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

Keiichi Mochida[1,2,*], Yukiko Uehara-Yamaguchi[1,2], Takuhiro Yoshida[2], Tetsuya Sakurai[2] and Kazuo Shinozaki[1,2]

[1]RIKEN Biomass Engineering Program, Yokohama, 230-0045 Japan
[2]RIKEN 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 Triticeae-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 Triticeae 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

(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 *RHM1* (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

Accession No.	Series name	DB	No. of CEL files
BB2	Expression profiling of wild type and mutants of Sultan 5 (Mla12) barley cultivar	PLEXDB	180
BB3	Transcription patterns during barley development	PLEXDB	63
BB4	Mla-specified transcriptional responses in barley–powdery mildew interactions	PLEXDB	108
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 7 d 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
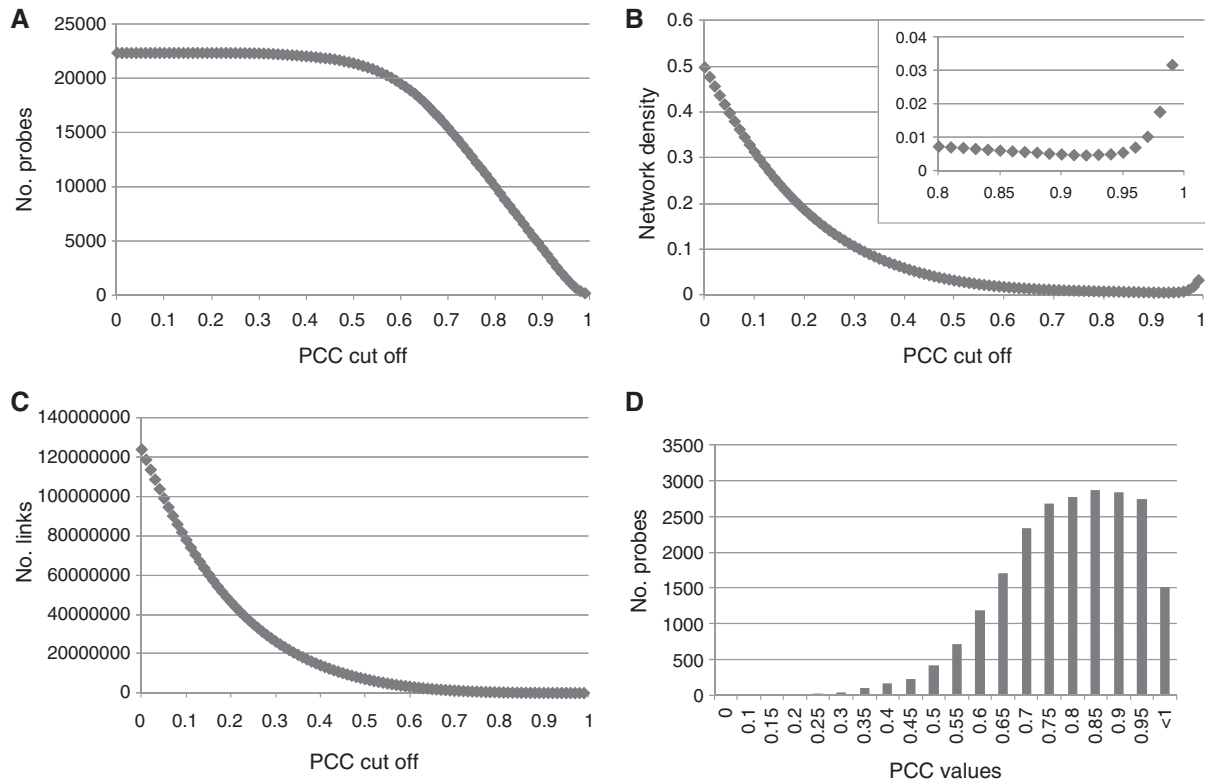
Fig. 1 Overview of the global features of the barley co-expression network. The number of probes and links, and network density were calculated along with the wPCC cut-off from 0 to 0.99 with an interval of 0.01. Network density is calculated by dividing the number of observed links by the number of possible maximum links. (A) The number of probes in the entire network at the positive PCC cut-off value. (B) The number of links in the entire network at the positive PCC cut-off value. (C) Network density in the entire network at the positive PCC cut-off value. The boxed plot shows a magnification of the area from the PCC cut-off range from 0.8 to 0.99. (D) Distributions of PCC values between each probe and its probes of strongest co-expressed gene.

minimal at a cut-off value of 0.92 of the PCC. This indicates that the PCC value was the cut-off value 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 transcriptomic 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 Triticeae 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).
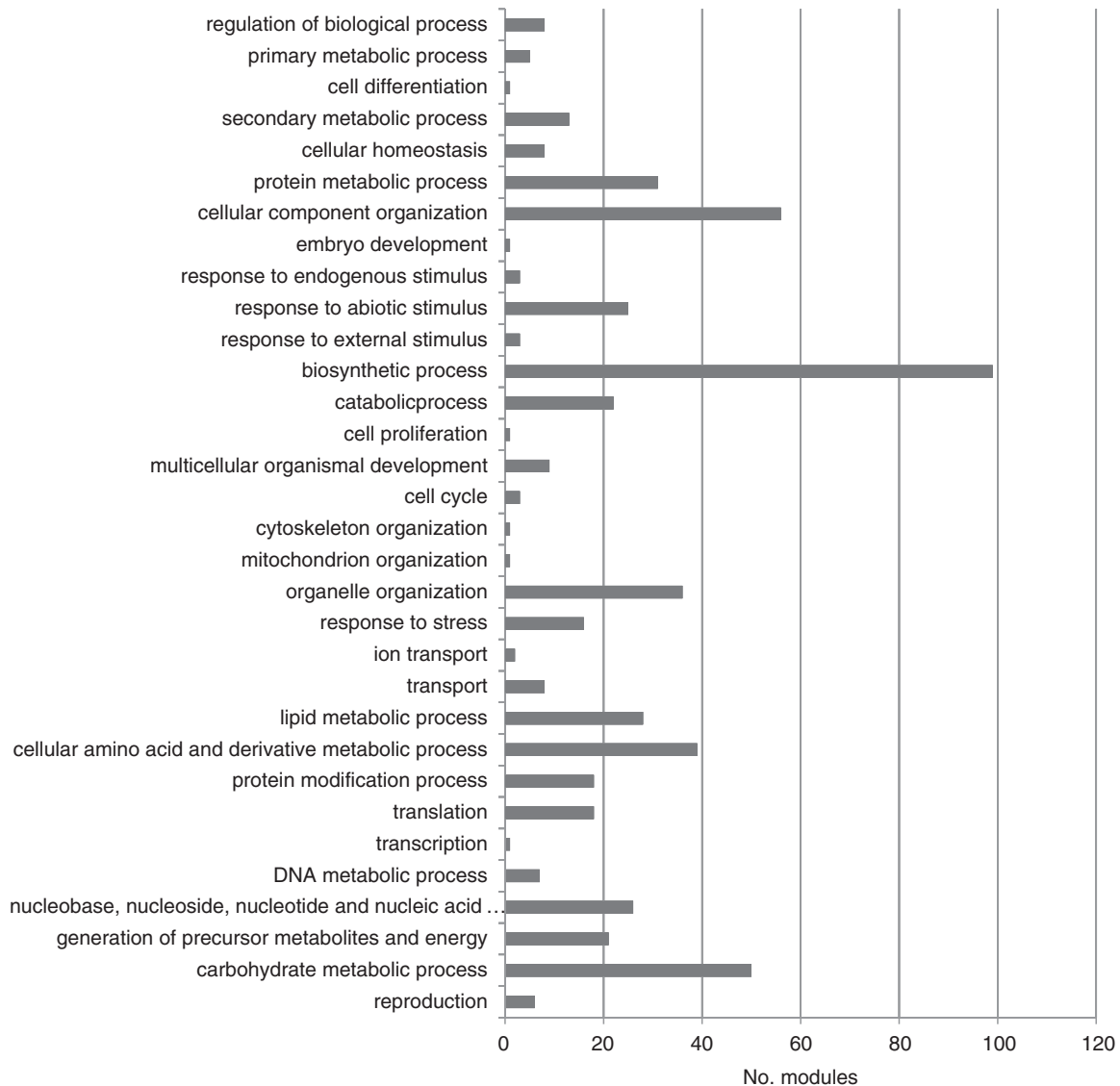
Fig. 2 Distribution of the GO category in the biological process significantly enriched in a co-expression network module based on topGO analysis with a cut-off threshold of $P < 1 \times 10^{-5}$.

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 hyper-cholesterolemia, 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.
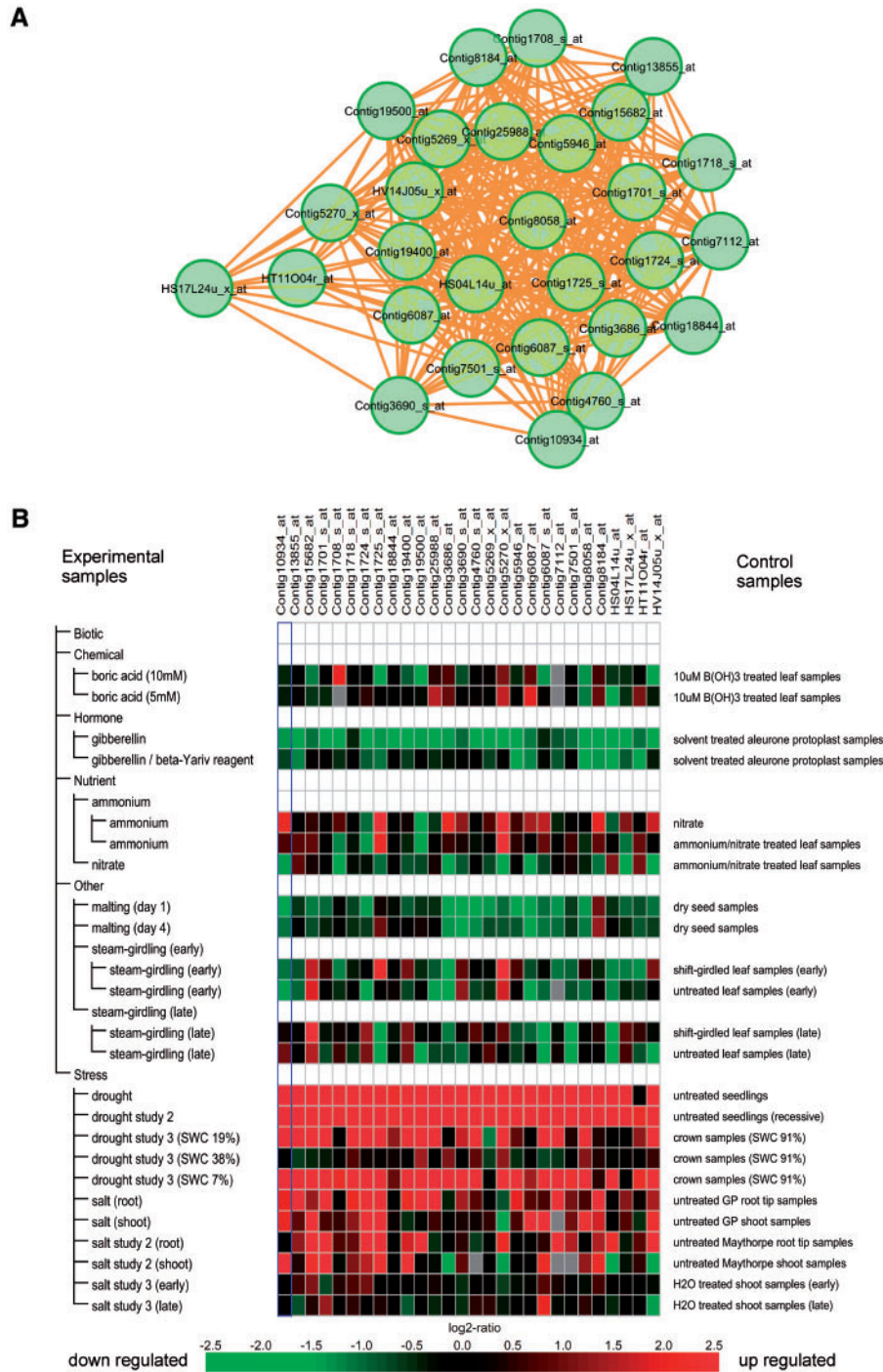
Fig. 3 A predicted co-expression network module in response to dry stress conditions in barley. (A) A graph showing the genes in the network module that respond to dry stress conditions. (B) Heat map of the expression pattern of each gene in the network module generated using Genevestigator.

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

Table 2 Predicted gene functions and sequence identifiers of best hit counterparts in other plant species of the probes of members in module 406

Probe name	Annotation from NetAffx	Bdi	Osa	Ath	TAGI	HVGI
Contig10934_at	AAN06848.1 2e-16 (AC099401) Putative ABA-induced protein [*Oryza sativa* (japonica cultivar group)]	Bradi4g07340.1 5e-32	Os06g0341300|AK107654 1e-32	AT3G22490.1 2e-16	TC416163 2e-155	TC223621 0.0
Contig13855_at	AAN05568.1 1e-29 hypothetical protein [*Oryza sativa* (japonica cultivar group)]	Bradi3g30320.1 5e-34	Os10g0505900|AK062588 6e-24		AW448829 5e-175	TC226835 0.0
Contig15682_at					TC439492 2e-34	TC231237 1e-122
Contig1701_s_at	AAD02254.1 1e-30 (AF043088) dehydrin 2 [*Hordeum vulgare*] [*Hordeum vulgare* subsp. *vulgare*]	Bradi4g22290.1 3e-29	Os11g0453900|AK109096 3e-24	AT5G66400.1 4e-20	TC381055 8e-149	TC211167 0.0
Contig1708_s_at	AAF01694.1 5e-27 (AF181456) dehydrin; DHN6 [*Hordeum vulgare*] [*Hordeum vulgare* subsp. *vulgare*]	Bradi4g22290.1 2e-30	Os11g0454300|AK121952 4e-30	AT5G66400.1 1e-22	TC395863 5e-120	TC197204 0.0
Contig1718_s_at	AAD02260.1 3e-30 (AF043094) dehydrin 9 [*Hordeum vulgare*] [*Hordeum vulgare* subsp. *vulgare*]	Bradi3g43870.1 3e-39	Os11g0454000|AK071366 2e-33	AT5G66400.1 1e-20	CJ596275 2e-103	TC232759 0.0
Contig1724_s_at	AAD02255.1 6e-19 (AF043089) dehydrin 3 [*Hordeum vulgare*] [*Hordeum vulgare* subsp. *vulgare*]	Bradi1g37410.1 6e-51	Os11g0454300|AK121952 7e-37	AT5G66400.1 2e-18	TC371537 0.0	TC195363 0.0
Contig1725_s_at	P12951 3e-29 DEHYDRIN DHN1 (B8) pir||S05544 dehydrin 8—barley	Bradi1g37410.1 2e-41	Os11g0454300|AK121952 5e-41	AT5G66400.1 8e-24	TC420367 0.0	TC198977 0.0
Contig18844_at	BAB39231.1 1e-79 (AP002869) hypothetical protein—similar to *Arabidopsis thaliana* chromosome 5	Bradi2g12390.1 4e-13	Os01g0329400|AK101813 3e-13	AT5G15640.1 7e-07	TC403160 6e-141	TC198553 0.0
Contig19400_at						TC235155 7e-83
Contig19500_at					TC447298 7e-27	TC233994 5e-90
Contig25988_at					TC423324 1e-41	TC233176 8e-152
Contig3686_at	T06978 4e-80 ABA-induced plasma membrane protein PM 19—wheat gb|AAB38504.1| ABA- induced plasma membrane protein PM 19 [*Triticum aestivum*]	Bradi1g00600.1 2e-21	Os05g0381400|AK102039 3e-15	AT1G04560.1 2e-11	TC414279 2e-41	TC231158 0.0
Contig3690_s_at	AAF29532.1 5e-36 plasma membrane-associated protein [*Hordeum vulgare*]				TC409808 5e-23	TC195357 0.0
Contig4760_s_at	BAB32715.1 1e-67 putative late embryogenesis abundant protein LEA14-A [*Oryza sativa* (japonica cultivar group)]	Bradi2g07480.1 4e-59	Os01g0225600|Os01g0225600 3e-58	AT1G01470.1 2e-39	TC393160 0.0	TC195540 0.0
Contig5269_x_at	NP_181073.1 4e-04 (NM_129082) similar to late embryogenesis abundant proteins; protein ID At2g35300.1	Bradi5g19870.1 3e-23	Os04g0589800|AK063682 2e-21	AT2G35300.1 2e-11	CA701261 1e-97	TC228899 2e-149
Contig5270_x_at					TC393321 4e-41	TC231848 1e-113
Contig5946_at	AAL84288.1 3e-04 (AC073556) putative embryo-specific protein [*Oryza sativa* (japonica cultivar group)]	Bradi1g73840.1 2e-19	Os03g0168100|AK121575 4e-13	AT2G18340.1 1e-08	TC383443 3e-66	TC208071 2e-111
Contig6087_at	NP_196350.1 1e-21 putative protein; protein ID: At5g07330.1 [*Arabidopsis thaliana*] pir||T49872 hypothetical protein T2I1.40—*Arabidopsis thaliana*					TC202822 6e-35

(continued)

Table 2 Continued

| Probe name	Annotation from NetAffx	Bdi	Osa	Ath	TAGI	HVGI
Contig6087_s_at	NP_196350.1 1e-21 putative protein; protein ID: At5g07330.1 [Arabidopsis thaliana] pir||T49872 hypothetical protein T2I1.40—Arabidopsis thaliana	Bradi4g17200.1 7e-27	Os11g0533400|AK063652 4e-20		TC381929 2e-152	TC202822 0.0
Contig7112_at	NP_201479.1 9e-19 putative protein; protein ID: At5g66780.1	Bradi1g10310.1 3e-35	Os03g0723400|AK107276 2e-36	AT5G66780.1 2e-20	TC419151 3e-130	TC218196 0.0
Contig7501_s_at	AAL77132.1 7e-82 Putative calcium-binding protein [Oryza sativa]	Bradi3g22640.1 4e-46	Os10g0177200|AK064016 3e-45	AT4G38810.1 9e-24	TC382986 0.0	TC206394 0.0
Contig8058_at	Q00747 7e-07 PROTEIN LE25 pir||S19253 gene le25 protein—tomato	Bradi4g01020.1 4e-30	Os06g0324400|AK063726 7e-25	AT5G06760.1 1e-07	TC379189 1e-126	TC204547 0.0
Contig8184_at	BAB44029.1 8e-48 WSI18 protein [Oryza sativa (japonica cultivar-group)] dbj|BAB86507.1| WSI18 protein [Oryza sativa (japonica cultivar-group)]	Bradi2g47700.1 4e-37	Os01g0705200|AK064074 1e-31	AT1G52690.1 4e-18	TC395010 0.0	TC216162 0.0
HS04L14u_at	BAB21159.1 .004 (AP002899) contains ESTs AU101349(E11470)				TC420953 2e-14	BQ662909 5e-72
HS17L24u_x_at		Bradi5g19880.1 4e-16	Os04g0589800|AK063682 4e-16	AT2G35300.1 4e-08	TC393321 5e-51	TC228899 2e-113
HT11O04r_at	P14928 1e-16 ABA-inducible protein PHV A1 pir||S08313 ABA-induced protein HVA-1—barley	Bradi2g18100.2 1e-12	Os01g0705200|AK064074 4e-08		TC391553 1e-52	TC195330 0.0
HV14J05u_x_at					TC393321 1e-28	TC228899 2e-92

Bdi, protein data set annotated on the *Brachypodium* genome, Bdi1.0 seached by BLASTP with a threshould of e-value <1e-10.
Osa, protein data set annotated on the rice genome, RAP2 seached by BLASTP with a threshould of e-value <1e-10.
Ath, protein data set annotated on the Arabidopsis genome, TAIR9 seached by BLASTP with a threshould 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.
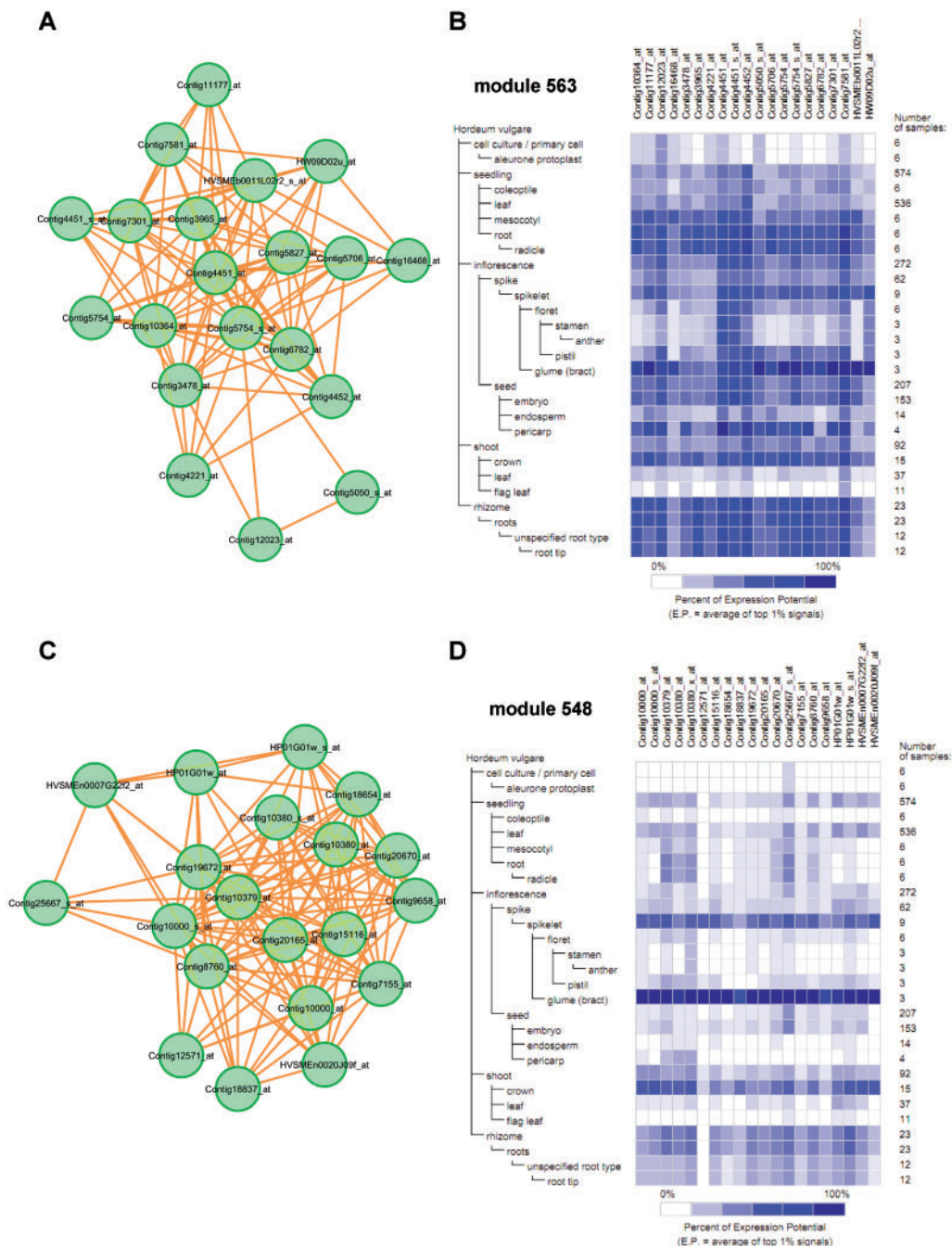
Fig. 4 Predicted co-expression network modules involved in cellulose biogenesis. (A) Module 563 consists of 20 nodes with 99 edges (PCC average = 0.802). It has also been determined that module 548 putatively involves cell wall biogenesis. (C) The module consists of 20 probes with 127 edges (PCC average = 0.816). (B) Heat map visualization of gene expression patterns from Genevestigator on members in module 563 and (D) module 548.

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 plumbaginifolia*, the mutation of *NpGUT1* 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 *NpGUT1* 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

Table 3 Predicted gene functions and sequence identifiers of best hit counterparts in other plant speceis of the probes of members in module 563

Probe name	Annotation from NetAffx	Bdi	Osa	Ath	TAGI	HVGI
Contig10364_at	NP_191301.1 9e-83 (NM_115602) putative protein; protein id: At3g57420.1 [Arabidopsis thaliana]	Bradi2g26610.1 1e-23	Os05g0391200|AK059645 6e-20	AT3G57420.1 4e-09	TC402249 0.0	TC198747 0.0
Contig11177_at	AAL77142.1 2e-43 (AC097447) Putative putative myosin-like protein [Oryza sativa]	Bradi1g68460.1 2e-14	Os10g0162400|AK064809 3e-11	TC418566 0.0	TC218109 0.0	
Contig12023_at	NP_177870.1 1e-69 (NM_106395) unknown protein; protein ID: At1g77460.1 [Arabidopsis thaliana]	Bradi1g45400.1 9e-55	Os06g0223800|AK105686 3e-54	AT2G22125.1 2e-46	TC399453 0.0	TC204331 0.0
Contig16468_at	NP_181761.1 4e-41 (NM_129794) unknown protein; protein ID: At2g42320.1 [Arabidopsis thaliana]	Bradi3g39880.1 1e-55	Os08g0505200|AK067190 5e-47	AT2G42320.1 2e-20	TC386351 1e-178	TC229472 0.0
Contig3478_at	AAF89964.1 e-131 cellulose synthase-4 [Zea mays]	Bradi1g54250.1 2e-47	Os07g0208500|AK072356 3e-46	AT5G05170.1 7e-45	TC369453 0.0	TC231134 0.0
Contig3965_at	BAC20928.1 1e-67 pectin-glucuronyltransferase [Nicotiana plumbaginifolia]	Bradi2g59380.1 3e-45	Os01g0926400|AK073976 6e-42	AT5G61840.1 1e-35	TC422805 9e-152	TC237466 0.0
Contig4221_at	AAF24189.1 e-123 phytochelatin synthetase-like protein [Zea mays]	Bradi1g08130.1 1e-14	AK102170|AK102170 2e-12		TC372535 3e-170	TC229966 0.0
Contig4451_at	AAF89963.1 e-124 cellulose synthase-3 [Zea mays]	Bradi2g34240.1 3e-42	Os05g0176100|AK100188 4e-41	AT5G17420.1 5e-33	TC383187 1e-162	TC220374 0.0
Contig4451_s_at	AAF89963.1 e-124 cellulose synthase-3 [Zea mays]				TC383187 5e-35	TC220374 4e-147
Contig4452_at	AAF89961.1 e-126 cellulose synthase-1 [Zea mays]	Bradi2g34240.1 9e-18	Os05g0176100|AK100188 7e-17	AT5G17420.1 5e-10	TC383550 1e-103	TC195373 1e-169
Contig5050_s_at	AAL83695.1 3e-17 (AC092263) hypothetical protein [Oryza sativa (japonica cultivar group)]	Bradi1g00560.2 7e-38	Os03g0861100|AK100706 2e-35	AT5G43310.1 5e-11	TC382479 7e-156	TC204688 0.0
Contig5706_at	AAL38530.1 e-121 CSLF6 [Oryza sativa]		Os08g0160500|AK065259 4e-13	TC374481 9e-100	TC212235 0.0	
Contig5754_at	NP_567774.1 1e-50 Expressed protein; protein ID: At4g27435.1	Bradi1g28490.1 8e-43	Os07g0462200|AK059964 4e-39	AT3G15480.1 2e-21	TC406502 2e-122	TC197865 0.0
Contig5754_s_at	NP_567774.1 1e-50 Expressed protein; protein ID: At4g27435.1	Bradi1g28490.1 5e-13	Os07g0462200|AK059964 1e-13	TC370973 3e-64	TC197865 2e-174	
Contig5827_at	BAC15948.1 1e-76 (AP003847) contains EST—hypothetical protein—similar to Arabidopsis thaliana chromosome 2	Bradi1g06290.1 1e-25	Os03g0788600|AK069046 4e-24	AT2G35860.1 4e-11	TC381073 1e-132	TC197354 0.0
Contig6782_at	BAB84493.1 3e-23 (AP004194) putative arabinogalactan-like protein [Oryza sativa (japonica cultivar group)]	Bradi2g16560.1 2e-32	Os05g0563600|AK099313 6e-34	AT1G44191.1 5e-14	TC410644 3e-146	TC200341 0.0
Contig7301_at	BAA90502.1 2e-90 (AP011111) maize EST AI621709	Bradi2g37720.1 7e-29	Os05g0119700|AK062482 3e-21	AT4G27595.1 6e-06	TC421218 0.0	TC203280 0.0
Contig7581_at	NP_683337.1 e-107 UDP-N-acetylglucosamine pyrophosphorylase-like protein; protein ID: At1g31070.2	Bradi5g21650.1 6e-49	Os04g0613700|AK064783 4e-50	AT2G35020.1 2e-44	TC373965 2e-168	TC199628 0.0
HVSMEb0011L02r2_s_at					TC389410 6e-23	TC229472 3e-41
HW09D02u_at	BAC10706.1 9e-12 P0431H09.18 [Oryza sativa (japonica cultivar group)]	Bradi2g47280.1 4e-16	Os01g0694900|AK063321 3e-19	TC418753 3e-105	AL502930 0.0	

Bdi, protein data set annotated on the *Brachypodium* genome, Bdi1.0 seached by BLASTP with a threshould of e-value <1e-10.
Osa, protein data set annotated on the rice genome, RAP2 seached by BLASTP with a threshould of e-value <1e-10.
Ath, protein data set annotated on the Arabidopsis genome, TAIR9 seached by BLASTP with a threshould of e-value <1e-5.
TAGl, 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 cellulosic 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 HvCslF 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 (Zhou 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.

Triticeae-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 Triticeae 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 Triticeae species.

We then applied the categorized probe data set of 8,586 genes classified into 606 network modules to screen for potentially Triticeae-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 Triticeae, 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 Triticeae-specific modules. For instance, module 243 consists of 24 probes that are in category 4. These probes are potentially specific to

Table 4 Predicted gene functions and sequence identifers of best hit counterparts in other plant speceis of the probes of members in module 548

| Probe name	Annotation from NetAffx	Bdi	Osa	Ath	TAGI	HVGI
Contig10000_at	BAA96759.1 4e-44 (AP002521) EST D25138(R3286) corresponds to a region of the predicted gene—similar to Pinus taeda	Bradi2g00220.1 1e-35	Os05g0163300|AK105598 8e-32	AT5G03170.1 6e-22	TC385361 2e-120	TC218744 7e-168
Contig10000_s_at	BAA96759.1 4e-44 (AP002521) EST D25138(R3286) corresponds to a region of the predicted gene—similar to Pinus taeda				CD877212 3e-20	TC218744 2e-111
Contig10379_at	BAB16875.1 8e-07 (AP002537) hypothetical protein [Oryza sativa (japonica cultivar group)]				TC412806 2e-149	TC218557 0.0
Contig10380_at	BAB16875.1 3e-55 (AP002537) hypothetical protein [Oryza sativa (japonica cultivar group)]	Bradi2g05930.1 1e-60	Os05g0199100|AK069019 1e-36	AT1G17620.1 1e-15	TC412806 0.0	TC218557 0.0
Contig10380_x_at	BAB16875.1 3e-55 (AP002537) hypothetical protein [Oryza sativa (japonica cultivar group)]	Bradi2g05930.1 4e-82	Os05g0199100|AK069019 2e-52	AT1G17620.1 7e-27	TC412806 0.0	TC218557 0.0
Contig12571_at	AAM74394.1 6e-62 (AC119149) Putative cytochrome 450 [Oryza sativa (japonica cultivar group)]	Bradi4g16560.1 2e-62	Os10g0317900|AK070442 1e-60	AT5G07990.1 2e-44	CK205504 0.0	TC210157 0.0
Contig15116_at	BAB67900.1 4e-42 (AP003237) putative cellulose synthase [Oryza sativa (japonica cultivar group)]	Bradi2g49910.1 3e-33	Os01g0750300|AK100475 1e-29	AT4G18780.1 1e-24	TC392907 6e-125	TC224574 0.0
Contig18654_at	AAD21424.1 1e-05 (AC005882) 69873 [Arabidopsis thaliana]	Bradi3g15340.1 7e-20	Os06g0681300|AK110977 1e-10		BJ261786 5e-95	TC203228 0.0
Contig18837_at	AAB17192.1 7e-22 (U73104) laccase [Liriodendron tulipifera]	Bradi2g23350.1 1e-22	Os05g0458600|AK068047 8e-23	AT5G60020.1 3e-21	TC420956 7e-97	GH228456 0.0
Contig19672_at	NP_187609.2 5e-27 unknown protein; protein ID: At3g09980.1	Bradi5g09130.1 3e-54	Os04g0408600|AK121434 2e-51	AT3G09980.1 6e-37	TC405047 0.0	TC218506 0.0
Contig20165_at	AAM26299.1 3e-67 (AY095297) cellulose synthase [Populus tremuloides]	Bradi4g30540.1 3e-77	Os09g0422500|AK121170 6e-75	AT5G17420.1 5e-70	CV767146 0.0	TC231301 0.0
Contig20670_at	XP_129042.1 4e-11 (XM_129042) similar to ring finger protein [Gallus gallus] [Mus musculus]	Bradi5g04540.1 2e-41	Os04g0243700|AK059146 1e-35	AT5G66160.2 2e-11	TC456103 1e-12	TC210145 0.0
Contig25667_s_at						BJ451823 2e-117
Contig7155_at	BAC01247.1 2e-50 P0019E03.5 [Oryza sativa (japonica cultivar group)]	Bradi2g17070.1 2e-59	Os05g0556400|AK058980 3e-57	AT3G61750.1 5e-33	TC427501 9e-180	TC229948 0.0
Contig8760_at	BAA90810.1 2e-71 (AP001168) ESTs AU082174(S13676)		Os08g0546100|Os08g0546100 6e-06	TC391122 2e-45	TC206614 0.0	
Contig9658_at	AAK27814.1 e-116 putative cellulose synthase [Oryza sativa (japonica cultivar group)]	Bradi3g28350.1 7e-66	Os10g0467800|AK072259 4e-63	AT5G44030.1 3e-60	TC387827 0.0	TC197870 0.0
HP01G01w_at	AAK15451.1 1e-18 (AC037426) unknown protein [Oryza sativa (japonica cultivar group)]	Bradi3g33070.1 1e-21	Os10g0554900|AK120846 7e-20		TC450405 9e-74	TC230675 2e-159
HP01G01w_s_at	AAK15451.1 1e-18 (AC037426) unknown protein [Oryza sativa (japonica cultivar group)]	Bradi3g33070.1 4e-10	Os10g0554900|AK120846 1e-09		TC450405 1e-87	TC230675 7e-143
HVSMEn0007G22f2_at	AAK15451.1 8e-15 (AC037426) unknown protein [Oryza sativa (japonica cultivar group)]	Bradi3g33070.1 1e-17	Os10g0554900|AK120846 8e-16		TC450405 9e-75	TC230675 0.0
HVSMEn0020J09f_at					TC218744 2e-41	

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 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.
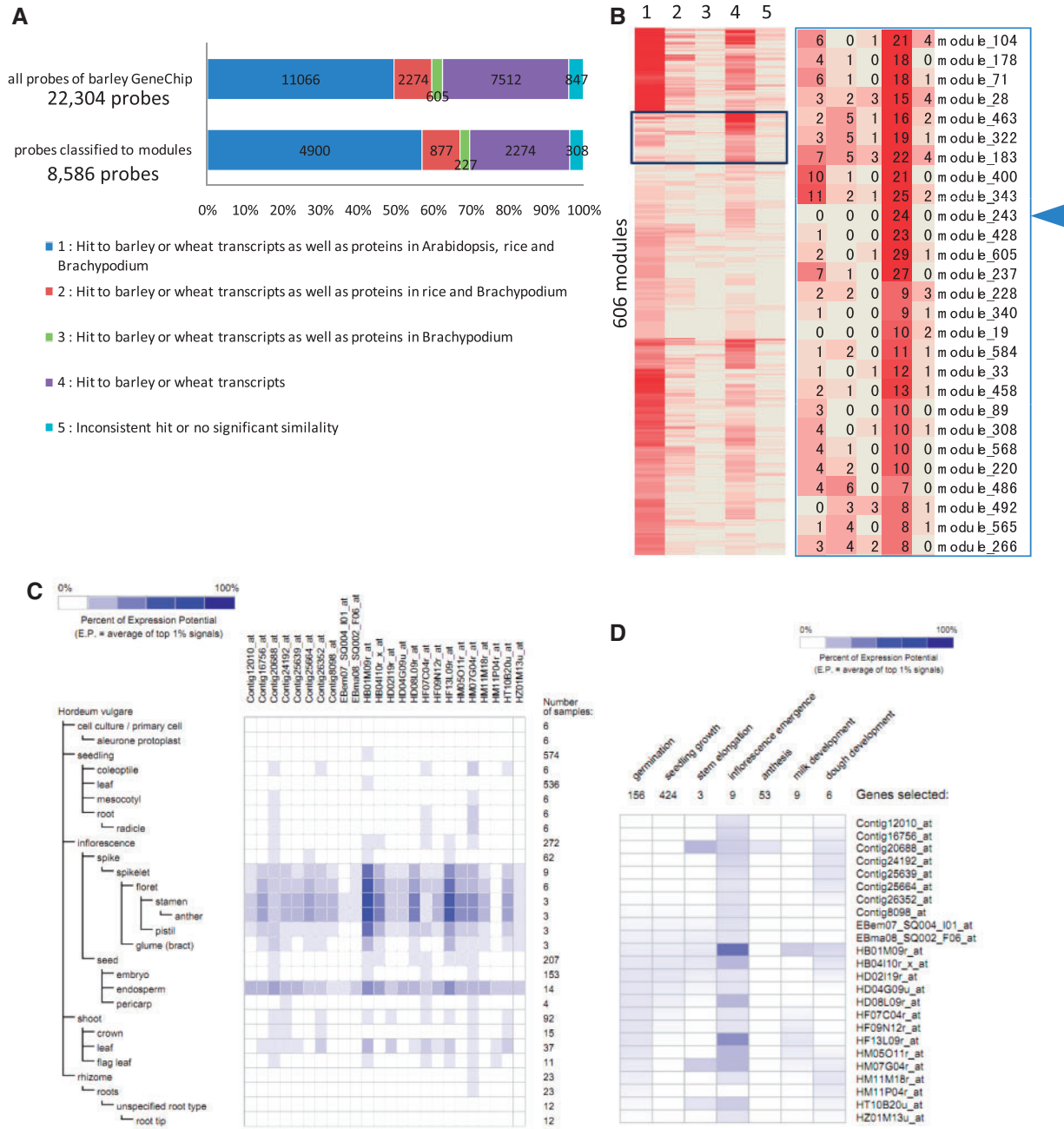
Fig. 5 Identification of putative lineage-specific network modules. (A) A bar graph showing numbers and proportions of the probes for all probes analyzed (22,043) and probes classified in modules (8,586). The graph indicates significant similarity among transcripts of barley and wheat, and/or proteins 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 of *Arabidopsis*, rice and *Brachypodium*, (2) hit to barley or wheat transcripts as well as proteins of rice and *Brachypodium*, (3) hit to barley or wheat transcripts as well as proteins of *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.

Table 5 Similarity search results of probes of target seqeunces in module 243 against protein data sets of *Brachypodium*, rice, Arabidopsis and transcript sequence data sets of wheat and barley

Probe name	Annotation from NetAffx	TAGI	HVGI
Contig12010_at	BAA93584.1 .040 (AB038621) ORF3 [TT virus]	CA616952 4e-17	TC231445 6e-165
Contig16756_at		TC452975 7e-17	TC223861 0.0
Contig20688_at			TC232361 6e-115
Contig24192_at			TC224488 0.0
Contig25639_at			TC234474 4e-100
Contig25664_at			TC229565 3e-179
Contig26352_at			TC225671 0.0
Contig8098_at			TC230464 3e-114
EBem07_SQ004_I01_at			BM369750 5e-178
EBma08_SQ002_F06_at		BE422961 2e-13	BM371330 0.0
HB01M09r_at			TC203396 0.0
HB04I10r_x_at			BU967505 3e-53
HD02I19r_at			BQ462026 5e-106
HD04G09u_at	AAM94290.1 9e-05 (AF527807) putative TNP-like transposable element [*Sorghum bicolor*]	TC446423 1e-18	BQ658594 0.0
HD08L09r_at			TC210581 0.0
HF07C04r_at			TC236704 4e-75
HF09N12r_at		TC417890 1e-30	TC208251 2e-67
HF13L09r_at		TC417890 2e-25	TC235158 2e-143
HM05O11r_at			TC234268 3e-82
HM07G04r_at			TC218602 0.0
HM11M18r_at			TC210566 9e-23
HM11P04r_at			TC222706 0.0
HT10B20u_at		AJ602722 2e-14	CA010503 0.0
HZ01M13u_at			BQ665987 6e-175

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.

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**).

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 Genevestigator 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 Bdi1 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 $<1 \times 10^{-10}$ between 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 blastn with a threshold e-value $<1 \times 10^{-10}$ 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.

References

Alexa, A., Rahnenfuhrer, J. and Lengauer, T. (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22: 1600–1607.

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

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F. et al. (2011) NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res.* 39: D1005–D1010.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C. et al. (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* 37: D885–D890.

Birnbaum, K., Jung, J.W., Wang, J.Y., Lambert, G.M., Hirst, J.A., Galbraith, D.W. et al. (2005) Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat. Methods* 2: 615–619.

Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinneny, J.R. et al. (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 318: 801–806.

Brady, S.M. and Provart, N.J. (2009) Web-queryable large-scale data sets for hypothesis generation in plant biology. *Plant Cell* 21: 1034–1051.

Burton, R.A., Collins, H.M., Kibble, N.A., Smith, J.A., Shirley, N.J., Jobling, S.A. et al. (2011) Over-expression of specific HvCslF cellulose synthase-like genes in transgenic barley increases the levels of cell wall (1,3;1,4)-β-D-glucans and alters their fine structure. *Plant Biotechnol. J.* 9: 117–135.

Burton, R.A., Jobling, S.A., Harvey, A.J., Shirley, N.J., Mather, D.E., Bacic, A. et al. (2008) The genetics and transcriptional profiles of the cellulose synthase-like HvCslF gene family in barley. *Plant Physiol.* 146: 1821–1833.

Burton, R.A., Wilson, S.M., Hrmova, M., Harvey, A.J., Shirley, N.J., Medhurst, A. et al. (2006) Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-β-D-glucans. *Science* 311: 1940–1942.

Close, T.J., Bhat, P.R., Lonardi, S., Wu, Y., Rostoks, N., Ramsay, L. et al. (2009) Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10: 582.

Close, T.J., Wanamaker, S.I., Caldo, R.A., Turner, S.M., Ashlock, D.A., Dickerson, J.A. et al. (2004) A new resource for cereal genomics: 22K barley GeneChip comes of age. *Plant Physiol.* 134: 960–968.

Demura, T. and Ye, Z.H. (2010) Regulation of plant biomass production. *Curr. Opin. Plant Biol.* 13: 299–304.

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

seed through barley development. *Funct. Integr. Genomics* 6: 202–211.

Fleury, D., Jefferies, S., Kuchel, H. and Langridge, P. (2010) Genetic and genomic tools to improve drought tolerance in wheat. *J. Exp. Bot.* 61: 3211–3222.

Gene Ontology Consortium. (2006) The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.* 34: D322–D326.

Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W. et al. (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J.* 55: 526–542.

Hamada, K., Hongo, K., Suwabe, K., Shimizu, A., Nagayama, T., Abe, R. et al. (2011) OryzaExpress: an integrated database of gene expression networks and omics annotations in rice. *Plant Cell Physiol.* 52: 220–229.

Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37: 1–13.

International Brachypodium Initiative. (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* 463: 763–768.

Iwai, H., Masaoka, N., Ishii, T. and Satoh, S. (2002) A pectin glucuronyltransferase gene is essential for intercellular attachment in the plant meristem. *Proc. Natl Acad. Sci. USA* 99: 16319–16324.

Kota, R., Varshney, R.K., Prasad, M., Zhang, H., Stein, N. and Graner, A. (2008) EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. *Funct. Integr. Genomics* 8: 223–233.

Lee, T.H., Kim, Y.K., Pham, T.T., Song, S.I., Kim, J.K., Kang, K.Y. 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.* 151: 16–33.

Lee, Y., Tsai, J., Sunkara, S., Karamycheva, S., Pertea, G., Sultana, R. et al. (2005) The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res.* 33: D71–D74.

Liepman, A.H., Wightman, R., Geshi, N., Turner, S.R. and Scheller, H.V. (2010) Arabidopsis—a powerful model system for plant cell wall research. *Plant J.* 61: 1107–1121.

McCarthy, F.M., Gresham, C.R., Buza, T.J., Chouvarine, P., Pillai, L.R., Kumar, R. et al. (2011) AgBase: supporting functional modeling in agricultural organisms. *Nucleic Acids Res.* 39: D497–D506.

Mochida, K., Saisho, D., Yoshida, T., Sakurai, T. and Shinozaki, K. (2008) TriMEDB: a database to integrate transcribed markers and facilitate genetic studies of the tribe Triticeae. *BMC Plant Biol.* 8: 72.

Mochida, K. and Shinozaki, K. (2010) Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol.* 51: 497–523.

Mochida, K., Yoshida, T., Sakurai, T., Ogihara, Y. and Shinozaki, K. (2009) TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. *Plant Physiol.* 150: 1135–1146.

Moreno-Risueno, M.A., Busch, W. and Benfey, P.N. (2010) Omics meet networks—using systems approaches to infer regulatory networks in plants. *Curr. Opin. Plant Biol.* 13: 126–131.

Obayashi, T., Hayashi, S., Saeki, M., Ohta, H. and Kinoshita, K. (2009) ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucleic Acids Res.* 37: D987–D991.

Obayashi, T. and Kinoshita, K. (2009) Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. *DNA Res.* 16: 249–260.

Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M. et al. (2007) ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis. *Nucleic Acids Res.* 35: D863–D869.

Ohtsu, K., Takahashi, H., Schnable, P.S. and Nakazono, M. (2007) Cell type-specific gene expression profiling in plants by using a combination of laser microdissection and high-throughput technologies. *Plant Cell Physiol.* 48: 3–7.

Okazaki, Y., Shimojima, M., Sawada, Y., Toyooka, K., Narisawa, T., Mochida, K. et al. (2009) A chloroplastic UDP-glucose pyrophosphorylase from Arabidopsis is the committed enzyme for the first step of sulfolipid biosynthesis. *Plant Cell* 21: 892–909.

Parkinson, H., Sarkans, U., Kolesnikov, N., Abeygunawardena, N., Burdett, T., Dylag, M. et al. (2011) ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res.* 39: D1002–D1004.

Ridley, B.L., O'Neill, M.A. and Mohnen, D. (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* 57: 929–967.

Rivera, C.G., Vakil, R. and Bader, J.S. (2010) NeMo: Network Module identification in Cytoscape. *BMC Bioinformatics* 11(Suppl 1), S61.

Rodriguez, E.M., Svensson, J.T., Malatrasi, M., Choi, D.W. and Close, T.J. (2005) Barley Dhn13 encodes a KS-type dehydrin with constitutive and stress responsive expression. *Theor. Appl. Genet.* 110: 852–858.

Rorat, T. (2006) Plant dehydrins—tissue location, structure and function. *Cell Mol. Biol. Lett.* 11: 536–556.

Rostoks, N., Mudie, S., Cardle, L., Russell, J., Ramsay, L., Booth, A. et al. (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Mol. Genet. Genomics* 274: 515–527.

Rostoks, N., Ramsay, L., MacKenzie, K., Cardle, L., Bhat, P.R., Roose, M.L. et al. (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proc. Natl Acad. Sci. USA* 103: 18656–18661.

Sasaki, E., Takahashi, C., Asami, T. and Shimada, Y. (2010) AtCAST, a tool for exploring gene expression similarities among DNA microarray experiments using networks. *Plant Cell Physiol.* 52: 169–180.

Sato, K., Shin, I.T., Seki, M., Shinozaki, K., Yoshida, H., Takeda, K. et al. (2009) Development of 5006 full-length cDNAs in barley: a tool for accessing cereal genomics resources. *DNA Res.* 16: 81–89.

Sato, K. and Takeda, K. (2009) An application of high-throughput SNP genotyping for barley genome mapping and characterization of recombinant chromosome substitution lines. *Theor. Appl. Genet.* 119: 613–619.

Sato, Y., Antonio, B.A., Namiki, N., Takehisa, H., Minami, H., Kamatsuki, K. et al. (2011) RiceXPro: a platform for monitoring gene expression in japonica rice grown under natural field conditions. *Nucleic Acids Res.* 39: D1141–D1148.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W. et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.

Schulte, D., Close, T.J., Graner, A., Langridge, P., Matsumoto, T., Muehlbauer, G. et al. (2009) The international barley sequencing consortium—at the threshold of efficient access to the barley genome. *Plant Physiol.* 149: 142–147.

Semenov, M.A. and Halford, N.G. (2009) Identifying target traits and molecular mechanisms for wheat breeding under a changing climate. *J. Exp. Bot.* 60: 2791–2804.

Shen, L., Gong, J., Caldo, R.A., Nettleton, D., Cook, D., Wise, R.P. et al. (2005) BarleyBase—an expression profiling database for plant genomics. *Nucleic Acids Res.* 33: D614–618.

Sreenivasulu, N., Graner, A. and Wobus, U. (2008) Barley genomics: an overview. *Int. J. Plant Genomics* 2008: 486258.

Suwabe, K., Suzuki, G., Takahashi, H., Shiono, K., Endo, M., Yano, K. et al. (2008) Separated transcriptomes of male gametophyte and tapetum in rice: validity of a laser microdissection (LM) microarray. *Plant Cell Physiol.* 49: 1407–1416.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H. et al. (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* 36: D1009–D1014.

Talame, V., Ozturk, N.Z., Bohnert, H.J. and Tuberosa, R. (2007) Barley transcript profiles under dehydration shock and drought stress treatments: a comparative analysis. *J. Exp. Bot.* 58: 229–240.

Tommasini, L., Svensson, J.T., Rodriguez, E.M., Wahid, A., Malatrasi, M., Kato, K. et al. (2008) Dehydrin gene expression provides an indicator of low temperature and drought stress: transcriptome-based analysis of barley (Hordeum vulgare L.). *Funct. Integr. Genomics* 8: 387–405.

Umezawa, T., Fujita, M., Fujita, Y., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2006) Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. *Curr. Opin. Biotechnol.* 17: 113–122.

Van Zee, K., Chen, F.Q., Hayes, P.M., Close, T.J. and Chen, T. (1995) Cold-specific induction of a dehydrin gene family member in barley. *Plant Physiol.* 108: 1233–1239.

Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L. and Van de Peer, Y. (2009) Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. *Plant Physiol.* 159: 535–546.

Vogel, J. (2008) Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* 11: 301–307.

Walia, H., Wilson, C., Wahid, A., Condamine, P., Cui, X. and Close, T.J. (2006) Expression analysis of barley (Hordeum vulgare L.) during salinity stress. *Funct. Integr. Genomics* 6: 143–156.

Wise, R.P., Caldo, R.A., Hong, L., Shen, L., Cannon, E. and Dickerson, J.A. (2007) BarleyBase/PLEXdb. *Methods Mol. Biol.* 406: 347–363.

Yamaguchi-Shinozaki, K. and Shinozaki, K. (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.* 57: 781–803.

Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R. et al. (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene–metabolite correlations in Arabidopsis. *Plant Cell* 20: 2160–2176.

Zhang, H., Sreenivasulu, N., Weschke, W., Stein, N., Rudd, S., Radchuk, V. et al. (2004) Large-scale analysis of the barley transcriptome based on expressed sequence tags. *Plant J.* 40: 276–290.

Zhou, J., Lee, C., Zhong, R. and Ye, Z.H. (2009) MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. *Plant Cell* 21: 248–266.