

Keiichi Mochida · Kanako Kawaura · Etsuo Shimosaka
Naoto Kawakami · Tadasu Shin-I · Yuji Kohara
Yukiko Yamazaki · Yasunari Ogihara

Tissue expression map of a large number of expressed sequence tags and its application to in silico screening of stress response genes in common wheat

Received: 15 July 2005 / Accepted: 14 March 2006 / Published online: 11 July 2006
© Springer-Verlag 2006

Abstract In order to assess global changes in gene expression patterns in stress-induced tissues, we conducted large-scale analysis of expressed sequence tags (ESTs) in common wheat. Twenty-one cDNA libraries derived from stress-induced tissues, such as callus, as well as liquid cultures and abiotic stress conditions (temperature treatment, desiccation, photoperiod, moisture and ABA) were constructed. Several thousand colonies were randomly selected from each of these 21 cDNA libraries and sequenced from both the 5' and 3' ends. By computing abundantly expressed ESTs, correlated expression patterns of genes across the tissues were monitored. Furthermore, the relationships between gene expression profiles among the stress-induced tissues were inferred from the gene expression patterns. Multi-dimensional analysis of EST data is analogous to microarray experiments. As

an example, genes specifically induced and/or suppressed by cold acclimation and heat-shock treatments were selected in silico. Four hundred and ninety genes showing fivefold induction or 218 genes for suppression in comparison to the control expression level were selected. These selected genes were annotated with the BLAST search. Furthermore, gene ontology was conducted for these genes with the InterPro search. Because genes regulated in response to temperature treatment were successfully selected, this method can be applied to other stress-treated tissues. Then, the method was applied to screen genes in response to abiotic stresses such as drought and ABA treatments. In silico selection of screened genes from virtual display should provide a powerful tool for functional plant genomics.

Keywords Global gene expression · Stress response gene · Virtual display · Common wheat

Communicated by R. Hagemann

Electronic Supplementary Material Supplementary material is available to authorised users in the online version of this article at <http://dx.doi.org/10.1007/s00438-006-0120-1>.

K. Mochida
Nagahama Institute of Bio-Science and Technology,
Shiga 526-0829, Japan

K. Kawaura · Y. Ogihara (✉)
Laboratory of Genetic Engineering, Kyoto Prefectural University,
Hangi-cho 1-5, Shimogamo, Kyoto 606-8522, Japan
E-mail: yogihara@kpu.ac.jp

E. Shimosaka
Laboratory of Plant Genecology,
Hokkaido National Agricultural Experiment Station,
Sapporo 062-8555, Japan

N. Kawakami
Laboratory of Plant Molecular Physiology,
Meiji University, Kawasaki 214-8571, Japan

T. Shin-I · Y. Kohara · Y. Yamazaki
Center for Genetic Resource Information,
National Institute of Genetics, Mishima 411-8540, Japan

Introduction

Plants regulate gene expression in response to environmental conditions, which allows adaptation to diverse breeding environments. As modular genes are triggered by environmental stresses, a cascade of other genes should be induced (Singh et al. 2002; Glombitza et al. 2004; Narusaka et al. 2004). Accordingly, plants display environmental adaptations against both abiotic and biotic stresses. Systematic monitoring of overall gene expression profiles can be achieved by functional genomic methods, such as cDNA microarrays or synthesized oligo-nucleotide chips (Schaffer et al. 2001; Rabbani et al. 2003; Seki et al. 2004).

In addition to analog methods of analysis, digital analysis of gene expression profiles can also be conducted by analyzing expressed sequence tags (ESTs) in cDNA libraries and gene expression patterns can be compared by transcript abundance in each cDNA

library. A large number of EST data from various cDNA libraries in combination with rigorous statistical analysis is required to assess digital gene expression profiling in each tissue (Ewing et al. 1999). We recently examined a large number of ESTs expressed in ten tissues throughout the wheat life cycle, and produced a body map of ESTs in these tissues by counting the numbers of sequenced clones belonging to each contig. This visualized profiling of gene expression patterns was designated virtual display (VD). These classified contigs represented each gene from the three homoeologous genomes of common wheat (Ogihara et al. 2003). Furthermore, we have developed a new SNP analysis system for hexaploid wheat using the PyrosequencingTM method in combination with the aneuploid series of hexaploid wheat in order to assign every transcript on each chromosome of hexaploid wheat and to clarify the distinct expression patterns of homoeologous genes among the three genomes (Mochida et al. 2003).

As members of Gramineae, wheat and rice show characteristic growth habits. Wheat, which originates in temperate zones, is a long-day plant (Zohary and Hopf 2001). On the other hand, rice, which is cultivated in tropical or subtropical areas, is a short-day plant (Oka 1988). Furthermore, wheat is characteristic of its polyploid nature, although polyploidy is more or less common in plants (Wendel 2000). The gene expression patterns of wheat in response to environmental stresses, particularly temperature stresses, are therefore of substantial interest.

In this context, we herein accumulate a large number of ESTs from cDNA libraries constructed from various stressed wheat tissues in addition to previous wheat ESTs. We have obtained substantial sequence data from 21 cDNA libraries, including those from abiotic-stressed tissues. These sequences were grouped into certain contigs using the phrap method. We show typical gene expression patterns with the "Virtual Display" in response to various stresses. Furthermore, as a model case, we have conducted in silico screening of genes in response to cold acclimation and heat-shock treatment. We are able to screen temperature-responsive genes and annotate these genes. The method applies to screen up- and down-regulated genes in response to abiotic stresses such as cold, drought and ABA treatment. This method offers a novel analysis system for gene expression patterns in response to environmental stresses.

Materials and methods

Plant materials and RNA extraction

Various tissues, such as whole developing seeds, embryos, seedlings, roots, anthers, calluses and liquid cultures from ten cultivars and/or lines of common wheat (*Triticum aestivum* L.) were subjected to stress treatment. After these tissues were subjected to abiotic

stresses, such as cold acclimation, heat shock, dehydration, and abscisic acid (ABA), total RNAs were extracted (Ogihara et al. 2003). In total, 21 cDNA libraries from these RNAs through tissue-treatment combinations were constructed. Detailed descriptions of the stress treatments and strains used are presented in Table 1 and S1.

Construction of cDNA library and sequencing of cDNA clones

Double-stranded cDNAs were synthesized as previously described (Ogihara et al. 2003). cDNAs were ligated with pBluescript SK(+) digested with *EcoRI* and *XhoI*. After transformation by electroporation, transformed bacterial cells were initially cultured in SOC medium for 1 h before culture at 37°C for 2 h in 2× LB medium. Cultured cells were stored in 20% glycerol at -80°C until use.

Transformed bacteria were randomly selected and plasmid DNAs were extracted (Ogihara et al. 2003). Inserted cDNAs were sequenced from both ends using the dye terminator cycle sequencing method (Applied Biosystems, AB, USA).

EST data mining

Grouping of wheat ESTs was performed as previously report (Ogihara et al. 2003). Related cDNA sequences from both ends were grouped as contigs using the phrap method (University of Washington Genome Center; <http://www.genome.washington-edu/UWGC>) under the following conditions: new ace -penalty, -5; -minmatch, 50; -minscore, 100. The resultant contigs were annotated according to DNA sequence homologies against the NCBI database using the BLAST method (Karlin and Altschul 1993). Contigs were then classified into related gene groups using the BLAST method (E value = $1e-30$). In order to carry out statistical analysis of gene expression profiles, contigs harboring 5 or more constituents were selected from among 38,048 contigs (Table 1, S1). Similarities between contigs or libraries were estimated using Pearson's correlation coefficient (Eisen et al. 1998). Hierarchical clustering (<http://www.rana.stanford.edu/clustering>) was applied to compare EST expression profiles among the 21 wheat tissues/treatments and libraries. Expression profiles are displayed based on the number of constituents in a contig (from 0 to 1,047; red intensity), along with increasing number of constituents (VD).

Homology search of wheat contigs

Homology of 38,048 contigs was searched against the Rice GD, and plant EST databases were searched using the BLAST method (E value = $1e-5$). Subsequently, contigs were classified into related gene groups using BLAST (E value = $1e-30$). Wheat gene clusters were examined for homology against rice ORFs in the rice

Table 1 Number of contigs from 21 wheat cDNA libraries constructed from the stress-induced tissues grouped by the phrap method

Tissue and stress treatment	CDNA library code	Wheat strain	No. of forward sequence	No. of reverse sequence	Total no. of sequences	Total no. of contigs after phrap analysis
Callus	whcs	CS ^a	4,960	5,256	10,216	3,861
Anther at booting stage	whgc	CS	4,503	4,702	9,205	3,220
Anther at booting stage	whhg	CSMT4BS ^c	4,411	4,597	9,008	3,745
Anther at booting stage	whpc	CSDT4BS ^d	4,613	4,811	9,424	3,915
Whole grain DPA5 ^b	whok	CS	5,133	5,328	10,461	3,898
Whole grain DPA5	whms	CSDT3DL	5,285	5,444	10,729	4,136
Whole grain DPA20	whdp	CS	5,589	5,719	11,308	3,951
Dormant embryo with water supply	whem	Kitakeil1354	4,912	5,243	10,155	4,375
Dormant embryo with cold treatment after water supply	whec	Kitakeil1354	5,064	5,146	10,210	4,250
Dormant embryo with water supply after wounded	whei	Kitakeil1354	5,061	5,156	10,217	3,884
Seedling root in dehydration	whrd	Kitakeil1354	5,224	5,411	10,635	4,272
Seedling with cold treatment after excision of grain part	whsc	Kitakeil1354	5,191	5,467	10,658	4,297
Upper seedling in dehydration	whsd	Kitakeil1354	4,818	5,069	9,887	3,704
Seedling grown under continuous illumination	whkp	CSKmppd ^e	5,533	5,821	11,354	4,153
Seedling grown under continuous illumination after cold treatment	whkv	CS	5,271	5,471	10,742	4,117
Seedling with cold treatment for 16 days	whv16	Valuevskaya	4,680	5,088	9,768	2,950
Seedling with cold treatment for 3 days	whv3n	Valuevskaya	4,436	4,747	9,183	2,825
Seedling with ABA treatment	whva	Valuevskaya	4,136	4,906	9,042	2,976
Seedling with heat shock	whvh	Valuevskaya	4,497	4,605	9,102	2,318
Seedling in 5 days dehydration	whvd	Valuevskaya	5,055	5,340	10,395	3,239
Cultured cells in B5 medium	whvs	Valuevskaya	5,289	5,516	10,805	4,343
MUGEST2003_all			103,661	108,843	212,504	38,048
MUTEST2003_all + MUGEST2002_all					361,180	53,976

^a *Triticum aestivum* cv. Chinese Spring

^b Days post-anthesis

^c Mono-telosomics 4BS of CS

^d Ditelosomics 4BS of CS

^e Near-isogenic line

pseudomolecules released from TIGR (<http://www.tigr.org/tdb/e2k1/osal>) using the BLAST method (E value = $1e-5$).

Results

Number of sequenced EST clones from various stress-treated tissues

We previously carried out molecular analysis of a large number of wheat ESTs (Ogihara et al. 2003). In total, 116,232 sequences from 10 tissues during the wheat life cycle were obtained. These ESTs were grouped into 25,971 contigs using the phrap method (Ogihara et al. 2003). To enrich the wheat EST database, we constructed an additional 21 cDNA libraries, as shown in Table 1. These cDNA libraries were grouped into three categories: (1) additional cDNA libraries from novel wheat tissues to enrich the body map of wheat gene expression, (2) cDNA libraries to compare differences in gene expression patterns between characteristic genotypes,

such as aneuploids versus euploids and isogenic lines versus pure lines, and (3) cDNA libraries constructed using abiotic-stressed tissues (Table 1, S1). As in previous cases, these libraries were merely cultured for stabilization in order to ensure that expressed gene members in these libraries more or less reflected the original mRNA populations in these tissues/treatments. Several thousand cDNA clones, which were randomly selected from each library, were sequenced from both ends of the inserted DNA fragments (Table 1). In total, 212,504 sequences were obtained. Those sequences were grouped into 38,048 contigs using the phrap method. We then determined the number of novel ESTs obtained in this study. In previous studies, we had four EST data sources: (1) wheat EST2002 (25,971 contigs; Ogihara et al. 2003), (2) MUGDB2000; ESTs derived from young spikelet (666 contigs), (3) Wh_SH; ESTs from seedling (2,378 contigs), and (4) Wh_FL; full-length cDNAs constructed from young spikelets (3,605 contigs; Ogihara et al. 2004). These contigs (designated MUGEST2002) and the contigs established in the present study (designated MUGEST2003) were combined, and grouped into 32,881 gene clusters using the BLAST method ($1e-60$),

as shown in Fig. 1. Of 22,955 gene clusters in MUGEST2003, 8,548 (37.2%) shared homology with those of MUGEST2002, thus demonstrating that the remaining 14,407 gene clusters (62.8%) were novel wheat genes cloned in the present study. This clearly indicates that cDNAs from distinct and stress-treated tissues are quite useful in obtaining new sources of ESTs.

Gene ontology of MUGEST2003

We searched the InterPro database (Apweiler et al. 2001) in order to characterize the 22,955 gene clusters of MUGEST2003. We were able to detect InterPro domains in most wheat gene clusters with molecular functions (21,167 genes, 92.2%; data not shown). Similar detection of protein domains was obtained in other categories, such as biological processes and cellular components. Protein domains found in wheat MUGEST2003 were categorized similarly as those in rice (Kikuchi et al. 2003). Because transcription factors are important and considered to be characteristic in wheat, constituents of the transcription factors were further identified through InterPro search (BLAST X; E value = $1e-30$). Total numbers of genes encoding transcription factors identified through InterPro search were similar between wheat and rice (Table 2). The number of genes for transcription factors specifically expressed in MUGEST2003 (S2) was always higher (about double) than that in MUGEST2002. This suggests that a number of transcription factors classified into each group of Table 2 were induced in response to biotic and abiotic stresses. Although the constituents of transcription factors in the wheat transcriptome were similar to those of rice (Table 2), some discrepancies in the constituents of tran-

scription factors between wheat and rice were identified: higher frequency of Zn-fingers in wheat, and lower frequencies of ERF and NAM. Although this may be partly due to the one-pass sequencing of wheat cDNA clones from both ends, transcription factors in wheat that are differentially expressed in rice should characterize wheat gene expression in response to environmental conditions.

Numerous wheat genes are homologous with genes in the rice genome

In order to analyze the proportion of wheat genes homologous to those of rice, the 32,881 wheat gene clusters were used as the query sequences in a homology search with BLASTX ($E < 1e-5$) against the 32,072 rice peptide sequences pooled in the KOME (<http://www.cdna01.dna.affrc.go.jp/cDNA/>). Only 20,127 of 32,881 wheat gene clusters (61.2 %) showed significant homology with rice full-length cDNAs, indicating that wheat ESTs cover a unique phase of the transcriptome in comparison with those of rice, and are a new data source for studying expression profiles in various tissues. Furthermore, we performed homology search against the rice pseudomolecules which includes ORFs and cDNAs of 61,250 gene models (OSA1 version 3) released from TIGR (<http://www.tigr.org/tdb/e2k1/osa1>) using the BLASTN and the BLASTX with a threshold of e value less than $1e-5$. In total, 22,578 sequences wheat gene clusters (68.7%) showed significant homology with the rice gene models. Among 10,303 sequences, which showed no homology with any rice gene models (OSA1 version 3), 564 sequences (1.7%) showed significant homology with certain genes according to the BLASTX search against the nr peptide database of NCBI ($E < 1e-5$). These wheat genes are possible candidates covering the unique functions disappeared in rice (S3).

Characterization of wheat-specific transcripts

In the 564 genes, some transcription factors, hormone response genes, stress-related genes, storage protein genes and other metabolic genes, as well as many genes with unknown functions and/or no homology with the DNA database, were involved (S3). Transcription factors are important to characterize gene expression specific to plant species. A Zn-finger family and a bZIP family were found among the wheat-specific transcripts. The genes were most homologous with those of *Arabidopsis*. Because 57 GATA transcription factors and 75 distinct members of the bZIP family were identified in the *Arabidopsis* genome, some of which showed no homology with those of rice (Jakovy et al. 2002; Reyes et al. 2004), these genes might play certain roles in determining characters common to both wheat and *Arabidopsis*, such as photo-period sensitivity. Transcription factor IIA (Li et al. 1999) and TATA-box binding protein (TBP)-associated factor might contribute to wheat-specific gene expression. Genes related to abiotic (Myb-related protein) and biotic (IciR-type transcription regulator

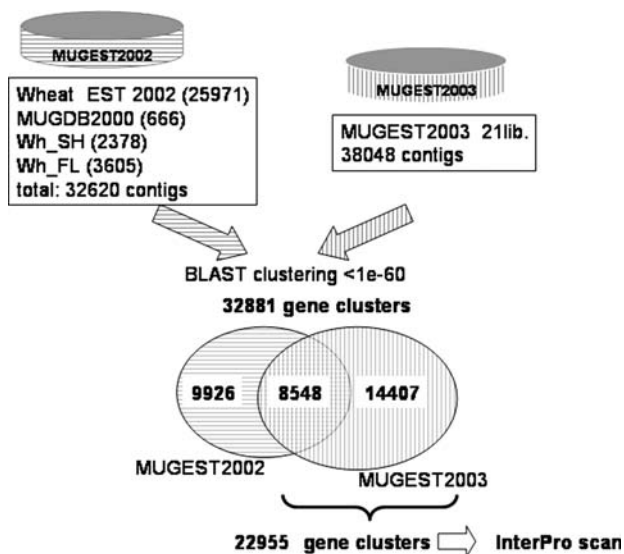


Fig. 1 Clustering of wheat ESTs from various cDNA libraries. Original ESTs from both ends of cDNA clones were grouped using the phrap method. The resultant contigs were then clustered using the BLAST method, as indicated in the figure

Table 2 Transcription factors identified in MUGEST2003 and MUGEST2002 through InterPro search

Category of domain	Total no. of wheat gene clusters	Specific in MUGEST2003 ^a	Specific in MUGEST2002 ^b	No. of rice FL-cDNA clones
Zn-finger	887 (58.7)	338 (44.1)	121 (22.0)	588 (44.6)
Myb	201 (13.3)	73 (44.5)	37 (28.9)	158 (12.0)
ERF	29 (1.9)	14 (56.0)	4 (26.7)	83 (6.3)
NAM	39 (2.6)	14 (41.2)	5 (20.0)	74 (5.6)
Homeobox	73 (4.8)	28 (47.5)	14 (31.1)	73 (5.5)
bZIP	56 (3.7)	17 (34.7)	7 (17.9)	63 (4.8)
AUX/IAA	40 (2.6)	17 (53.1)	8 (34.8)	47 (3.6)
WRKY	29 (1.9)	18 (72.00)	4 (36.4)	46 (3.5)
TFB3	49 (3.2)	18 (40.9)	5 (16.1)	27 (2.1)
GRAS	23 (1.5)	11 (55.0)	3 (25.0)	27 (2.1)
MADS	13 (0.9)	4 (33.3)	1 (11.1)	26 (2.0)
HSF	11 (0.7)	5 (55.5)	2 (33.3)	25 (1.9)
Tubby	12 (0.8)	5 (41.7)	0 (0)	24 (1.8)
BRCT	17 (1.1)	4 (26.7)	2 (15.4)	19 (1.4)
SBP	10 (0.7)	5 (71.4)	3 (60.0)	17 (1.3)
Jumonji	18 (1.2)	7 (46.7)	3 (27.3)	11 (0.8)
TCP	4 (0.3)	1 (100)	3 (100)	10 (0.8)
Total	1,511	579 (44.9)	222 (23.8)	1,318

The transcription domain of wheat gene clusters was searched against the InterPro data base, and the search of transcription factors was also carried out with the BLASTX (E value = $1e-30$) against the rice KOME data base. The candidates through both search systems are listed in the table. Frequencies are given in parentheses

^a These numbers were divided by the gene cluster number in MUGEST2003

^b These numbers were divided by the gene cluster number MUGEST2002

against pathogen attack; Thomson et al. 1999) plant stresses were detected. Furthermore, transcription factors that regulate plant development, such as MADS-box gene (Murai et al. 1997), JUMONJI (Kim et al. 2003) and SWIRM (chromosomal proteins with chromatin-modifying activities; Arabind and Iyer 2002) were also present in the data pool of wheat-specific genes. In addition, 5 hormone-responsive genes and 40 stress-related genes were confirmed in the wheat-specific gene data pool. These genes might control wheat homeostasis in response to environmental conditions. As expected, genes encoding for storage proteins accumulated in wheat seeds (glutenins and gliadins) showed sequence specificity without homology to other plants. However, it should be noted that we could not discern the functions of many genes (465 genes, 57%) based on mere sequence homology.

Virtual display and in silico screening of stress-response genes in wheat

In an effort to assess global expression profiles of ESTs in these 21 different tissues/treatments, contigs showing abundant expression were selected (Ogihara et al. 2003). A total of 9,459 contigs with 5 or more members were selected, and comparative expression profiles of these contigs among the 21 tissues/treatments were displayed. The number of EST constituents assigned to 21 cDNA libraries was scored for each contig, producing a two-way expression profile, i.e., contig versus library. Based on the EST constituent matrix, hierarchical clustering was performed as reported previously (Eisen et al. 1998).

A VD of the expression profile is shown in Fig. 2. In addition to gene clusters expressed in a tissue/treatment-specific manner, genes commonly expressed throughout the 21 tissues/treatments were found (Fig. 2). The 9,459 contigs were classified into 9 major groups. A remarkable feature of the investigation is to identify genes specifically expressed in certain tissues/treatments. VD enables us to extract certain genes in response to abiotic stresses, such as temperature, drought, abscisic acid (ABA) treatment and wounding, as well as biotic stresses like tissue and/or cell cultures (Table 1; Fig. 2). As an example, we extracted contigs expressed in six tissues/treatments related to temperature, drought and ABA; whkp, whv3n, whv16, whvh, whvd, and whva (S1, S4). Comparison of overall expression patterns with the 5,676 contigs revealed that gene expression patterns of seedlings incubated under cold conditions for 3 days and 16 days were similar, while those under heat-shock treatment were grouped into those incubated under cold conditions, constituting a cluster. On the other hand, expression patterns of seedlings under drought and ABA treatment were grouped into another cluster differing from seedlings under temperature treatment (S4). These results support data obtained from cDNA microarray experiments (Rabbani et al. 2003; Seki et al. 2004; Shinozaki et al. 2003). Furthermore, using display data, we are able to extract genes specifically expressed under certain treatments. This method can be designated for in silico screening of responsive genes, as in the case of cDNA microarray. At first, of the six tissues/treatments, we compared expression profiles among three cDNA libraries concerning temperature-responsive genes. The whkp

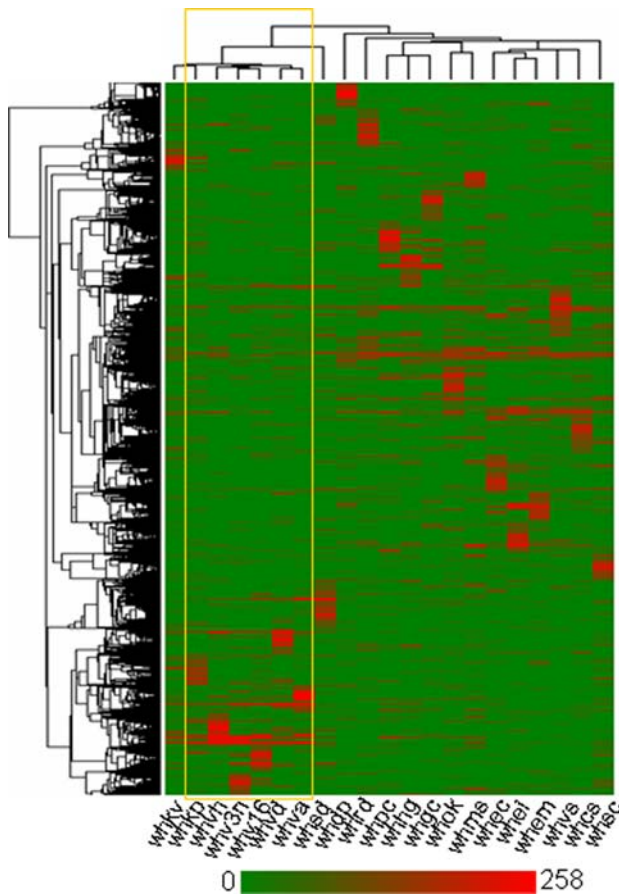


Fig. 2 Clustered correlation display of wheat ESTs in MUGEST2003. A total of 9,459 contigs with 5 or more members were selected to construct the clustered correlation map [virtual display (VD)] of cDNA libraries from the 21 wheat tissues/treatments. At the bottom of the panel, abbreviations for the 21 tissues from which cDNA libraries were constructed are shown (for abbreviations, see Table 1). The expression profile of each gene is represented by a single row of red-colored boxes, and that of each library is represented by a single column

contigs (cDNAs derived from seedlings grown under continuous illumination) was adopted as a control for comparison. The whv16 contigs were used for selection of genes responsive to cold acclimation, while the whvh contigs were applied for selection of heat-shock genes. In comparison to expression levels of control genes, contigs

showing fivefold increases or decreases were selected. The number of selected genes is presented in Fig. 3. These contigs can be categorized into two groups: (1) contigs specifically induced or suppressed in response to temperature, and (2) contigs commonly expressed under temperature-stress conditions, regardless of temperature. In total, 708 genes changed their expression levels. For temperature-induced genes, 269 and 175 genes were specifically up-regulated during cold and heat treatment, respectively. In addition, 46 genes were induced under both temperature conditions. On the other hand, 46 genes were specifically suppressed under cold or heat treatment, while 126 genes were down-regulated under both temperature conditions (Fig. 3). These gene expression patterns should reflect their roles classified into each category (S5). The functions of these 708 genes were inferred with Gene Ontology, as presented in S6. In these categories, genes classified into carbohydrate metabolism, energy metabolism, folding-sorting-degradation and translation were more frequently found than those grouped into other categories. Among the fivefold up-regulated genes after cold acclimation, a number of WCOR genes, such as WCOR14a, WCOR80, WCOR413 and WCOR719, as well as cold-responsive LEA/RAB-related genes (Tsuda et al. 2000), drought-related genes (Ohno et al. 2003) and pathogen-resistance-related genes (Koike et al. 2002; Schultheiss et al. 2003), were found. Chaperones, such as *dnaJ*, *dnaK*, HSP16.9, HSP70, HSP80, which are induced by heat-shock treatment (Wang et al. 2004), were also found among the 708 genes. The list of these genes strongly supports the reliability of in silico monitoring of temperature-responsive genes. Furthermore, a number of transcription factors in response to temperature treatment were detected (S5), among which bHLH, Zn-finger proteins, bZIP, Myb, MADS, and APF1 proteins were included.

In addition to temperature-responsive genes, overlapping of up- and down-regulated genes in response to abiotic stresses, namely cold acclimation (whv16), drought (whvd) and ABA (whva) treatment was examined (Fig. 4). Of up-regulated genes (625 genes in total, as presented in S7), 14 genes were commonly induced by three abiotic stresses. In these abiotic-stress-induced genes, genes related to systemic resistance (lipid transfer protein: Guidendoni et al. 2002; Maldonado et al. 2002;

Fig. 3 Number of contigs selected in silico from the VD in response to temperature treatment. **a** Number of contigs induced by cold acclimation and heat-shock. **b** Number of contigs suppressed by cold acclimation and heat-shock. *Overlapping regions* indicate the number of contigs induced or suppressed in both cold acclimation and heat-shock

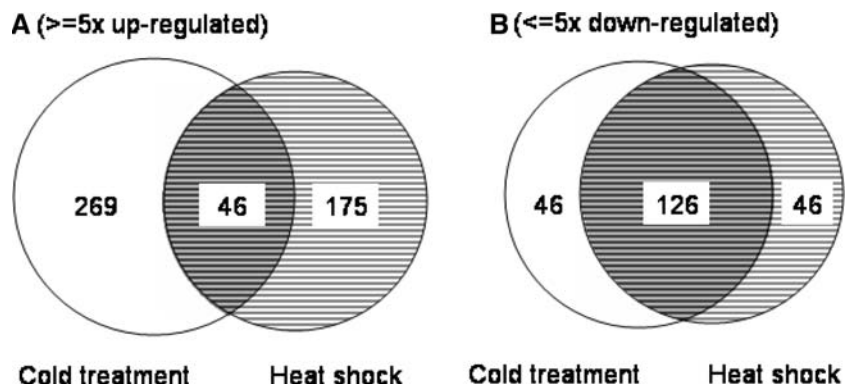
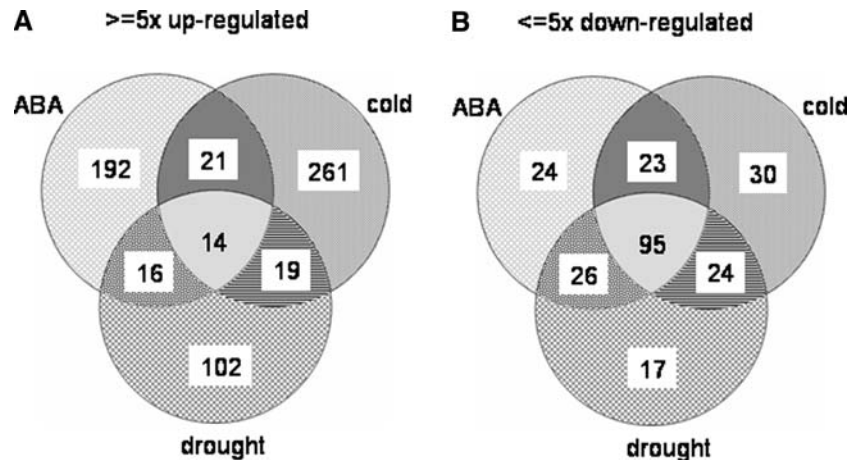


Fig. 4 Number of contigs selected in silico from the VD in response to abiotic/biotic stress. **a** Number of contigs induced by three stress treatments, i.e., ABA, cold acclimation and drought treatment. **b** Number of contigs suppressed by three stresses, namely ABA, cold acclimation and drought treatment. Contigs showing fivefold changes of expression levels were selected



Jung et al. 2003), abiotic stresses (s-adenosyl methionine synthase: Sanchez-Aguay et al. 2004, and phosphoethanolamine methyltransferase: Mou et al. 2002; Charron et al. 2002), reactive oxygen species (ROS; metallothionein, thiol protease and thioredoxin: Wong et al. 2004; Buchanan and Balmer 2005) and brown plant hopper susceptibility protein (Yang et al. 2005) were involved. A number of 21, 19 and 16 genes were overlapped between ABA treatment and cold acclimation, cold acclimation and drought response, and drought response and ABA treatment, respectively. In genes commonly induced by ABA treatment and cold acclimation, fatty acid desaturase (Kodama et al. 1995; Matsuda et al. 2005) and ubiquitin-related genes were found. ROS (ascorbate peroxidase), alanine aminotransferase (Kikuchi et al. 1999), aquaporin PIP1 and glucose 6-phosphate dehydrogenase genes were highly induced in drought and ABA treatments of wheat seedling. In cold acclimation and drought, LEA/RAB-related COR, COR615, ROS-related (thioredoxin/transketolase), dehydrin-5 and amino acid selective channel protein genes were up-regulated. More than 100 genes were specifically induced in response to individual abiotic stresses (Fig. 4a). On the other hand, 95 out of 239 genes were commonly down-regulated in response to 3 abiotic-stresses (Fig. 4b). In these genes (S8), histone genes for H2s and H4, transcription factor containing MADS domain, telomerase reverse transcriptase, RNA-binding protein, translation-related protein, proteinase inhibitors, ROS such as glutathione s-transferase, glutathione peroxidase and superoxide dismutase (SOD), stress-related protein such as cytochrome P450 and HSP70, cytoskeleton-related genes such as tublin β and endo-1,4- β -glucanase, mitochondria-related genes such as mitochondrial carrier protein, cytochrome c oxidase and reductase were detected. A number of 23, 24 and 26 genes were overlapped between ABA treatment and cold acclimation, cold acclimation and drought response, and drought response and ABA treatment, respectively. In genes commonly down-regulated between ABA treatment and cold acclimation, genes for translation initiation factor 5A, stress-related proteins such as nodulin MtN3 family protein, DNA-J protein

and long-chain-fatty acid CoA ligase, tublin α protein, and ABC transporter were involved. Genes for elongation factor, small Ras-related GTP-binding protein, tublin α , and sucrose synthesis-related enzymes were commonly suppressed in drought and ABA treatments. In genes commonly down-regulated between cold acclimation and drought treatment, genes that are stress-related such as benzothiadiazole-induced protein, drought-inducible 22 kD protein, lipid-related genes such as sulfolipid synthase and phospholipase were found. Additionally, 2-oxoglutarate/malate translocator was found in this category. Functional analyses of the genes monitored by in silico profiling might provide a powerful approach for identifying gene networks that respond to abiotic stresses.

Discussion

We demonstrated the in silico screening of stress-response genes in common wheat by analyzing a large number of ESTs from 21 independent cDNA libraries (Table 1). We previously released 148,676 cDNA sequences from 4 EST sources (Ogihara et al. 2003, 2004), collectively designated MUGEST2002, as shown in Fig. 1. In addition to MUGEST2002, we released 212,504 cDNA sequences in the present investigation. As of November 11, 2005, 600,039 wheat ESTs are accumulated in the public database (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). The registered number of ESTs is the second highest among all plant species. Consequently, we have contributed 44.5% of the deposited wheat ESTs to the public database. Our ESTs were collectively grouped into contigs, and these contigs were further grouped into 32,881 gene clusters (unique genes; Fig. 1). The number of genes in the rice genome was recently estimated to be less than 40,000 (Benetzen et al. 2004), and thus we were able to capture more than 80% of the wheat genome. The EST data pool of common wheat is sufficient to construct a cDNA microarray/DNA chip. Furthermore, these substantial genomic

resources must provide the opportunity for wheat improvement through reverse genetics, such as the TILLING method (Slade et al. 2005).

Expressed sequence tag constituents from stress-treated tissues were unique and about 63% of the gene clusters in MUGEST2003 did not overlap those in MUGEST2002 (Fig. 1). This uniqueness is also reflected in the constituents of transcription factors (Table 2). It is expected that a number of transcription factors should be expressed in response to environmental stresses (Madlung and Comai 2004). In fact, transcription factors, such as Zn-fingers, Myb, ERF, NAM, bZIP and WRKY, were specifically expressed in the stress-treated tissues (Table 2). Careful traces of these expression patterns might provide information of novel gene-network systems in response to environmental stresses.

Based on sequence homology, 816 of 32,881 wheat gene clusters (2.5%) were inferred to be unique in comparison with the rice genome. Because wheat and rice show characteristic growth habits, these distinct genes might explain the differences in the growth habits at the molecular level. Furthermore, mapping of these distinct genes on the wheat chromosomes should indicate the hypervariable regions of wheat chromosomes. DNA sequencing of these chromosome regions might then clarify the origins of these novel genes throughout evolution.

Characteristics of gene expression patterns in certain tissues and/or stress treatments can be monitored in silico as a VD (Fig. 2). Genes expressed in response to temperature treatments (Fig. 3) and other abiotic stresses (Fig. 4) were successfully selected in the present study. As in the case of cDNA microarray/DNA chip experiments (Narusaka et al. 2004), expression of genes was distinguishable among the different treatments: (1) genes specifically induced or suppressed by stress treatments, (2) genes commonly induced or suppressed by stress treatments. Reliability of VD has been confirmed by the identification of expressed genes in response to temperature conditions and stress reactions. The advantage of in silico selection through VD is that we can freely add and/or omit the tissues/treatments from the VD panel. We actually constructed a mini-VD constituting six wheat tissues/treatments in order to focus gene expression profiles among these tissues/treatments (S4). Accordingly, VD will become a powerful tool in constructing body maps of expressed genes during the plant life cycle and in monitoring the expression of plant genes in response to stress treatment.

Acknowledgments The authors would like to express their cordial thanks to members of the Wheat Consortium in Japan for supplying of original wheat RNAs; Dr. K. Murai, Fukui Prefectural University, Drs. E. Himi and K. Noda, Okayama University, Dr. S. Nasuda, Kyoto University, Drs. H. Miura and K. Kato, Obihiro University, Dr. T. Koba, Chiba University, and Drs. S. Takumi and C. Nakamura, Kobe University. This work was supported by Grants-in-Aid for Scientific Research on Priority Areas (C) "Genome Science" and basic research (A) (Nos. 13202055 and 13356001) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. The sequence data from this study have been submitted to DDBJ under accession nos. CJ493525–CJ736986 for the EST sequences.

References

- Apweiler R et al (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* 29:37–40
- Arabind L, Iyer LM (2002) The SWIRM domain: a conserved module found in chromosomal proteins to novel chromatin-modifying activities. *Genome Biol* 3: RESEARCH 0039
- Bennetzen JL, Coleman C, Liu R, Ma J, Ramakrishna W (2004) Consistent over-estimation of gene number in complex plant genomes. *Curr Opin Plant Biol* 7:732–736
- Buchanan BB, Balmer Y (2005) REDOX REGULATION: a broadening horizon. *Ann Rev Plant Biol* 56:187–220
- Charron J-BF, Breton G, Danyluk J, Muzac I, Ibrahim RK, Sarhan F (2002) Molecular and biochemical characterization of a cold-regulated phosphoethanolamine *N*-methyltransferase from wheat. *Plant Physiol* 129:363–373
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868
- Ewing R.M, Kahla AB, Poirot O, Lopez F, Audic S, Claverie JM (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res* 9:950–959
- Glombitza S et al (2004) Crosstalk and differential response to abiotic and biotic stresses reflected at the transcriptional level of effector genes from secondary metabolism. *Plant Mol Biol* 54:817–835
- Guideroni E, Cordero MJ, Bignols F, Garcia-Garrido JM, Lescot M, Tharreau D, Meynard D, Ferriere N, Notteghem JL, Delseny M (2002) Inducibility by pathogen attack and developmental regulation of the rice *Ltp1* gene. *Plant Mol Biol* 49:683–699
- Jakovy M, Weissshaar B, Droge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F (2002) bZIP transcription factors in *Arabidopsis*. *Trends Plant Sci* 7:106–111
- Jung HW, Kim W, Hwang BK (2003) Three pathogen-inducible genes encoding lipid transfer protein from pepper are differentially activated by pathogens, abiotic, and environmental stresses. *Plant Cell Environ* 26:915–928
- Karlin S, Altschul SF (1993) Applications and statistics for multiple high scoring segments in molecular sequences. *Proc Natl Acad Sci USA* 90:5873–5877
- Kikuchi H, Hirose S, Toki S, Akama K, Takaiwa F (1999) Molecular characterization of a gene for alanine aminotransferase from rice (*Oryza sativa*). *Plant Mol Biol* 39:149–159
- Kikuchi S et al. (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. *Science* 301:376–379
- Kim TG, Kraus JC, Chen J, Lee Y (2003) JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. *J Biol Chem* 278:42247–42255
- Kodama H, Horiguchi G, Nishiuchi T, Nishimura M, Iba K (1995) Fatty acid desaturation during chilling acclimation is one of the factors involved in conferring low-temperature tolerance to young tobacco leaves. *Plant Physiol* 107:1177–1185
- Koike M, Okamoto T, Tsuda S, Imai R (2002) A novel plant defensin-like gene of winter wheat is specifically induced during cold acclimation. *Biochem Biophys Res Commun* 298:46–53
- Li YF, LeGourierrec J, Torki M, Kim YJ, Guerinneau F, Zhou DX (1999) Characterization and functional analysis of *Arabidopsis* TFIIA reveal that the evolutionally conserved region of the large subunit has a transcription factor domain. *Plant Mol Biol* 39:515–525
- Madlung A, Comai L (2004) The effect of stress on genome regulation and structure. *Ann Bot* 94:481–495
- Maldonado AM, Doerner P, Dixan RA, Lamb CJ, Cameron RK (2002) A putative lipid transfer protein involved in systemic resistance in *Arabidopsis*. *Nature* 419:399–403
- Matsuda O, Sakamoto H, Hashimoto T, Iba K (2005) A temperature-sensitive mechanism that regulates post-translational stability of a plastidial omega-3 fatty acid desaturase (FAD8) in *Arabidopsis* leaf tissues. *J Biol Chem* 280:3597–3604

- Mochida K, Yamazaki Y, Ogihara Y (2003) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol Gen Genomics* 270:371–377
- Mou Z, Wang X, Fu Z, Dai Y, Han C, Ouyang J, Bao F, Hu Y, Li J (2002) Silencing of phosphoethanolamine *N*-methyl transferase results in temperature-sensitive male sterility and salt hypersensitivity in *Arabidopsis*. *Plant Cell* 14:2031–2043
- Murai K, Murai R, Ogihara Y (1997) Wheat MADS box genes, a multigene family dispersed throughout the genome. *Genes Genet Syst* 72:317–321
- Narusaka Y, Narusaka M, Seki M, Umezawa T, Ishida J, Nakajima M, Enju A, Shinozaki K (2004) Crosstalk in the responses to abiotic and biotic stresses in *Arabidopsis*: analysis of gene expression in cytochrome P450 gene superfamily by cDNA microarray. *Plant Mol Biol* 55:327–342
- Ogihara Y, Mochida K, Nemoto Y, Murai K, Yamazaki Y, Shin-I T, Kohara Y (2003) Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. *Plant J* 33:1001–1011
- Ogihara Y, Mochida K, Kawaura K, Murai K, Seki M, Kamiya A, Shinozaki K, Carninci P, Hayashizaki Y, Shin-I T, Kohara Y, Yamazaki Y (2004) Construction of a full-length cDNA library from young spikelets of hexaploid wheat and its characterization by large-scale sequencing of expressed sequence tags. *Genes Genet Syst* 79:227–232
- Ohno R, Takumi S, Nakamura C (2003) Kinetics of transcript and protein accumulation of a low-molecular-weight wheat LEAD-11 dehydrin in response to low temperature. *J Plant Physiol* 160:193–200
- Oka HI (1988) *Origin of cultivated rice*. Elsevier, Amsterdam
- Rabbani MA, Maruyama L, Abe H, Khan MA, Katsura K, Ito Y, Yashiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiol* 133:1755–1767
- Reyes JC, Muro-Pastor MI, Florencio FJ (2004) The GATA family of transcription factors in *Arabidopsis* and rice. *Plant Physiol* 134:1718–1732
- Sanchez-Aguay I, Rodriguez-Galan JM, Garcia R, Torreblanca J, Pardo JM (2004) Salt-stress enhances xylem development and expression of s-adenosyl-L-methionine synthase identifying tissues of tomato plants. *Planta* 220:278–285
- Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E (2001) Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *Plant Cell* 13:113–123
- Schultheiss H, Dechert C, Kogel KH, Huelclhoven R (2003) Functional analysis of barley RAC/ROP G-protein family members in susceptibility to the powdery mildew fungus. *Plant J* 36:589–601
- Seki M et al (2004) RIKEN *Arabidopsis* full-length (RAFL) cDNA and its application for expression profiling under abiotic stress conditions. *J Exp Bot* 55:213–223
- Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulation network of gene expression in the drought and cold stress responses. *Curr Opin Plant Biol* 6:410–417
- Singh K, Foley RC, Onate-Sanchez L (2002) Transcription factors in plant defense and stress responses. *Curr Opin Plant Biol* 5:430–436
- Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat Biotech* 23:75–81
- Thomson NR, Nasser W, McGowan S, Sebahia M, Salmond G.P (1999) Erwinia carotovora has two KdGR-like proteins belonging to the IciR family of transcriptional regulators: identification and characterization of the RexZ activator and the KdGR repressor of pathogenesis. *Microbiology* 145:1531–1545
- Tsuda K, Tsvetanov S, Takumi S, Mori N, Atanassov A, Nakamura C (2000) New members of a cold-responsive group-3 *Lea*/*Rab*-related *Cor* gene family from common wheat. *Genes Genet Syst* 75:179–188
- Wang W, Vinocur B, Shoseyov O, Altman A (2004) Role of plant heat-shock proteins and molecular chaperons in the abiotic stress response. *Trends Plant Sci* 9:244–252
- Wendel JF (2000) Genome evolution in polyploids. *Plant Mol Biol* 42:225–249
- Wong HL, Sakamoto T, Kawasaki T, Umemura K, Shimamoto K (2004) Down-regulation of metallothionein, a reactive oxygen scavenger, by the small GTPase OsRac1 in rice
- Yang Z, Zhang F, He Q, He G (2005) Molecular dynamics of detoxification and toxin-tolerance genes in brown plant hopper (*Nilaparvata lugens* Stal., Homoptera: Delphacidae) feeding on resistant rice plants. *Arch Insect Biochem Physiol* 59:59–66
- Zohary D, Hopf M (2001) *Domestication of plants in the old world*, 3rd edn. Oxford University Press, New York