip designed using tentatively transcribed consensus sequences that were obtained mainly from clustered ESTs (Close et al. 2004). For example, the gene expression profile obtained under salt stress can be used to gain an understanding of the global changes that occur in the transcriptome and to screen for genes that respond to stress conditions using the GeneChip (Walia et al. 2006). An atlas of gene expression patterns that occur throughout the development of the barley plant was constructed with the aim of characterizing global expression patterns of each gene as well as determining the transcriptome characteristics that differ among different tissues during development (Druka et al. 2006). Omics information has been obtained from the development of genomic resources for cereals [especially with respect to the recent release of genomic sequences for sorghum, corn and Brachypodium following that of rice in Poaceae (Mochida and Shinozaki 2010)]. The roles of genomic resources and their combinatorial use in barley have also been attracting increasing attention with respect to diploid crop species in the Triticeae group, as well as comparable species which have established genomic resources. These efforts will accelerate discovery of genes and lead to elucidation of the biological features responsible for agronomic traits.

This knowledge will be valuable for molecular breeding efforts (Sreenivasulu et al. 2008). Recently, gene expression data of barley have been rapidly accumulated during the course of various experiments (Shen et al. 2005), and have provided us with an opportunity to construct a co-expression network for the barley transcriptome. This network can be used to discover genes involved in specific regulatory networks. The gene expression data may also be used in comparative analyses with the objective of discovering homologous genes in other Triticeae crops. In this study, we analyzed > 1,000 data sets related to gene expression of the barley transcriptome. These data sets were collected using the Affymetrix barley GeneChip, which is available in the public domain. To construct a co-expression network of barley, we calculated the Pearson correlation coefficient (PCC) value for all combinations of available probes, and then identified co-expression modules associated with various biological processes. These modules were identified, annotated based on network topology, and the gene functions of the members of each module were predicted. The modules were characterized by coupling with expression patterns in various tissues or conditions. Enrichment analysis of gene ontology (GO) terms was also applied to estimate the functional category for each particular co-expression module. Furthermore, we attempted to define the relationships between conserved cDNAs in barley and wheat based on comparative analysis. We also performed comparative analysis in attempts to discover potentially Triticeae-specific network modules. This study demonstrates that co-expression analysis facilitates the identification of gene function and discovery of functionally related genes. This study also promotes knowledge exchange among Triticeae species by the combinatorial use of genomic information resources.

Results and Discussion

Construction of a global co-expressed gene network

We retrieved a total of 1,347 CEL files of a transcriptome data set of barley derived from PLEXdb, NCBI Gene Expression Omnibus (GEO) and EBI ArrayExpress, and then used these data sets to construct a co-expression gene network for barley (Table 1). To construct co-expression networks, weighted PCC values for all pair-wise combinations of the available probes were calculated. To obtain an overview of the topological features of the entire barley co-expression network, we conducted a survey on the number of links, number of probes and network density with respect to the probes with at least one link at the specified PCC cut-offs (Fig. 1). The number of links and probes decreases when the increasing PCC cut-off threshold increases (Fig. 1A, B). The network density, however, displays a minimal value at a PCC of 0.92, and shows a slight increase at a PCC cut-off greater than this value (Fig. 1C). The result of the calculation of the network density was found to be...
Table 1 List of experiment series of barley GeneChip used for the co-expression analysis

<table>
<thead>
<tr>
<th>Accession No.</th>
<th>Series name</th>
<th>DB</th>
<th>No. of CEL files</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB2</td>
<td>Expression profiling of wild type and mutants of Sultan 5 (Mla12) barley cultivar</td>
<td>PLEXDB</td>
<td>180</td>
</tr>
<tr>
<td>BB3</td>
<td>Transcription patterns during barley development</td>
<td>PLEXDB</td>
<td>63</td>
</tr>
<tr>
<td>BB4</td>
<td>Mla-specified transcriptional responses in barley–powdery mildew interactions</td>
<td>PLEXDB</td>
<td>108</td>
</tr>
</tbody>
</table>
| BB5            | A Ca
\(^{2+}\)/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning | PLEXDB | 4                |
| BB7            | mlo5-mediated resistance responses in barley | PLEXDB | 4                |
| BB9            | barley cv. Morex inoculated with Fusarium graminearum and water as mock control | PLEXDB | 44               |
| BB10           | Transcription profiling of barley plants containing variants of Mla1 and Mla6 powdery mildew resistance genes | PLEXDB | 144              |
| BB20           | Genotype-dependent gene expression in barley | PLEXDB | 24               |
| BB21           | Genetics of gene expression in barley | PLEXDB | 41               |
| BB22           | Developing seeds of M955 low phytic acid barley 7d after anthesis | PLEXDB | 9                |
| BB28           | Expression profiling of Morex and rpr1 mutant | PLEXDB | 6                |
| BB46           | Comparison of wild-type and cell death mutant of barley plants containing Mla6 powdery mildew resistance gene | PLEXDB | 72               |
| BB47           | Transcriptome analysis of Bowman vs. four tillering mutants at four developmental stages | PLEXDB | 60               |
| BB49           | Barley stem rust interaction | PLEXDB | 68               |
| BB50           | Carbohydrate accumulation in barley leaves leads to senescence and protease gene up-regulation | PLEXDB | 21               |
| BB52           | Transcriptome analysis of trichothecene-induced gene expression in barley | PLEXDB | 18               |
| BB53           | Functional genomic analysis of barley (Hordeum vulgare L.) grain protein accumulation | PLEXDB | 24               |
| BB62           | barley host response to the direct application of the trichothecene mycotoxin deoxynivalenol | PLEXDB | 24               |
| BB63           | Late response to boron toxicity in barley leaves | PLEXDB | 9                |
| BB65           | Transcriptome analysis of cold acclimation in barley albina and xantha mutants | PLEXDB | 30               |
| BB71           | Microarray analysis of the interaction between Rhopalosiphum padi and partially resistant or susceptible barley lines | PLEXDB | 24               |
| BB73           | Comparative transcriptional profiling of organs of the barley spike | PLEXDB | 12               |
| BB74           | Response of barley roots during the host interaction with the plasmodiophorid virus vector Polymyxa graminis | PLEXDB | 6                |
| BB75           | Response of barley roots during the non-host interaction with the plasmodiophorid virus vector Polymyxa betae | PLEXDB | 6                |
| BB76           | Structural and functional characterization of a winter malting barley | PLEXDB | 12               |
| BB79           | Pseudomonas aeruginosa virulent factor to barley | PLEXDB | 3                |
| BB80           | ABA experiment | PLEXDB | 9                |
| BB81           | Low temperature stress in cv. Dicktoo | PLEXDB | 12               |
| BB82           | Transcriptome analysis of barley anthers: effect of mannitol treatment on microspore embryogenesis | PLEXDB | 6                |
| BB83           | Mercury toxicity in barley roots | PLEXDB | 6                |
| BB84           | Differentially expressed genes between drought-tolerant and drought-sensitive barley genotypes | PLEXDB | 35               |
| BB85/BB86      | Expression data from barley maturing grains | PLEXDB | 32               |
| BB87           | Expression data from malting barley seeds | PLEXDB | 22               |
| BB89           | Gene expression in the barley spike during drought stress | PLEXDB | 24               |
| BB91           | Transcriptome analysis of a breeding program pedigree | PLEXDB | 84               |
| GSE6990        | Barley drought stress | GEO | 9                |
| GSE6993        | Barley low temperature stress | GEO | 3                |
| GSE8712        | The effects of Yariv-reagent on barley aleurone gibberellic acid signaling | GEO | 6                |
| GSE10332       | Transcriptome analysis of cold acclimation in barley albina and xantha mutants1 | GEO | 30               |
| GSE11182       | Steptoe × Morex seedling leaf comparison | GEO | 6                |
| GSE18758       | Microarray data from barley aleurone | GEO | 15               |
| GSE20034       | Diurnal expression data from developing barley caryopses | GEO | 12               |
| GSE23775       | Transcriptome analysis of the barley fast neutron mutants nec3 | GEO | 6                |
| E-MEXP-301     | Transcription profiling of barley roots during adaptation to abiotic stress conditions | ArrayExpress | 5 |
| E-MEXP-729     | Transcription profiling of barley in response to nitrate, ammonium or both | ArrayExpress | 9 |

Total 1,347
The PCC value was used to construct the network with the lowest density (Fig. 1C). We attempted to use the PCC value as well as other procedures to identify co-expression network modules. The distribution of PCC values between each probe and its nearest neighbor probe sorted by PCC value was calculated. The distribution of the PCC value was found to have a broad peak within the range of PCC values from 0.75 to 0.95 (Fig. 1D). According to the number of probes along with the PCC threshold, at 0.75 of the PCC, more than half of the probes have been validated as having at least one link (Fig. 1A, D). As a threshold for construction of the co-expression network, we used the top 50 probes for each of the probes with a PCC value > 0.75 for the cut-off as a roughly defined threshold.

**Identification and functional predication of co-expressed modules**

Using a PCC value > 0.75 and the top 50 probes, the constructed co-expression network included 12,791 probes with 180,406 links. In regulatory networks in the transcriptome estimated on the basis of a large-scale data set such as the data set obtained from co-expression analysis, groups of densely connected genes are frequently interpreted as having important biological functions. A dense subnetwork of genes often represents functional modules of a regulatory network and/or members involved in a coherent biological process.

Since the barley co-expression network elucidated with the threshold remains quite complex, to identify network modules that are associated with biological functions in the barley transcriptome, we applied the NeMo algorithm that was recently developed as a Cytoscape plug in (Rivera et al. 2010). NeMo provides a method that combines a unique neighbor-sharing score with hierarchical agglomerative clustering to identify diverse network communities. This approach has been found to be better than or competitive with leading approaches developed thus far (Rivera et al. 2010). As a result of the NeMo analysis, 2,344 network modules could be identified. The redundancy of the members of each module was then removed by using the single linkage clustering method. Finally, 606 unified modules of co-expressed genes were identified, including 8,586 probes of barley genes (Supplementary Table S1). The number of member probes per module ranged from four to 72 (Supplementary Fig. S1).

The GO annotations are remarkably useful in the process of mining significant biological functions from large-scale data.
sets, such as transcriptome data (Gene Ontology Consortium 2006). The relationships between the functional information based on GO annotation and the expression patterns of a set of genes can help to better understand the underlying biological phenomena. Therefore, we attempted to build putative GO annotations of barley based on similarity searches with the protein data set of Arabidopsis as well as those of *Brachypodium distachyon*. *Brachypodium distachyon* is an emerging model species for the temperate grasses, which include important cereals such as barley and wheat (International Brachypodium Initiative. 2010). As a result of a homology search-based GO annotation of the genes corresponding to barley GeneChip probes, it was found that a total of 11,620 probes could be assigned to at least one GO term. A total of 103 and 3,467 probes were specifically identified by the annotation of Arabidopsis in TAIR and that of *Brachypodium* by InterProScan analysis, respectively. A total of 9,673 probes were identified by the annotation of both data sets. Among the applications of the GO annotation, gene annotation enrichment analysis is a promising high-throughput strategy that increases the likelihood of investigators identifying biological processes most pertinent to their study (Huang da et al. 2009). In order to estimate functional features of each of the modules, GO term enrichment analysis was performed for members of each module (Vandepoele et al. 2009). We adopted topGO, an R Bioconductor program package that allows us to perform GO term enrichment analysis of a selected gene set against a custom annotated gene set (Alexa et al. 2006). A topGO analysis using the custom data set of GO annotations of the probe set of the barley GeneChip consisting of 1,603 terms of the biological process identified 101 modules with at least one GO term of 322 terms in the biological process ontology that was significantly enriched with the threshold $P$-value $<1 \times 10^{-5}$ of Fisher’s test. To provide an overview of the distribution of modules in the broad range of the biological process ontology, re-mapping of each GO term to the GO Slim terms was performed using the AgBase web service (Fig. 2). The 101 modules with significantly enriched GO terms were covered by 32 slim terms in the applied threshold for the analysis, and these slim terms covered major cellular processes such as biosyntheses, cellular component organization and various metabolic processes, as well as cellular response processes. The relationships between each module and the enriched GO terms are summarized in Supplementary Table S2. The co-existence of the GO terms in certain modules should provide clues for predicting functional features of the module and differentiation among modules that have similar roles and consist of similar genes.

**Co-expression modules in functional categories of biological processes**

**Co-expression module in response to drought stress.** As one of the demonstrable results of co-expression analysis, we have focused on gene network modules in response to drought stress conditions. Environmental stresses such as drought, high salt and low temperature affect plant growth and cause extreme decreases in crop productivity (Umezawa et al. 2006, Yamaguchi-Shinozaki and Shinozaki 2006). It is important to improve the stress tolerance of crops to increase crop yields under stress conditions. Drought tolerance is a key trait for increasing and stabilizing worldwide barley productivity in dry areas (Talame et al. 2007, Tommasini et al. 2008). Identification of the genes responsible for drought tolerance in barley (*H. vulgare L.*) will improve our understanding of the molecular mechanisms of drought tolerance and also facilitate the genetic improvement of barley by molecular breeding. The GOSlim term ‘response to abiotic stimulus’ is assigned to eight modules. The significantly enriched GO terms of module 406 under the GOSlim term were identified as ‘cold acclimation’, ‘response to water deprivation’ and ‘response to water’. The module includes 28 probes with 305 links (average PCC value = 0.854), some of which are densely allocated to the central area of a network graph and are tightly connected to each other as a result of the higher correlation of the expression pattern (Fig. 3A). The expression pattern of each of the probes was analyzed by using Genevestigator. This analysis showed that gene expression responds to drought stress conditions. This demonstrates good accordance with the co-expression analysis results (Fig. 3B). The module includes five probes of genes encoding dehydrin (DHN), three ABA-inducible genes and two genes encoding late embryogenesis abundant protein according to the annotation derived from NetAffx of Affymetrix, and these genes should be in good agreement with transcriptional instances induced in response to drought stress conditions (Table 2). The DHNs are a family of intrinsically unstructured proteins that have high water solubility and accumulate during late seed development under low temperature or water deficit conditions (Van Zee et al. 1995, Rorat 2006). These proteins are thought to play a protective role in freezing and drought tolerance in plants. Although drought stress-inducible promoters that function in Triticaceae should be important for engineering drought-tolerant crops, thus far there has been only one report on the genomic sequence of the promoter region of the barley *Dhn13* gene (Rodriguez et al. 2005). The members of module 406 may represent candidates for discovery of drought stress-inducible promoters. These sequences could also be compared to discover various types of drought stress-inducible promoters for fine-tuning of drought tolerance in barley. Our comparative analysis of target sequences corresponding to all probes of the barley GeneChip against the modeled proteome data set in Arabidopsis, rice and *Brachypodium*, as well as the data set of clustered cDNA sequences of wheat and barley in the TIGR Gene Index should provide useful information for predicting functions from knowledge collected from model plants. The knowledge gained for barley with respect to the drought stress response can be applied to studies of wheat. Drought tolerance is also a key trait for worldwide wheat production (Semenov and Halford 2009, Fleury et al. 2010).
Therefore, the co-expression module of barley should also provide a useful data set for gene discovery and lead to enhanced molecular engineering of drought stress tolerance in wheat. The high-throughput sequencing platform should enable us to easily perform large-scale collections of genomic fragments from non-sequenced crops with large complex genomes such as those of Triticeae crops (Schmutz et al. 2010). Therefore, to accelerate the discovery of promoters and genes that can promote molecular breeding of barley and the rationale design of promoters for combinatorial use, it is believed that co-expression data sets related to important crops should play more significant roles together with other accumulated comprehensive genomic and information resources.

Co-expression module involved in cellulose biogenesis. As another demonstrable result of this co-expression analysis, we focused on modules involved in cellulose biogenesis. Cellulose is the world’s most abundant biopolymer and a key structural component of the plant cell wall. Cell walls in cereal and grasses such as barley are characterized by the presence of (1,3;1,4)-β-D-glucans (Burton et al. 2011). These polysaccharides are beneficial constituents of human diets and reduce the risk of hypercholesterolemia, type II diabetes, obesity and colorectal cancer (Burton et al. 2011). Furthermore, cell walls of grass species are expected to become a major source of soft cellulose biomass to enhance biomass deconstruction and generate biofuels (Demura and Ye 2010). The biosynthesis of the cell wall (1,3;1,4)-β-D-glucans in the Poaceae is mediated, at least in part, by members of the cellulose synthase-like family (CslF) of genes (Burton et al. 2006). Our co-expression analysis and module identification identified two co-expression modules that are expected to be involved in cellulose biogenesis.

Fig. 2 Distribution of the GO category in the biological process significantly enriched in a co-expression network module based on topGO analysis with a cut-off threshold of $P < 1 \times 10^{-5}$.
These modules were suggested by GO term enrichment analysis and the putative function of each member gene. Module 563 consists of 20 probes with 99 links (PCC average = 0.802), whose members are abundantly expressed in most tissues (Fig. 4A, B). Module 563 includes five members of probes of genes putatively encoding cellulose synthases or cellulose synthase-like proteins (Table 3). Furthermore, the module includes two probes of genes that have significant homology with respect to putative arabinogalactan-like proteins and with respect to pectin-glucuronyltransferase. Arabinogalactan is a structural protein found in the primary plant cell wall (Liepman et al. 2010). Pectins, also one of the main components of the plant cell wall, are important in cell wall biosynthesis and development.
Table 2 Predicted gene functions and sequence identifiers of best hit counterparts in other plant species of the probes of members in module 406

<table>
<thead>
<tr>
<th>Probe name</th>
<th>Annotation from NetAffx</th>
<th>Bdi</th>
<th>Osa</th>
<th>Ath</th>
<th>TAGI</th>
<th>HVG1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig30934_at</td>
<td>AAN06848.1 2e-16 (AC099401) Putative ABA-induced protein [Oryza sativa (japonica cultivar group)]</td>
<td>Bradi4g07340.1 5e-32</td>
<td>Os06g0341300</td>
<td>AK107654 1e-32</td>
<td>AT3G22490.1 2e-16</td>
<td>TC416163 2e-155</td>
</tr>
<tr>
<td>Contig33855_at</td>
<td>AAN05568.1 1e-29 hypothetical protein [Oryza sativa (japonica cultivar group)]</td>
<td>Bradi3g03020.1 5e-34</td>
<td>Os10g0509300</td>
<td>AK062588 6e-24</td>
<td>AW448829 5e-175</td>
<td>TC226835 0.0</td>
</tr>
<tr>
<td>Contig15682_at</td>
<td></td>
<td>TC439492 2e-34</td>
<td></td>
<td>TC231237 1e-122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig1701_s_at</td>
<td>AAD02254.1 1e-30 (AF043088) dehydrin 2 [Hordeum vulgare] [Hordeum vulgare subsp. vulgare]</td>
<td>Bradi4g22290.1 3e-29</td>
<td>Os11g0453900</td>
<td>AK109096 3e-24</td>
<td>AT5G66400.1 4e-20</td>
<td>TC381055 8e-149</td>
</tr>
<tr>
<td>Contig1708_s_at</td>
<td>AAF01694.1 5e-27 (AF181456) dehydrin; DHN6 [Hordeum vulgare] [Hordeum vulgare subsp. vulgare]</td>
<td>Bradi4g22290.1 2e-30</td>
<td>Os11g0454300</td>
<td>AK121952 4e-30</td>
<td>AT5G66400.1 1e-22</td>
<td>TC395863 5e-120</td>
</tr>
<tr>
<td>Contig1718_s_at</td>
<td>AAD02260.1 3e-30 (AF043094) dehydrin 9 [Hordeum vulgare] [Hordeum vulgare subsp. vulgare]</td>
<td>Bradi3g43870.1 3e-39</td>
<td>Os11g0454000</td>
<td>AK071366 2e-33</td>
<td>AT5G66400.1 1e-20</td>
<td>CJS96275 2e-103</td>
</tr>
<tr>
<td>Contig1724_s_at</td>
<td>AAD02255.1 6e-19 (AF043089) dehydrin 3 [Hordeum vulgare] [Hordeum vulgare subsp. vulgare]</td>
<td>Bradi4g23740.1 6e-51</td>
<td>Os11g0454300</td>
<td>AK121952 7e-37</td>
<td>AT5G66400.1 2e-18</td>
<td>TC371537 0.0</td>
</tr>
<tr>
<td>Contig1725_s_at</td>
<td>P129511 3e-29 DEHYDRIN DHN1 (B8) pir</td>
<td></td>
<td>S05544 dehydrin 8—barley</td>
<td>Bradi4g3740.1 2e-41</td>
<td>Os11g0454300</td>
<td>AK121952 3e-41</td>
</tr>
<tr>
<td>Contig1884_at</td>
<td>BAB39231.1 1e-79 (AP002869) hypothetical protein—similar to Arabidopsis thaliana chromosome 5</td>
<td>Bradi2g12390.1 4e-13</td>
<td>Os10g0329400</td>
<td>AK101813 3e-13</td>
<td>AT5G15640.1 7e-07</td>
<td>TC403160 6e-141</td>
</tr>
<tr>
<td>Contig19400_at</td>
<td></td>
<td>TC447298 7e-27</td>
<td></td>
<td>TC233994 5e-90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig19500_at</td>
<td></td>
<td>TC423324 3e-41</td>
<td></td>
<td>TC233376 8e-152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig25998_at</td>
<td>T06978 4e-80 ABA-induced plasma membrane protein PMA 19—wheat gb</td>
<td></td>
<td>AAB38504.1 [ABA-induced plasma membrane protein PM 19 [Triticum aestivum]]</td>
<td>Bradi4g00600.1 2e-21</td>
<td>Os05g0381400</td>
<td>AK102039 3e-15</td>
</tr>
<tr>
<td>Contig3690_s_at</td>
<td>AAF2532.1 5e-36 plasma membrane-associated protein [Hordeum vulgare]</td>
<td>Bradi4g09808.1 5e-23</td>
<td></td>
<td>TC195357 0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig4760_s_at</td>
<td>BAB3275.1 5e-67 putative late embryogenesis abundant protein LEA4-A [Oryza sativa (japonica cultivar group)]</td>
<td>Bradi2g07480.1 4e-59</td>
<td>Os01g0225600</td>
<td>Os01g0225600 3e-58</td>
<td>AT1G01470.1 2e-39</td>
<td>TC393160 0.0</td>
</tr>
<tr>
<td>Contig5269_s_at</td>
<td>NP_181073.1 4e-04 (NM_129082) similar to late embryogenesis abundant proteins; protein ID AAt2g35300.1</td>
<td>Bradi5g18970.1 3e-23</td>
<td>Os04g0589800</td>
<td>AK063682 2e-21</td>
<td>AT2G35300.1 2e-11</td>
<td>CA701261 1e-97</td>
</tr>
<tr>
<td>Contig5270_s_at</td>
<td></td>
<td>TC393321 4e-41</td>
<td></td>
<td>TC231848 1e-113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig5946_at</td>
<td>AAL8488.1 3e-04 (AC073556) putative embryo-specific protein [Oryza sativa (japonica cultivar group)]</td>
<td>Bradi1g73840.1 2e-19</td>
<td>Os03g0168100</td>
<td>AK121575 4e-13</td>
<td>AT2G18340.1 1e-08</td>
<td>TC383443 3e-66</td>
</tr>
<tr>
<td>Contig6087_at</td>
<td>NP_196390.1 1e-21 putative protein; protein ID: At5g07330.1 [Arabidopsis thaliana] pir</td>
<td></td>
<td>T49872 hypothetical protein T2II.40—Arabidopsis thaliana</td>
<td>Bradi1g73840.1 2e-19</td>
<td>Os03g0168100</td>
<td>AK121575 4e-13</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Probe name</th>
<th>Annotation from NetAffx</th>
<th>Bdi</th>
<th>Osa</th>
<th>Ath</th>
<th>TAGI</th>
<th>HVGI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig6087_s_at</td>
<td>NP_196350.1 1e-21 putative protein; protein ID: At5g07330.1 [Arabidopsis thaliana]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>hypothetical protein T211.40—Arabidopsis thaliana</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig7112_at</td>
<td>NP_201479.1 9e-19 putative protein; protein ID: At5g65780.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig7501_s_at</td>
<td>AAJ7132.1 7e-82 Putative calcium-binding protein [Oryza sativa]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig8058_at</td>
<td>Q00747 7e-07 PROTEIN LE25 pir</td>
<td>S19253 gene le25 protein—tomato</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig8184_at</td>
<td>BAB44029.1 8e-48 WSI18 protein [Oryza sativa (japonica cultivar-group)] dbj[BAB86507.1] WSI18 protein [Oryza sativa (japonica cultivar-group)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H504L14u_at</td>
<td>BAB21159.1 004 (AP002899) contains ESTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H517L24u_x_at</td>
<td>BAB21159.1 004 (AP002899) contains ESTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT11004r_at</td>
<td>P14928 1e-16 ABA-inducible protein PHV A1 pir</td>
<td>S08313 ABA-induced protein HVA-1—barley</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bdi, protein data set annotated on the Brachypodium genome, Bdi1.0 seached by BLASTP with a threshold of e-value < 1e-10.
Osa, protein data set annotated on the rice genome, RAP2 seached by BLASTP with a threshold of e-value < 1e-10.
Ath, protein data set annotated on the Arabdopsis genome, TAIR9 seached by BLASTP with a threshold of e-value < 1e-5.
TAGI, TIGR Gene Index of wheat searched by BLASTN with a threshold of e-value < 1e-10.
HVGI, TIGR Gene Index of barley searched by BLASTN with a threshold of e-value < 1e-10.
primary cell wall, are complex polysaccharides containing homogalacturonan (HG), rhamnogalacturonan I (RG-I) and rhamnogalacturonan II (RG-II) regions (Ridley et al. 2001). In the analysis of pectin-glucuronyltransferase of Nicotiana plum-baginifolia, the mutation of NpGUTI was found to cause defects in the glucuronic acid of RG-II of pectin. This drastically reduced the formation of borate cross-links with RG-II, and it was concluded that NpGUTI is involved in pectin biosynthesis and essential for intercellular attachment to plant meristems and tissues (Iwai et al. 2002). Module 563, which was identified in this study, might be a part of the regulatory network for formation of the primary cell wall in barley, which has a highly correlated expression pattern. Because elucidation of the regulatory networks of genes involved in primary cell wall

Fig. 4 Predicted co-expression network modules involved in cellulose biogenesis. (A) Module 563 consists of 20 nodes with 99 edges (PCC average = 0.802). It has also been determined that module 548 putatively involves cell wall biogenesis. (C) The module consists of 20 probes with 127 edges (PCC average = 0.816). (B) Heat map visualization of gene expression patterns from Genevestigator on members in module 563 and (D) module 548.
Table 3  Predicted gene functions and sequence identifiers of best hit counterparts in other plant species of the probes of members in module 563

<table>
<thead>
<tr>
<th>Probe name</th>
<th>Annotation from NetAffx</th>
<th>Bdi</th>
<th>Osa</th>
<th>Ath</th>
<th>TAGI</th>
<th>HVGI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig10364_at</td>
<td>NP_191301.1</td>
<td>9e-83</td>
<td>(NM_115602) putative protein; protein id: At3g57420.1</td>
<td>[Arabidopsis thaliana]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig11177_at</td>
<td>AAL77142.1</td>
<td>2e-43</td>
<td>(AC097447) Putative myosin-like protein</td>
<td>[Oryza sativa]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig12023_at</td>
<td>NP_177870.1</td>
<td>1e-69</td>
<td>(NM_106395) unknown protein; protein ID: At1g77460.1</td>
<td>[Arabidopsis thaliana]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig16468_at</td>
<td>NP_181761.1</td>
<td>4e-41</td>
<td>(NM_129794) unknown protein;</td>
<td>[Arabidopsis thaliana]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig3478_at</td>
<td>AAF89964.1</td>
<td>e-131</td>
<td>cellulose synthase-4</td>
<td>[Zea mays]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig3965_at</td>
<td>BAC20928.1</td>
<td>1e-67</td>
<td>pectin-glucuronyltransferase</td>
<td>[Nicotiana plumbaginifolia]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig4221_at</td>
<td>AAF24189.1</td>
<td>e-123</td>
<td>phytochelatin synthetase-like protein</td>
<td>[Zea mays]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig4451_at</td>
<td>AAF89963.1</td>
<td>e-124</td>
<td>cellulose synthase-3</td>
<td>[Zea mays]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig4452_at</td>
<td>AAF89961.1</td>
<td>e-126</td>
<td>cellulose synthase-1</td>
<td>[Zea mays]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig5050_s_at</td>
<td>AAL83695.1</td>
<td>3e-17</td>
<td>hypothetical protein</td>
<td>[Oryza sativa (japonica cultivar group)]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig5706_at</td>
<td>AAL38530.1</td>
<td>e-121</td>
<td>CSLF6</td>
<td>[Oryza sativa]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig5754_at</td>
<td>NP_56774.1</td>
<td>1e-50</td>
<td>Expressed protein; protein ID: At4g27435.1</td>
<td>[Arabidopsis thaliana]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig5754_s_at</td>
<td>NP_56774.1</td>
<td>1e-50</td>
<td>Expressed protein; protein ID: At4g27435.1</td>
<td>[Arabidopsis thaliana]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig5827_at</td>
<td>BAC15948.1</td>
<td>1e-76</td>
<td>contains EST—hypothetical protein—similar to Arabidopsis thaliana chromosome 2</td>
<td>[Arabidopsis thaliana]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig7301_at</td>
<td>BAA90502.1</td>
<td>2e-90</td>
<td>maize EST AI621709</td>
<td>[Arabidopsis thaliana]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig7581_at</td>
<td>NP_683337.1</td>
<td>e-107</td>
<td>UDP-N-acetylglucosamine pyrophosphorylase-like protein; protein ID: At1g31070.2</td>
<td>[Arabidopsis thaliana]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bdi, protein data set annotated on the Brachypodium genome, Bdi1.0 searched by BLASTP with a threshold of e-value < 1e-10. Osa, protein data set annotated on the rice genome, RAP2 searched by BLASTP with a threshold of e-value < 1e-10. Ath, protein data set annotated on the Arabidopsis genome, TAIR9 searched by BLASTP with a threshold of e-value < 1e-5. TAGI, TIGR Gene Index of wheat searched by BLASTN with a threshold of e-value < 1e-10. HVGI, TIGR Gene Index of barley searched by BLASTN with a threshold of e-value < 1e-10.

biogenesis should promote discovery of useful genes to engineer cellulolic biomass productivity in barley, we expect that co-expression data sets will provide informative clues and contribute to approaches toward biomass engineering of grasses by integration with other biomass grass species.

Module 127 has also been identified as being putatively involved in cell wall biogenesis. The module consists of 20 probes with 127 links (PCC average = 0.816) (Fig. 4C, D). This module includes three probes that are homologous to genes encoding putative cellulose synthases (Table 4). The expression patterns of member genes in the module are quite different from the expression patterns of the members of module 563 based on the heat map of the expression pattern from Genevestigator (Fig. 4B, D). The genes are mainly expressed in the spikelet, crown and root. The module might be a part of a tissue-specific subnetwork related to cell wall biogenesis in barley. Regarding the barley cellulose synthase-like gene family, seven HvCsIF genes were identified in a similarity search of the homologs of rice. The expression and genetic relationships with quantitative trait loci (QTLs) with respect to the (1,3;1,4)-β-D-glucan content of grain were analyzed (Burton et al. 2008). In addition to this candidate gene approach, co-expression analysis and module identification of co-expression networks should provide collateral figures not only of target genes but also of potentially related genes. The modules identified by co-expression analysis might also be able to identify further candidate genes associated with related QTLs. Interestingly, the module also includes probes of genes encoding laccase, a putative member of the family of secondary cell wall-associated proteins (Zhao et al. 2009). The known genes associated with the primary and secondary cell wall coupled with expression patterns should provide informative clues for building a hypothesis for further analysis. Based on a comparative analysis of the genes of model plants, some members of the co-expressed genes in the module were found to be inter-related with putative homologs. However, the cell walls of grasses differ dramatically from the cell walls of dicots in terms of the major structural polysaccharides present, polysaccharide linkages and the abundance and importance of pectins, proteins and phenolic compounds (Vogel 2008). Therefore, omics-based approaches with insights obtained from systems biology should provide significant information that will increase our understanding of the characteristics of cell wall biogenesis in grasses.

Triticaceae-specific transcriptional network modules. Species-specific transcriptional networks might be associated with specific cellular systems and may also be involved in characteristic biological properties of each species. In a typical approach, genome-scale sequence comparisons based on comparative genomics have allowed us to identify conserved genes as well as potentially species-specific or lineage-specific genes that might be related to specific regulatory networks. In particular, in applied organisms, comparative analyses have been recognized as a useful means for introducing knowledge from other model organisms and to screen for potentially species-specific genes. Therefore, we performed comparative analysis among the data sets of target sequences of all of the 22,304 probes on the barley GeneChip analyzed in this study, and the sequence data set of proteins in Arabidopsis, rice and Brachypodium, as well as those of clustered transcripts in barley and wheat based on sequence similarity. We then classified all probes of the barley GeneChip into five categories based on the hit characteristics as follows: category 1, hit to barley and/or wheat transcripts as well as proteins in Arabidopsis, rice and Brachypodium (these should be potentially conserved genes among higher plants); category 2, hit to barley and/or wheat transcripts as well as proteins in rice and Brachypodium (these should be potentially conserved in Gramineae plants); category 3, hit to barley and/or wheat transcripts as well as proteins in Brachypodium (these should be potentially conserved in Pooideae species); category 4, hit only to barley and/or wheat transcripts (these should be potentially specific to Triticaceae species); and category 5, inconsistent hit characteristics in terms of phylogenetic relationships among these plants or no significant similarity to the data sets employed (Fig. 5A). According to this genome-scale comparative analysis, the barley GeneChip consists of about half of the probes (category 1, 11,066 probes) of genes showing significant similarity to Arabidopsis as well as Gramineae species, approximately 10% (category 2, 2,274 probes) and 2.7% (category 3, 605 probes) of those potentially specific to Gramineae and Pooideae species, respectively. Probes of genes showing significant similarity to barley and/or wheat transcripts (category 4, 7,512 probes) are approximately 34% of all probes analyzed. Although some limitations may remain due to the EST-based probe design in the barley GeneChip, the results suggest that there is a considerable number of probes that might represent specific or divergent genes in Triticaceae species.

We then applied the categorized probe data set of 8,586 genes classified into 606 network modules to screen for potentially Triticaceae-specific transcriptional modules (Fig. 5A). The 8,586 probes of genes with the network modules were classified into 4,900, 877, 227, 2,274 and 308 probes in categories 1, 2, 3, 4 and 5, respectively (Fig. 5A). To screen for network modules that are potentially specific to Triticaceae, the numbers of probes assigned to each category were counted in each module, and these data were hierarchically clustered and visualized as a heat map (Fig. 5B). The heat map visually assisted us in identifying clustered groups consisting of network modules that include probes assigned to category 4. As a result of the clustering based on numbers of member probes in each of the categories, we identified a group that consists of 27 modules preferentially including probes of category 4 (Fig. 5B). In Fig. 5, the module 492 consists of probes of category 2, 3 and 4, which indicate Gramineae-specific modules. This group includes two modules that consist only of probes of category 4 (modules 243 and 19); these could be potentially Triticaceae-specific modules. For instance, module 243 consists of 24 probes that are in category 4. These probes are potentially specific to
<table>
<thead>
<tr>
<th>Probe name</th>
<th>Annotation from NetAffx</th>
<th>Bdi Osa Ath TAGI HVGI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig10003_at</td>
<td>BAA65396.1 At (AP010217) EST D23183(R2868)</td>
<td></td>
</tr>
<tr>
<td>Contig10003_s_at</td>
<td>BAA65396.1 At (AP010217) EST D23183(R2868)</td>
<td></td>
</tr>
<tr>
<td>Contig10379_at</td>
<td>BAB16875.1 At (AP002537) hypothetical protein</td>
<td></td>
</tr>
<tr>
<td>Contig10380_at</td>
<td>BAB16875.1 At (AP002537) hypothetical protein</td>
<td></td>
</tr>
<tr>
<td>Contig10380_x_at</td>
<td>BAB16875.1 At (AP002537) hypothetical protein</td>
<td></td>
</tr>
<tr>
<td>Contig12571_at</td>
<td>AAM74394.1 At (AC119149) Putative cytochrome P450</td>
<td></td>
</tr>
<tr>
<td>Contig15116_at</td>
<td>BAB67900.1 At (AP003237) putative cellulose synthase</td>
<td></td>
</tr>
<tr>
<td>Contig18654_at</td>
<td>AAD21424.1 At (AC005882) hypothetical protein</td>
<td></td>
</tr>
<tr>
<td>Contig18837_at</td>
<td>AAB17192.1 At (U73104) laccase</td>
<td></td>
</tr>
<tr>
<td>Contig19672_at</td>
<td>NP_187609.2 At (NP_187609) hypothetical protein</td>
<td></td>
</tr>
<tr>
<td>Contig20165_at</td>
<td>AAM26299.1 At (AY095297) hypothetical protein</td>
<td></td>
</tr>
<tr>
<td>Contig20670_at</td>
<td>XP_129042.1 At (XM_129042) similar to ring finger protein</td>
<td></td>
</tr>
<tr>
<td>Contig25667_s_at</td>
<td>BJ451823 At (P0019E03.5) hypothetical protein</td>
<td></td>
</tr>
<tr>
<td>Contig9658_at</td>
<td>AAK27814.1 At (AC037426) unknown protein</td>
<td></td>
</tr>
<tr>
<td>HP01G01w_at</td>
<td>AAK15451.1 At (AC037426) unknown protein</td>
<td></td>
</tr>
<tr>
<td>HP01G01w_s_at</td>
<td>AAK15451.1 At (AC037426) unknown protein</td>
<td></td>
</tr>
<tr>
<td>HVSMEn0007G22f2_at</td>
<td>AAK15451.1 At (AC037426) unknown protein</td>
<td></td>
</tr>
<tr>
<td>HVSMEn0020J09f_at</td>
<td>AAK15451.1 At (AC037426) unknown protein</td>
<td></td>
</tr>
</tbody>
</table>

Bdi, protein data set annotated on the Brachypodium genome, B410 searched by BLASTP with a threshold of e-value < 1e-10.
Osa, protein data set annotated on the rice genome, RAP2 searched by BLASTP with a threshold of e-value < 1e-10.
Ath, protein data set annotated on the Arabidopsis genome, TAIR9 seached by BLASTP with a threshold of e-value < 1e-5.
TAGI, TIGR Gene Index of wheat searched by BLASTN with a threshold of e-value < 1e-10.
HVGI, TIGR Gene Index of barley searched by BLASTN with a threshold of e-value < 1e-10.

797

Fig. 5 Identification of putative lineage-specific network modules. (A) A bar graph showing numbers and proportions of the probes for all probes analyzed (22,043) and probes classified in modules (8,586). The graph indicates significant similarity among transcripts of barley and wheat, and/or proteins of Arabidopsis, rice and Brachypodium. The hit patterns in the similarity search are classified into five categories: (1) hit to barley or wheat transcripts as well as proteins in Arabidopsis, rice and Brachypodium, (2) hit to barley or wheat transcripts as well as proteins in rice and Brachypodium, (3) hit to barley or wheat transcripts as well as proteins in Brachypodium, (4) hit to barley or wheat transcripts and (5) inconsistent hit or no significant similarity. (B) A heat map representing the number of probes that are classified into the five categories described in A in each of 606 modules, which are hierarchically clustered for identification of modules consisting of probes of genes, which are putatively lineage specific. The magnified image of the boxed area shows modules, each of which abundantly include probes of Triticeae-specific genes (category 4). Each number shown in each cell indicates the number of probes assigned to each category in each module. The arrowhead indicates module 243, which consists of 24 probes of putative Triticeae-specific genes. (C) The heat map visualization of gene expression patterns from Genevestigator on members in module 243 in anatomical barley tissues and (D) in developmental stages.
Triticeae species (Fig. 5B). According to the annotation of these probes, HD04G09u_at is the only probe showing weak similarity to the putative TNP-like transposable element of Sorghum, and all probes of this module have significant similarity to transcripts of barley and/or wheat (Table 5). Therefore, this module consists of probes of genes specific to Triticeae in terms of at least a currently available sequenced data set.

In order to investigate expression patterns of genes of module 243 in barley tissues, we used Genevestigator. The gene expression patterns along with anatomical tissues and developmental stages of barley showed preferential expression patterns in the spikelet and endosperm and in the early stage of seed development (Fig. 5C, D). These stage- and organ-specific expression patterns might suggest that module 243 is a network module playing specific roles in early stage development of barley seeds. Combinatorial use of comprehensive expression profiles across developmental stages and/or various types of tissues and the data set of co-expression analysis coupled with results of comparative analysis across plant lineages should allow us to narrow the targets to be analyzed with the aim of elucidating such potentially Triticeae-specific transcriptional modules. In Fig. 5B, most of the modules (538 modules) include at least one probe of genes in category 4. Some of these observations may be because of the fact that barley GeneChip probes were designed mainly from EST data, which are often partially sequenced. However, some of the probes might include transcriptional networks that consist of conserved genes as well as Triticeae-specific genes. The participation of species-specific genes may enable diversification of transcriptional regulatory networks as a form of evolution of cellular systems, and this could cause changes in the biological properties of each species.

Access to the data sets of the co-expressed gene network in barley. The data set of co-expressed genes and the annotations generated in this study should present useful information for promoting research in barley. Therefore, we developed a web site to provide access to co-expressed genes and related information for barley. The web site provides a web-based search interface to search for probes of genes on the barley GeneChip by using keywords as well as probe ID numbers and GO terms as query terms (Supplementary Fig. S2A).

<table>
<thead>
<tr>
<th>Probe name</th>
<th>Annotation from NetAffx</th>
<th>TAGI</th>
<th>HVGI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig12010_at</td>
<td>BAA93584.1.040 (AB038621) ORF3 [TT virus]</td>
<td>CA616952 4e-17</td>
<td>TC231445 6e-165</td>
</tr>
<tr>
<td>Contig16756_at</td>
<td></td>
<td>TC452975 7e-17</td>
<td>TC223861 0.0</td>
</tr>
<tr>
<td>Contig20688_at</td>
<td></td>
<td>TC232361 6e-115</td>
<td></td>
</tr>
<tr>
<td>Contig24192_at</td>
<td></td>
<td>TC224488 0.0</td>
<td></td>
</tr>
<tr>
<td>Contig25639_at</td>
<td></td>
<td>TC234474 4e-100</td>
<td></td>
</tr>
<tr>
<td>Contig25664_at</td>
<td></td>
<td>TC229565 3e-179</td>
<td></td>
</tr>
<tr>
<td>Contig26352_at</td>
<td></td>
<td>TC225671 0.0</td>
<td></td>
</tr>
<tr>
<td>Contig8098_at</td>
<td></td>
<td>TC230464 3e-114</td>
<td></td>
</tr>
<tr>
<td>EBem07_SQ004_i01_at</td>
<td></td>
<td>BM369750 5e-178</td>
<td></td>
</tr>
<tr>
<td>EBma08_SQ002_F06_at</td>
<td></td>
<td>BE422961 2e-13</td>
<td>BM371330 0.0</td>
</tr>
<tr>
<td>HB01M09r_at</td>
<td></td>
<td>TC203396 0.0</td>
<td></td>
</tr>
<tr>
<td>HB04I10r_x_at</td>
<td></td>
<td>BU967505 3e-53</td>
<td></td>
</tr>
<tr>
<td>HD02I19r_at</td>
<td></td>
<td>BQ662026 5e-106</td>
<td></td>
</tr>
<tr>
<td>HD04G09u_at</td>
<td>AAM94290.1 9e-05 (AF527807) putative TNP-like transposable element [Sorghum bicolor]</td>
<td>TC446423 1e-18</td>
<td>BQ658594 0.0</td>
</tr>
<tr>
<td>HD08L09r_at</td>
<td></td>
<td>TC210581 0.0</td>
<td></td>
</tr>
<tr>
<td>HF07C04r_at</td>
<td></td>
<td>TC236704 4e-75</td>
<td></td>
</tr>
<tr>
<td>HF09N12r_at</td>
<td></td>
<td>TC417890 1e-30</td>
<td>TC208251 2e-67</td>
</tr>
<tr>
<td>HF13L09r_at</td>
<td></td>
<td>TC417890 2e-25</td>
<td>TC235758 2e-143</td>
</tr>
<tr>
<td>HM05O11r_at</td>
<td></td>
<td>TC34268 3e-82</td>
<td></td>
</tr>
<tr>
<td>HM07G04r_at</td>
<td></td>
<td>TC218602 0.0</td>
<td></td>
</tr>
<tr>
<td>HM11M18r_at</td>
<td></td>
<td>TC210566 9e-23</td>
<td></td>
</tr>
<tr>
<td>HM11P04r_at</td>
<td></td>
<td>TC222706 0.0</td>
<td></td>
</tr>
<tr>
<td>HT10B20u_at</td>
<td></td>
<td>AJ602722 2e-14</td>
<td>CA010503 0.0</td>
</tr>
<tr>
<td>HZ01M13u_at</td>
<td></td>
<td>BQ665987 6e-175</td>
<td></td>
</tr>
</tbody>
</table>

TAGI, TIGR Gene Index of wheat searched by BLASTN with a threshold of e-value <1e-10.
HVGI, TIGR Gene Index of barley searched by BLASTN with a threshold of e-value <1e-10.
Users can browse related information including results of similarity searches against protein data sets of Arabidopsis, rice and Brachypodium and against clustered transcripts of barley and wheat used in this study. Hyperlinks to Genevigator enable users to browse for expression patterns and obtain lists of associated probes that show correlated expression (Supplementary Fig. S2B, S2C). The web site also provides a user interface for downloading archived data sets of the co-expressed barley genes (Supplementary Fig. S2D). The web site is available to the public at http://coexpression.psc.riken.jp/barley/.

Co-expression analysis has become one of the most significant approaches based on large-scale data sets of the transcriptome of various species. Recently, transcriptome analysis has been widely expanded as a key platform for investigations of biological properties based on transcriptome changes. For instance, microarray analyses coupled with laser capture microdissection or with fluorescence-activated cell sorting have been applied to monitor the transcriptome of specific cells of interest in plants (Birnbaum et al. 2005, Ohtsu et al. 2007). These approaches were used, for example, in studies of cell differentiation of Arabidopsis root, and of transcriptomes of the male gametophyte and tapetum in rice (Brady et al. 2007, Suwabe et al. 2008). Furthermore, transcriptome data have been recently collected from plants grown in field environments (Sato et al. 2011). Such microscopic and macroscopic collections of transcriptome data sets and efficient integration with other information resources will be essential for improving our understanding of cellular systems associated with biological properties of plants.

In this study, we demonstrated that efficient use of large-scale transcriptome data sets collected for barley can be used to construct a co-expression network for gene discovery. Although comparative genomics should aid in predicting gene function based on significant similarity against well analyzed genes in model plants, the functions of several genes remain unknown. Co-expression network analysis should provide clues for development of hypotheses in the determination of gene functions from the relationships with other known genes classified in the same modules. According to the recent accumulation of data sets from genomic resources available in Gramineae species, the number of transcriptome data sets of barley will also increase. The barley genome project is ongoing (http://barleygenome.org/). After the completion of the barley genome project, we will have opportunities to discover regulatory networks from not only transcriptome profiles but also promoter motifs featured in the members of co-expressed genes to elucidate regulatory networks of cellular systems of interest combined with genome-wide profiles of chromatin modification and/or binding of transcription factors by Chip-seq. Furthermore, combinatorial approaches together with metabolomics efforts should result in more attention being focused on barley for molecular elucidation of gene expression networks coupled with metabolic networks. This will accelerate the discovery of genes that play important roles in practical molecular breeding in Triticeae crops. The combinatorial use of other genomic resources such as high-density genetic maps, mutants and molecular phenotypes will also provide integrated information resources for barley. Combinatorial approaches integrating multiple omics researches in crops should play important roles to promote plant systems breeding toward sustainable agricultural production as well as green innovation based on plant functionalities.

Materials and Methods

Raw expression data

The barley transcriptome data set for the construction of the co-expression network was obtained from the barley1 GeneChip data in the PLEXdb, the data set of platform accession number GPL1340 in the NCBI Gene Expression Omnibus (GEO) and from the array design identifier A-AFFY-31 in EBI ArrayExpress (Wise et al. 2007, Barrett et al. 2011, Parkinson et al. 2011). The array platform for the retrieved data set consists of experimental samples from the assay using the Affymetrix GeneChip Barley Genome [Barley1] (http://www.affymetrix.com/estore/browse/products.jsp?productid=131420 &categoryID=35878). The redundant data sets among these three databases were unified. The data sets of cross-species hybridization, overexpressor and genotyping were manually removed. In total, 1,347 CEL files of samples in the barley1 GeneChip data remained for construction of the network (Supplementary Table S1). These CEL files were normalized using the MAS 5.0 algorithm by the Affy library in the Bioconductor package. The control probes from the barley1 GeneChip were removed from the samples. Then, the MAD (the median of the absolute deviations from the median) and the mean of the signal intensities of the probe sets among samples were normalized from 1 to 0. The data set was used to construct co-expression data.

Co-expression analysis

Possible redundancies and biases based on PCCs between samples were calculated to avoid biased correlation due to the large series of similar samples. This method was originally applied to calculations of gene co-expression data in ATTED-II (Obayashi et al. 2007). The weight value for each sample was calculated based on sample redundancy derived as the summation of the pair-wise sample redundancies between the sample and each of the samples including the sample itself. The weight value of each sample was used to calculate the weighted PCC (wPCC) of a pair of gene expression patterns (Obayashi and Kinoshita 2009). The number of probes and links, and network density were calculated along with the wPCC cut-off to evaluate the general topology of the barley co-expression network and to determine the cut-off threshold of wPCC for network analysis. Network density is calculated as follows: (number of observed links)/(number of possible maximum links).
Database for prediction of gene function
The GO annotations of probes of the barley1 GeneChip were provided according to their sequence similarity to genes annotated in Arabidopsis and Brachypodium. The GO annotation of Arabidopsis genes was derived from TAIR9 (Swarbreck et al. 2008). The GO annotations of Brachypodium genes were derived from the results of an InterProScan search of the protein data set of the Bd1 genome annotation. A similarity search to identify homologous genes in Arabidopsis and Brachypodium was performed by an NCBI BLAST using blastp with a threshold e-value $\leq 1 \times 10^{-10}$ against the data set of target sequences of the barley 1 GeneChip as the query and the protein data sets of each species as the database. A similarity search to identify a corresponding transcript in barley and a putative homologous transcript in wheat was performed by conducting an NCBI BLAST using blastx with a threshold e-value $\leq 1 \times 10^{-10}$ against a clustered transcript data set of barley and wheat from the TIGR Gene Index (Lee et al. 2005). All results of probe annotation are provided in Supplementary Table S3 and at http://coexpression.psc.riken.jp/barley/. This website includes a search interface.

Network analysis
To identify subnetwork modules from the co-expression network data set, the NeMo plug in of Cytoscape was applied (Rivera et al. 2010). Possible co-expression network modules were estimated by the NeMo plug in. Then, all modules including at least one link were clustered using the single linkage method to remove the redundancy of members.

Enrichment analysis of GO terms
To annotate the gene function of the identified co-expression subnetwork modules, we performed an enrichment analysis of GO terms. The TopGO package (http://www.bioconductor.org/packages/2.5/bioc/html/topGO.html) of the Bioconductor was applied to perform the enrichment analysis coupled with GO annotation data of barley probes generated from the homology search of the Brachypodium and Arabidopsis proteins. The summarization of GO terms enriched in each module into GO Slim terms was performed by using the GOSlimViewer web service in AgBase (McCarthy et al. 2011).

Supplementary data
Supplementary data are available at PCP online.

Funding
This work was supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan [Grant-in-Aid for Young Scientists (B) (21780011) to K.M.]; Bio-oriented Technology Research Advancement Institution (BRAIN) [the Research and Development Program for New Bio-industry Initiatives].

Acknowledgments
The authors thank Dr. Lam-Son Phan Tran of the RIKEN Plant Science Center for helpful discussions. We also thank Y. Nishizuka of the RIKEN Plant Science Center for technical assistance.

References
Druka, A., Muehlbauer, G., Druka, I., Caldo, R., Baumann, U., Rostoks, N. et al. (2006) An atlas of gene expression from seed to


