Use of a large-scale Triticeae expressed sequence tag resource to reveal gene expression profiles in hexaploid wheat (*Triticum aestivum* L.)

S. Chao, G.R. Lazo, F. You, C.C. Crossman, D.D. Hummel, N. Lui, D. Laudencia-Chingcuanco, J.A. Anderson, T.J. Close, J. Dubcovsky, B.S. Gill, K.S. Gill, J.P. Gustafson, S.F. Kianian, N.L.V. Lapitan, H.T. Nguyen, M.E. Sorrells, P.E. McGuire, C.O. Qualset, and O.D. Anderson

Abstract: The US Wheat Genome Project, funded by the National Science Foundation, developed the first large public Triticeae expressed sequence tag (EST) resource. Altogether, 116 272 ESTs were produced, comprising 100 674 5′ ESTs and 15 598 3′ ESTs. These ESTs were derived from 42 cDNA libraries, which were created from hexaploid bread wheat (*Triticum aestivum* L.) and its close relatives, including diploid wheat (*T. monococcum* L. and *Aegilops speltoides* L.), tetraploid wheat (*T. turgidum* L.), and rye (*Secale cereale* L.), using tissues collected from various stages of plant growth and development and under diverse regimes of abiotic and biotic stress treatments. ESTs were assembled into 18 876 contigs and 23 034 singletons, or 41 910 wheat unigenes. Over 90% of the contigs contained fewer than 10 EST members, implying that the ESTs represented a diverse selection of genes and that genes expressed at low and moderate to high levels were well sampled. Statistical methods were used to study the correlation of gene expression patterns, based on the ESTs clustered in the1536 contigs that contained at least 10 5′ EST members and thus representing the most abundant genes expressed in wheat. Analysis further identified genes in wheat that were significantly upregulated ($p < 0.05$) in tissues under various abiotic stresses when compared with control tissues. Though the function annotation cannot be assigned for many of these genes, it is likely that they play a role associated with the stress response. This study predicted the possible functionality for 4% of total wheat unigenes, which leaves the remaining 96% with their functional roles and expression patterns largely unknown. Nonetheless, the EST data generated in this project provide a diverse and rich source for gene discovery in wheat.

Key words: Expressed sequence tags, ESTs, gene expression profiles, wheat, Triticeae.

Résumé : Le projet américain sur le génome du blé, financé par la « National Science Foundation », a développé la première grande collection publique d'EST chez les hordées. Au total, 116 272 EST ont été produits à partir de 42

Received 30 June 2005. Accepted 28 December 2005. Published on the NRC Research Press Web site at http://genome.nrc.ca on 8 June 2006.

Corresponding Editor: G.J. Scoles.

S. Chao.[1,2] US Department of Agriculture - Agricultural Research Service (USAD-ARS), Western Regional Research Center, 800 buchanan St., Albany, CA 94170, USA; and Genetic Resources Conservation Program, University of California, Davis, CA 95616, USA.
G.R. Lazo, F. You,[3] **C.C. Crossman, D.D. Hummel, N. Lui, D. Laudencia-Chingcuanco, and O.D. Anderson.** US Department of Agriculture-Agricultural Research Service (USDA-ARS), Western Regional Research Center, 800 Buchanan Street, Albany, CA 94710, USA.
P.E. McGuire and C.O. Qualset. Genetic Resources Conservation Program, University of California, Davis, CA 95616, USA.
J. Dubcovsky. Department of Plant Sciences, University of California, Davis, CA 95616, USA.
J.A. Anderson. Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA.
T.J. Close. Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA.
B.S. Gill. Department of Plant Pathology, Wheat Genetics Resource Center, Kansas State University, Manhattan, KS 66506, USA.
K.S. Gill. Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA.
J.P. Gustafson. USDA-ARS Plant Genetics Research Unit, Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA.
S.F. Kianian. Department of Plant Sciences, North Dakota State University, Fargo ND 58105, USA.
N.L.V. Lapitan. Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523, USA.
H.T. Nguyen. Division of Plant Sciences, University of Missouri, Columbia MO 65211, USA.
M.E. Sorrells. Department of Plant Breeding, Cornell University, Ithaca, NY 14853, USA.

[1]Corresponding author (e-mail: chaos@fargo.ars.usda.gov).
[2]Present address: USDA-ARS Biosciences Research Lab, 1605 Albrecht Blvd, Fargo, ND 58105-5674, USA.
[3]Present address: Department of Plant Sciences, University of California, Davis, CA 95616, USA.

 doi:10.1139/G06-003

banques d'ADNc du blé hexaploïde (*Triticum aestivum* L.) ou de ses proches parents dont des blés diploïdes (*T. monococcum* L. et *Aegilops speltoides* L.), tétraploïdes (*T. turgidum* L.) et le seigle (*Secale cereale* L.). Les ADNc étaient issus de tissus prélevés à divers stades de croissance ou de développement et soumis à différents stress biotiques ou abiotiques. Les EST ont été assemblés pour former 18 876 contigs et 23 034 singulons pour un total de 41 910 unigènes. Plus de 90 % des contigs comprenaient moins de 10 EST, ce qui implique que les EST représentaient une vaste gamme de gènes et que les gènes modérément ou faiblement exprimés étaient bien échantillonnés. Des analyses statistiques ont été employées pour étudier la corrélation entre les motifs d'expression génique en s'appuyant sur les EST compris au sein des 1 536 contigs qui comptaient au moins 10 EST en 5′ (ceux qui représentent vraisemblablement les gènes le plus fortement exprimés chez le blé). Des analyses ont également permis d'identifier des gènes dont l'expression était significativement plus élevée ($p < 0,05$) au sein de tissus ayant subi divers stress par rapport aux tissus témoins. Bien qu'on ne puisse assigner aucune fonction à plusieurs de ces gènes, il est probable qu'ils jouent un rôle dans la réponse aux stress. Cette étude a permis de prédire une fonction pour 4 % des unigènes chez le blé, ce qui laisse 96 % de gènes pour lesquels la fonction demeure largement inconnue. Néanmoins, les données produites sur les EST au cours de ce travail constituent une vaste et riche ressource pour la découverte de gènes chez le blé.

Mots clés : étiquettes de séquences exprimées, EST, profils d'expression génique, blé, hordées.

[Traduit par la Rédaction]

Introduction

The first partial cDNA sequencing project was started in 1991 to generate a large number of expressed sequence tags (ESTs) as a comprehensive resource to survey expressed genes in the human genome (Adams et al. 1991). Since then, over 30 million ESTs have been generated and made publicly available (National Center for Biotechnology Information's dbEST, October 2005), representing a wide variety of fungi, plants, and animals. The surge in EST databases is largely due to the advancement of high-throughput, automated sequencing technology, making large-scale EST sequencing cost-effective for producing large amounts of sequence information from the transcribed portion of genomes. Large-scale EST data have proven critically important for plants whose whole-genome sequences are known, e.g., *Arabidopsis thaliana* and rice (*Oryza sativa* L.), to assist in locating gene regions and splice junctions.

Among plants, hexaploid bread wheat (*Triticum aestivum* L.) currently has the second highest number of ESTs deposited in dbEST, at >600 000. However, this is a relatively recent situation, since as late as May 2000 there were only 9 ESTs recorded. Hexaploid wheat, even though it is one of the world's foundation food plants, will not be completely sequenced in the near future because of its very large genome size (16 000 Mb). Thus, this huge amount of EST-derived sequence information is important for wheat, particularly as markers, as sources of candidate genes, and as a valuable resource for systematic study of the functional portion of the wheat genome.

This report provides details of the US Wheat Genome Project, which developed the first large public wheat EST resource and the first use of these ESTs in large-scale genome mapping and organization studies (Gustafson et al. 2004). The project has contributed over 116 000 *Triticeae* ESTs to the public databases (Lazo et al. 2004), of which over 7800 were mapped physically to deletion bins to all 21 chromosomes. This mapping revealed and confirmed the presence of gene-rich regions of the wheat genomes (8 papers in the October 2004 issue of *Genetics*). Through sequence comparisons, chromosome synteny was confirmed between wheat and rice chromosomes (Sorrells et al. 2003). On the basis of the mapped EST information, wheat chromosome organization and evolution in terms of gene duplications and deletions were characterized and evaluated with respect to recombination rates along chromosome arms (Akhunov et al. 2003*a*, 2003*b*; Dvorak and Akhunov 2005).

With the dramatic increase of EST data released to the public in recent years, electronic analyses based on statistical methods have been exploited for gene expression profiling studies. Quantitative analysis of gene expression has been achieved by calculating the frequency of EST appearances from random ESTs generated by nonbiased mRNA representations (Okubo et al. 1992). The results indicated that transcript abundance can be inferred directly from EST counts. Normalization and subtraction can help improve gene discovery (Zhang et al. 2004), but relative transcript abundance has been altered in the process, thus the resulting populations of ESTs will not be suitable for our purposes. To interpret, reliably, differences of EST counts as an indication of differential gene expression, Audic and Claverie (1997) developed a statistical method taking into account the sample size. This test has been used as a tool to discover genes associated with metabolic pathways in *Arabidopsis* (Ohlrogge and Benning 2000), cell wall biosynthesis in loblolly pine (*Pinus taeda* L.) (Whetten et al. 2001), and tuber development in potato (*Solanum tuberosum* L.) (Ronning et al. 2003). To assess global gene expression profiles, Ewing et al. (1999) used statistical clustering to group genes (or contigs) with similar functions or cDNA libraries (or tissues) with similar expression patterns based on EST contig assembly results. This method should be widely applicable for analyzing large-scale EST data derived from cDNA libraries constructed using distinct and diverse tissue types such as ours.

Previously, gene expression profiling using ESTs from 10 different wheat tissues revealed genes expressed preferentially in specific tissues and those expressed commonly within all tissues (Ogihara et al. 2003). In that analysis, over 116 000 ESTs containing similar amounts of both 5′ and 3′ ESTs derived from normal tissues at different growth and developmental stages were used. The present project focused on generating mainly 5′ ESTs (>100 000) derived from 42 cDNA libraries constructed with a wide range of tissues at

different growth stages and under various abiotic and biotic stresses. In this report we present the aspects of gene expression profiling of wheat by EST analysis.

The results reported here provide an overview of possible changes of wheat gene expression patterns among different tissue types at different stages of tissue development and under various abiotic stress growth conditions in wheat. Also identified were genes that showed significant upregulation under different stress treatments and provided insight into the genes likely responsible for stress tolerance in wheat.

Materials and methods

cDNA libraries

Forty-two cDNA libraries were constructed, 33 of them from tissues of hexaploid wheat (*T. aestivum*, AABBDD, $2n = 6x = 42$). To improve EST diversity and to sample conditions well characterized in specific *Triticeae* species, libraries were also included from close relatives of wheat. These included 3 libraries from *T. monoccocum* (A^mA^m, related to the A genome donor of hexaploid wheat, $2n = 14$), 2 libraries from *Aegilops speltoides* (SS, a likely B genome donor, $2n = 14$), 1 library from tetraploid durum wheat (*T. turgidum*, AABB, $2n = 4x = 28$), and 3 libraries from rye (*Secale cereale*, RR, $2n = 14$). To assess the effectiveness in reducing sequence redundancy, 5 libraries were modified by normalization and subtraction. Detailed methods on library construction and protocols used for normalization and subtraction have been previously described (Zhang et al. 2004). These authors' evaluations of the quality of the normalized and subtracted libraries have indicated that the ESTs derived from the normalized library TA006E3N and the subtracted library TA007E3S had nonwheat cDNA contamination (Zhang et al. 2004). Those ESTs were subsequently retracted from the NCBI's dbEST. Although ESTs from the 2 libraries were included in the contig assembly, they were excluded in the gene expression profiling analysis described in this report. Table 1 lists the libraries, the laboratories of origin, and information on tissue and stress or control condition.

Large-scale EST generation, data processing, annotation, and assembly

The procedures used to generate EST data and the bioinformatics pipeline developed for large-scale EST data processing were reported in Lazo et al. (2004). EST assembly was carried out using the program Phrap (University of Washington, Seattle, http://www.genome.washington.edu/UWGC/analysistools/Phrap.cfm). The parameters used to run Phrap were "-penalty -5 -minmatch 50 -minscore 100 -phrapview", which will allow any 2 sequences with at least 90% sequence similarity over 100 bases to form a contig. The stringency of the parameters used for contig assembly was tested using 1500 ESTs derived from a wheat endosperm library (TA001E1X) mixed with 6 known and well-characterized genome specific glutenin sequences encoding ax, ay, bx, by, dx, and dy subunits (Anderson and Greene 1989). By varying the "penalty", "minmatch", and "minscore", an optimal set of parameters was found, stringent enough to preclude sequences from different homoeologous chromosomes from forming a contig. The decision to adopt stringent parameters for this study was mostly due to the polyploid nature of wheat and the lack of empirical data allowing an adequate threshold for the sequence similarity level to be determined that would permit homoeologous wheat gene sequences to be clustered together. The contigs found to be misassembled because of the presence of chimeric ESTs were manually checked. Chimeric ESTs were removed from the processed EST data set used for final assembly.

The final assembly results were written into a phrapview file, which was then parsed by Perl scripts written in-house. The resulting file included the information of contig number, EST names, their library origin, and the number of ESTs from each library assembled in each contig. This file was the source used for the gene expression profile analysis described next. The complete EST and contig annotation results can be queried at http://wheat.pw.usda.gov/westsql/.

Statistical analysis of EST data

Cluster analysis used to analyze gene expression profiling was based on the method described by Ewing et al. (1999). Cluster analysis was performed using the PROC CLUSTER function in SAS (SAS Institute Inc., Cary, N.C.). Hierarchical clustering was done using the unweighted pair-group method with arithmetic mean (UPGMA) algorithm as part of the PROC CLUSTER function in SAS. Two dendrograms were generated to display the mean distance among cDNA libraries or among contigs analyzed. The contigs were then reorganized in the same order as shown in the dendrogram, the resulting reorganized contig list being the basis for the gene expression profile analysis. A 2-dimensional graphic display of gene expression profile was generated by importing reorganized data into TreeView software, as described by Eisen et al. (1998). For differential gene expression analysis, pair-wise comparison was performed on the basis of EST counts, following the method of Audic and Claverie (1997). The program for computing the probability of differential expression was downloaded from http://igs-server.cnrs-mrs.fr/~audic/significance.html and compiled in the Unix environment.

Results and discussion

Large-scale EST data development

During the course of this project 100 674 5′ ESTs and 15 598 3′ ESTs derived from 42 cDNA libraries were generated. Table 1 lists the numbers of 5′ and 3′ processed ESTs produced from each of the libraries after chimeric ESTs and ESTs of rRNA origin had been manually removed. The total number of 5′ ESTs generated varied among libraries, ranging from 268 to more than 11 000. Sequencing depth for each library was monitored according to the quality of the library and level of redundancy of the library. The mean high-quality sequence read length ranged from 396 to 676 bases with an overall mean of 497 bases. The 3′ ESTs were generated to facilitate selection and identification of unique or nonredundant ESTs for use as probes for wheat physical mapping (Qi et al. 2004). Details on EST mapping probe selection and validation, and physical mapping results from 7800 unique ESTs from this EST set were described in a series of papers in the October 2004 issue of *Genetics*.

Table 1. cDNA libraries and their tissue source used for EST production, number of 5′ and 3′ ESTs produced in each library, and mean quality sequence read length for 5′ ESTs.

Library name, source*	Tissue and treatment†	No. of 5′ ESTs	Mean quality read length (bases)	No. of 3′ ESTs
TA001E1X[1]	Endosperm ('Cheyenne')	2 663	458	1 216
TA001E1S[1]	Endosperm[b] ('Cheyenne')	268	406	—
TA005E1X[2]	Dehydrated seedling	788	491	92
TA006E1X[2]	Etiolated shoot	2 244	498	264
TA006E2N[3]	Etiolated shoot[a]	1 686	447	174
TA006E3N[3]	Etiolated shoot[a]	1 670	556	332
TA007E1X[2]	Cold-stressed seedling	935	477	301
TA007E3S[3]	Cold-stressed seedling[b]	1 555	422	687
TA008E1X[2]	Etiolated seedling root	4 099	457	658
TA008E3N[3]	Etiolated seedling root[a]	4 294	498	979
TA009XXX[4]	*Fusarium*-infected spike (Sumai 3)	10 253	511	1 077
TA012XXX[5]	ABA-treated embryo (Brevor)	1 916	469	—
TA015E1X[2]	Heat-stressed seedling	818	408	291
TA016E1X[2]	Vernalized crown	2 278	424	703
TA017E1X[2]	20–45 DPA spike	1 073	476	177
TA018E1X[2]	5–15 DPA spike	2 857	450	709
TA019E1X[2]	Preanthesis spike	11 175	442	3 071
TA027E1X[2]	Drought-stressed leaf (TAMW101)	904	396	136
TA031E1X[2]	Heat-stressed flagleaf	971	427	341
TA032E1X[2]	Heat-stressed spike	1 009	406	309
TA036E1X[2]	Drought-stressed leaf	639	443	180
TA037E1X[2]	Salt-stressed sheath	897	431	248
TA038E1X[2]	Salt-stressed crown	940	459	286
TA047E1X[2]	Root tip	953	598	75
TA048E1X[2]	Aluminum-stressed root tip (BH1146)	991	555	55
TA049E1X[2]	ABA-treated dormant embryo (Brevor)	2 904	562	148
TA054XXX[6]	Early meiotic stage anther	9 132	525	—
TA055E1X[2]	Drought-stressed root	1 309	645	—
TA056E1X[2]	Aluminum-stressed root tip	1 032	618	—
TA058E1X[2]	Unstressed root at tiller stage	1 021	640	—
TA059E1X[2]	Developing whole grain (Butte86)	3 639	649	—
TA065E1X[7]	Salt-stressed root	2 051	667	—
TA066E1X[7]	Mixed tissue	1 404	640	—
TM011XXX[8]	Vegetative apex (Dv92)	3 020	428	—
TM043E1X[8]	Early reproductive apex (Dv92)	2 639	448	930
TM046E1X[8]	Vernalized shoot apex (G3116)	3 361	586	—
TT039E1X[2]	Whole plant (Langdon16)	1 191	451	284
SC010XXX[9]	Aluminum-stressed root tip (Blanco)	1 198	437	—
SC013XXX[9]	Root tip (Blanco)	778	423	—
SC024E1X[2]	Anther (Blanco)	4 614	421	1 071
AS040E1X[2]	Anther	2 462	464	804
AS067E1X[7]	Anther	1 043	676	—
Total		100 674		15 598
Mean			497	

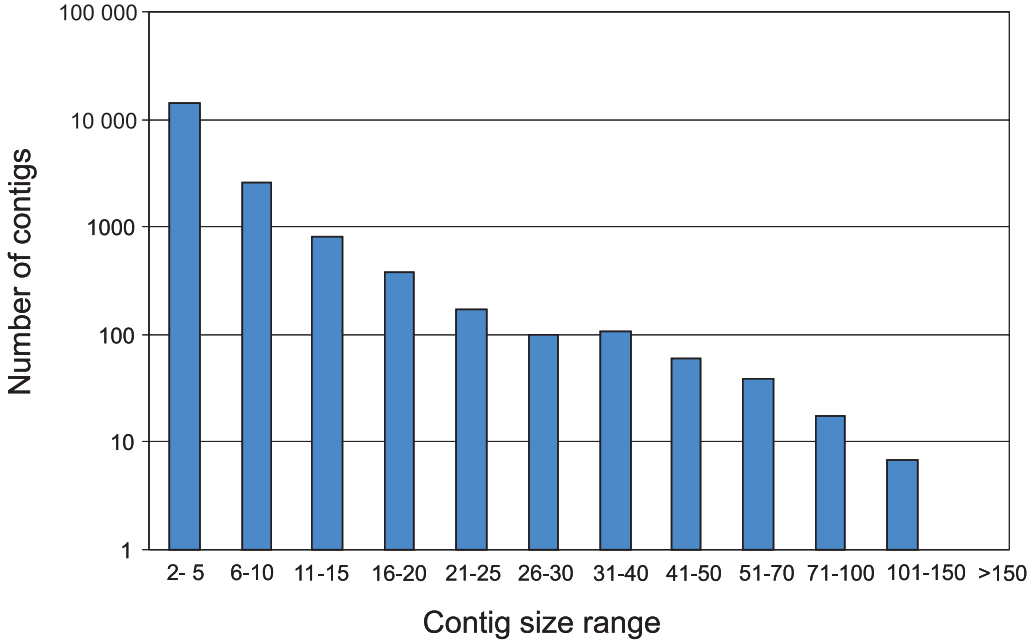
*Library source: 1, O. Anderson, USDA–ARS, Albany, Calif.; 2, T. Close, U.C. Riverside; 3, H. Nguyen, University of Missouri; 4, G. Muehlbauer, University of Minnesota; 5, K. Simmons, USDA–ARS; 6, P. Langridge, Waite Institute, Adelaide, Australia; 7, J. Dvořák, U.C. Davis; 8, J. Dubcovsky, U.C. Davis; 9, P. Gustafson, USDA–ARS, Columbia, Mo. The species source of the library is indicated by the first 2 letters of the name; TA indicates *Triticum aestivum*, TM is *T. monococcum*, TT is *T. turgidum*, SC is *Secale cereale*, and AS is *Aegilops speltoides*.

†Library treatment: [a]normalized library, [b]subtracted library. All of the TA libraries are from the hexaploid wheat 'Chinese Spring' genotype except where indicated otherwise in parentheses; RNA for both *Ae. speltoides* libraries came from F_2 plants from the cross 2-12-4-8-1-1-1-(1) by PI36909-12-811-(1); DPA, days postanthesis.

EST assembly

In addition to the project ESTs already described, 728 5′ ESTs generated from a cDNA library of Sumai 3 wheat spikes infected with the head blight fungal pathogen *Fusarium graminearum* (library TA006G1X from John Fellers, US Department of Agriculture, Agricultural Research Service (USDA–ARS), Manhattan, Kans.) and 510 5′ ESTs derived from a 'Cheyenne' wheat endosperm library (O.D. Anderson, USDA-ARS, Albany, Calif.) previously submitted to the International Triticeae EST Cooperative (ITEC

Fig. 1. Contig size distribution. *x* axis is the range of number of ESTs in each contig. *y* axis is number of contigs shown in a logarithmic scale.



at http://wheat.pw.usda.gov/genome, Gustafson et al. 2004) were included for EST assembly. Altogether, a total of 117 510 ESTs were used.

The results showed that 94 476 ESTs collapsed into 18 876 contigs, and the remaining 23 034 ESTs were singletons. The total number of wheat unigenes was 41 910 with an overall EST redundancy of 65%. The contig consensus sequence lengths ranged from 120 to 3452 bases with a mean length of 809 bases. The number of contigs having fewer than 10 EST members was 17 176, which accounted for 91% of the total contigs (Fig. 1). This clearly reflected that the generated EST data represented a diverse set of genes and that genes expressed at low/moderate to high levels were probably well sampled at the sequencing depth adopted in this project. Forty-one contigs were found with 60 or more EST members, representing the most abundant genes expressed in wheat and rye. Gene annotation based on BLASTX best alignments and numbers of EST members for these contigs are listed in Table 2. No contig was found containing ESTs originating from all 42 cDNA libraries. Possible causes are the diverse range of gene expression patterns in some libraries or among all libraries, shallow EST sampling, and (or) the stringent assembly parameters used.

For the original EST assembly and EST mapping (Lazo et al. 2004), 3′ EST data were included. While 5′ EST sequencing can facilitate gene identification during the annotation process, 3′ EST information can extend a contig in its 3′ direction and assist in eliminating redundant ESTs. Of the 15 598 3′ ESTs used for assembly, 4509 (29%) were singletons and failed to assemble into any contigs containing their corresponding 5′ sequences. The remainder of the 11 089 3′ ESTs were assembled into 9049 contigs, from which 3321 contigs, or 4071 3′ ESTs, did not have their 5′ ESTs assembled in the same contig. Taken together, the assembly results indicated that about 55% of all the 3′ ESTs (8580/15 598) used for assembly were not able to assemble with their corresponding 5′ ESTs because of insufficient sequence overlap. This may have been the result of the strategy used in selecting clones from each contig for 3′ sequencing, which strongly favored 5′ ESTs that could extend each contig farthest toward the 5′ end. This strategy was used previously in *Drosophila* to successfully enrich clones containing full-length open reading frames (Rubin et al. 2000). Although this strategy ensures that the longest clones are selected for each gene candidate, in our case the additional annotation and inclusion of full-length ESTs for assembly were necessary to help in linking all the 5′ ESTs and their corresponding 3′ ESTs together.

Gene expression profile in wheat

Cluster analysis was used to examine global gene expression profiles in wheat (Ewing et al. 1999). To ensure that the ESTs used for this analysis were from libraries produced from unbiased transcripts, the ESTs generated from the 5 normalized and subtracted libraries were excluded from the analysis. Because a 5′ EST and its corresponding 3′ EST were derived from the same cDNA clone, all 3′ ESTs were removed from the analysis to avoid having the same EST (clone) counted more than once in the same contig. In this analysis, only contigs with at least 10 EST members were considered. It was found that the majority of the contigs with fewer than 9 EST members often had only 1 or 2 ESTs contributed from different libraries. The gene expression profiling analysis, based on statistical methods, would not yield statistically meaningful results if ESTs with lower levels of expressions were included.

After removing contigs with fewer than 10 members, 29 525 ESTs clustered in 1536 contigs derived from 37 libraries representing the most abundant genes expressed in wheat were compiled for the analysis. The gene expression data were generated by first tabulating the number of ESTs derived from each library for each of the 1536 contigs. Gene

Table 2. The most abundant contigs (60 or more ESTs in each contig) and their gene annotations based on BLASTX best hit (cutoff 10^{-5}).

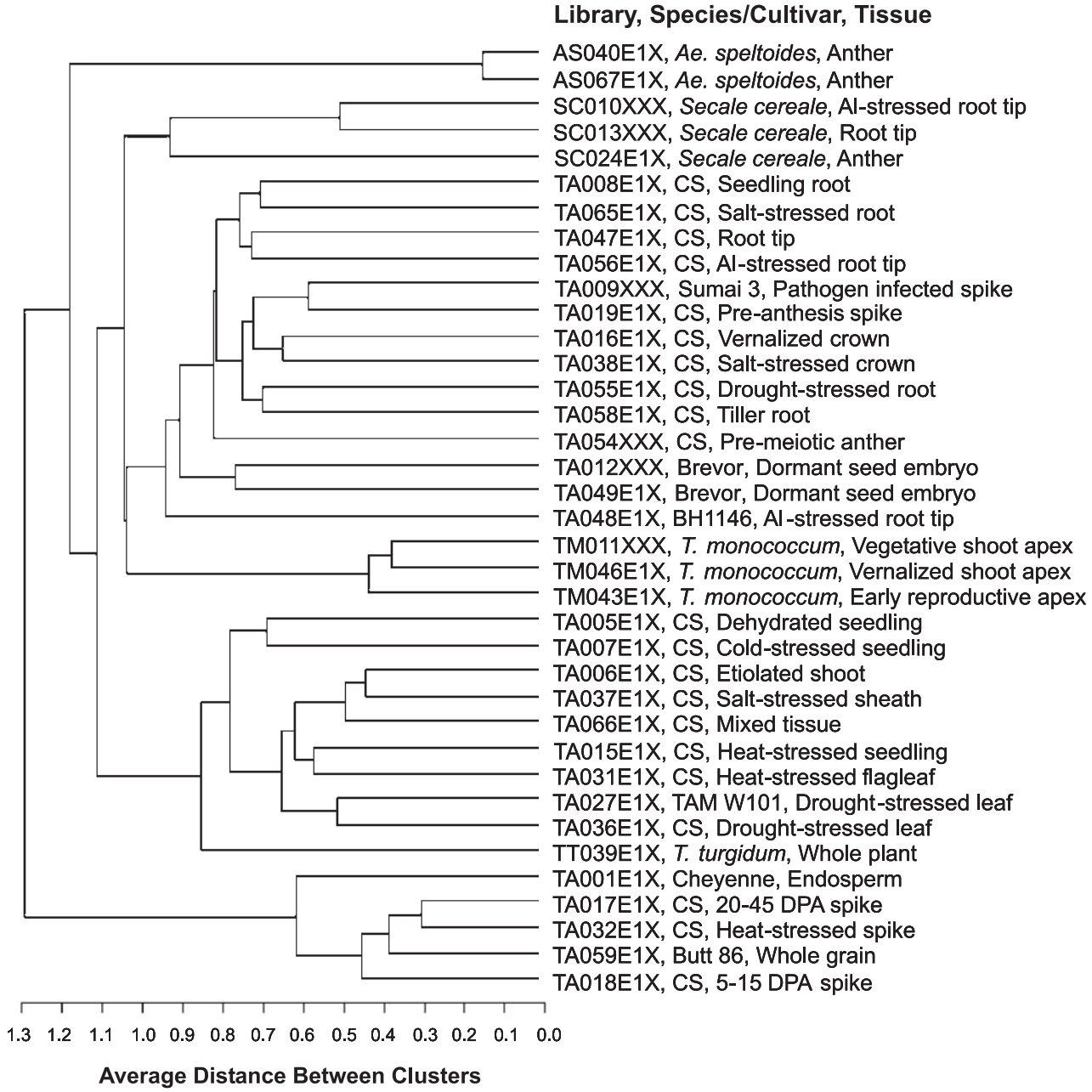
Contig	No. of ESTs	Gene annotation (BLASTX best hit, cutoff 10^{-5})
18876	318	Protein translation factor SU11 homolog
18875	145	γ-gliadin
18874	129	Unknown protein
18873	117	α-gliadin
18872	113	Ribulose bisphosphate carboxylase/oxygenase activase A, chloroplast precursor
18871	110	γ-gliadin
18870	109	Translation initiation factor 5A
18869	103	Elongation factor 1-α
18868	96	Tubulin α-2 chain
18867	93	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit
18866	92	Glyceraldehyde 3-phosphate dehydrogenase, cytosolic
18865	91	*S*-adenosylmethionine decarboxylase precursor
18864	90	*S*-adenosylmethionine decarboxylase precursor
18862	88	Lipid transfer protein
18863	88	Glutenin HMW subunit 1Dy
18861	86	UTP-glucose-1-phosphate uridylyltransferase
18860	84	ADP, ATP carrier protein, mitochondrial precursor
18859	83	*S*-adenosylmethionine synthetase 1
18858	81	HSP80-2
18856	80	Metallothionein-like protein
18857	80	Glyceraldehyde 3-phosphate dehydrogenase, cytosolic
18855	78	40S ribosomal protein S8
18854	77	ADP-ribosylation factor
18853	74	Avenin precursor
18851	72	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit
18852	72	Heat shock protein 70
18850	70	40S ribosomal protein S4
18848	69	Ribulose bisphosphate carboxylase/oxygenase activase A, chloroplast precursor
18849	69	Metallothionein-like protein
18846	68	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit
18847	68	Glyceraldehyde 3-phosphate dehydrogenase, cytosolic
18844	67	40S ribosomal protein S4
18845	67	α-amylase/trypsin inhibitor CM3 precursor
18843	66	Tubulin α-2 chain
18842	64	Metallothionein-like protein
18838	63	Glycine-rich RNA-binding protein GRP1
18839	63	D-TDP-glucose dehydratase
18840	63	Unknown protein
18841	63	Elongation factor 1-α
18836	62	HSP80-2
18837	62	Low-molecular-weight glutenin subunit

expression levels were then calculated as the fraction of ESTs relative to the total number of 5′ ESTs produced in any particular library. Use of the EST fraction in the analysis rather than the actual EST number avoided sampling size bias caused by libraries having different numbers of ESTs generated. Although the depth of library sampling may influence the EST fraction numbers, ESTs included in the analysis were mostly derived from transcripts expressed at moderate to high levels and thus were less sensitive to sampling depth bias than rare ESTs. The resulting 1536 contigs × 37 libraries matrix was the basis for the cluster analysis.

Library clustering and gene expression profile among different tissues of wheat

The similarity of gene expression patterns among the 37 libraries was estimated by Pearson's correlation coefficient for all pairwise library comparisons. Hierarchical clustering of libraries based on a pairwise Euclidean distance matrix derived from the 37 × 37 Pearson correlation matrix was then performed using an UPGMA (unweighted pair group method with arithmetic averaging). A dendrogram displaying library similarity is shown in Fig. 2. The results showed that, in general, all green-tissue libraries, such as seedling, leaf, sheath, and flagleaf, share similar overall gene expression profiles. When their relatedness was assessed in detail, different tissues grown under the same abiotic stress conditions were found to be similar, such as seedling (TA015E1X) and flagleaf (TA031E1X), both under heat stress. Two leaf-tissue libraries treated with a similar drought-stress condition but made from different cultivars, such as TA027E1X from

Fig. 2. The library dendrogram based on the cluster analysis data.



'TAMW101' (a drought-tolerant winter wheat) and TA36E1X *Triticum aestivum* 'Chinese Spring' (CS), were also found to share similar expression patterns. The etiolated shoot library (TA006E1X) grouped with all the green-tissue libraries, indicating that many genes expressed in the green tissues were also expressed under etiolated conditions.

Interestingly, 3 shoot apex libraries derived from *T. monococcum* clustered together: vegetative apex (TM011XXX), early stage of reproductive apex (TM043E1X), and vernalized apex (TM046E1X); however, libraries made from similar tissue but at different developmental stage generally clustered separately. For example, 2 root libraries made from seedling stage (TA008E1X) and tiller stage (TA058E1X) tissue had very different expression profiles. Among 3 spike libraries, spike 20–45 d postanthesis (DPA) (TA017E1X) and spike 5–15 DPA (TA018E1X) were more similar to endosperm (TA001E1X) and developing whole grain (TA059E1X) libraries than to the preanthesis spike library (TA019E1X), which was more similar to another spike library, TA009XXX.

The TA009XXX library was created from an anthesis-stage spike infected with the head blight fungal pathogen *F. graminearum*. Both TA019E1X and TA009XXX were more closely related to root, root-tip, embryo, and crown libraries than to libraries made from spikes in later developmental stages. These results indicate that the gene expression pattern in the preanthesis spike was complex and that many genes expressed during the early stage of spike development were turned off later. Gene content and gene expression pattern in the TA009XXX library have been analyzed in

greater detail previously (Kruger et al. 2002), and genes preferentially expressed during the wheat and fungal pathogen interaction have been identified.

The relations among the 4 root libraries were complex. Seedling root (TA008E1X) expression was similar to salt-stressed seedling root (TA065E1X). While tiller root (TA058E1X) was similar to drought-stressed tiller root (TA055E1X), they were both closer to other tissue types, such as crown and spike, than to the root libraries. The diverse expression patterns were most prominent among root-tip libraries under aluminum stress between species and cultivars. The 2 rye root-tip libraries, control SC013XXX and aluminum-stressed SC010XXX, clustered together but separately from 2 CS wheat root-tip libraries, the control TA047E1X and the aluminum-stressed TA056E1X, which were closer together; however, another aluminum-stressed root-tip library, TA048E1X, made from a highly aluminum-tolerant spring wheat cultivar 'BH1146', was clustered in a separate clade from all other root-tip libraries. This indicates that the TA048E1X library had a distinctively different expression profile from all the other root-tip libraries analyzed. The results also showed that different tissues can respond very differently to the same stress condition, as found among CS salt-stressed root (TA065E1X), CS salt-stressed crown (TA038E1X), and CS salt-stressed sheath (TA037E1X) libraries, implying that the stress-response genes may be expressed in a tissue-specific manner.

Anther tissue also showed complex gene expression pattern. Three species were used to construct anther libraries (rye, *Ae. speltoides*, and CS wheat). The 2 *Ae. speltoides* libraries clustered together, as expected, since similar starting material was used. However, the CS anther library was more similar to root, root-tip, crown, and spike libraries of hexaploid wheat. In contrast, the rye anther library was more similar to the 2 rye root-tip libraries than to the diploid and hexaploid wheat anther libraries. Though these anther libraries from 3 different species were made from premeiotic anthers, the expression profiles were drastically different from each other. This huge difference is attributed most likely to genetic variation among species. The cluster analysis grouped libraries from the same species closer than with other species, as observed among 3 shoot apex libraries from *T. monococcum* (TM011XXX, TM043E1X, and TM046E1X) and 3 libraries from rye (two root-tip libraries, SC010XXX and SC013XXX, and the anther library, SC024E1X). This is the result of the parameters used for the EST assembly, which allowed for sequences sharing greater than 90% similarity to form a contig, while orthologous sequences with less than 90% similarity were assigned to separate contigs. Taken together, the results clearly indicated that gene expression profile studies using cluster analysis based on contig assembly data were best carried out on a species-specific basis. Therefore, the gene expression profile analyses presented here will be informative and accurately inferred only for ESTs derived from hexaploid wheat.

Contig clustering, coexpression of genes, and differential gene expression in wheat

To identify ESTs clustered in different contigs that were coexpressed and shared a similar expression pattern among various tissues, cluster analysis used for library clustering was applied for 1536 contigs. A dendrogram based on a 1536 × 1536 distance matrix was generated. Fig. 3 is the 2-dimensional representation displaying the relatedness of ESTs in 1536 contigs and their differential expression profiles among 37 libraries. The results revealed that genes with related functions expressed similarly among different libraries tended to be grouped together. The level of gene expression in each contig was expressed as the EST fraction contributed from each library, as indicated by the shades of color (Fig. 3). Genes expressed predominantly in a tissue-specific manner appeared in higher intensity of color on the display.

To view differential gene expression in greater detail, the contig dendrogram was divided into 21 major clusters according to the level of differential expression. Each of the resulting 21 clusters represented a group of genes that were coexpressed preferentially in a specific tissue. Five clusters, 5, 6, 18, 19, and 20, containing genes expressed primarily in species other than hexaploid wheat were excluded from further discussion. The remaining 16 clusters showing differential gene expression pattern among different wheat tissues are summarized in Fig. 4. It is obvious that 2 clusters, 1 (Fig. 4B) and 21 (Fig. 4C), corresponding to 2 different groups of tissues, contain genes expressed at a very high level. Cluster 1 contained genes expressed predominantly in green tissues. The gene annotations indicated that a majority of these genes were found to be associated with photosynthesis and chloroplast biosynthesis. Cluster 21 contained genes that were highly expressed in spikes, endosperm, and developing grain tissues. The 20–45 DPA spike (TA017E1X) library alone contributed 48% of its total ESTs to cluster 21, whereas the endosperm (TA001E1X) and developing grain (TA059E1X) libraries contributed moderate numbers of ESTs, at 25% and 22%, respectively. In contrast, 15% of ESTs were from the 5–15 DPA spike (TA018E1X) library, and less than 5% came from the preanthesis spike (TA019E1X) library. Most of these genes encoded for storage proteins and proteins involving seed development. The dynamic changes in gene expression levels observed on the basis of EST count are closely correlated with actual changes of gene activities during different stages of seed development, as observed by Clarke et al. (2000).

The anther (TA054XXX) and preanthesis spike (TA019E1X) libraries contained genes expressed in most of the clusters but at very low levels, reflecting the sequence diversity found in these libraries and the relatively lower number of larger contigs. A complete list of contigs organized by cluster with EST member tabulation in fraction and their gene annotations is provided in the supplementary material (Table S1)[4].

Of the 16 clusters analyzed, 12 contained genes expressed predominantly in 1 of the 12 specific tissues. For example, clusters 8, 11, 13, 15, 16, and 17 (Fig. 4A) correspond to genes expressed in salt-stressed seedling root, drought-stressed tiller root, tiller root, root-tip, aluminum-stressed CS root-tip, and aluminum-stressed 'BH1146' root-tip. Clus-

[4] Supplementary data for this article are available on the Web site or may be purchased from the Depository of Unpublished Data, Document Delivery, CISTI, National Research Council Canada, Building M-55, 1200 Montreal Rd., Ottawa, ON K1A 0R6, Canada. DUD 5049. For more information on obtaining material refer to http://cisti-icist.nrc-cnrc.gc.ca/irm/unpub_eshtml.

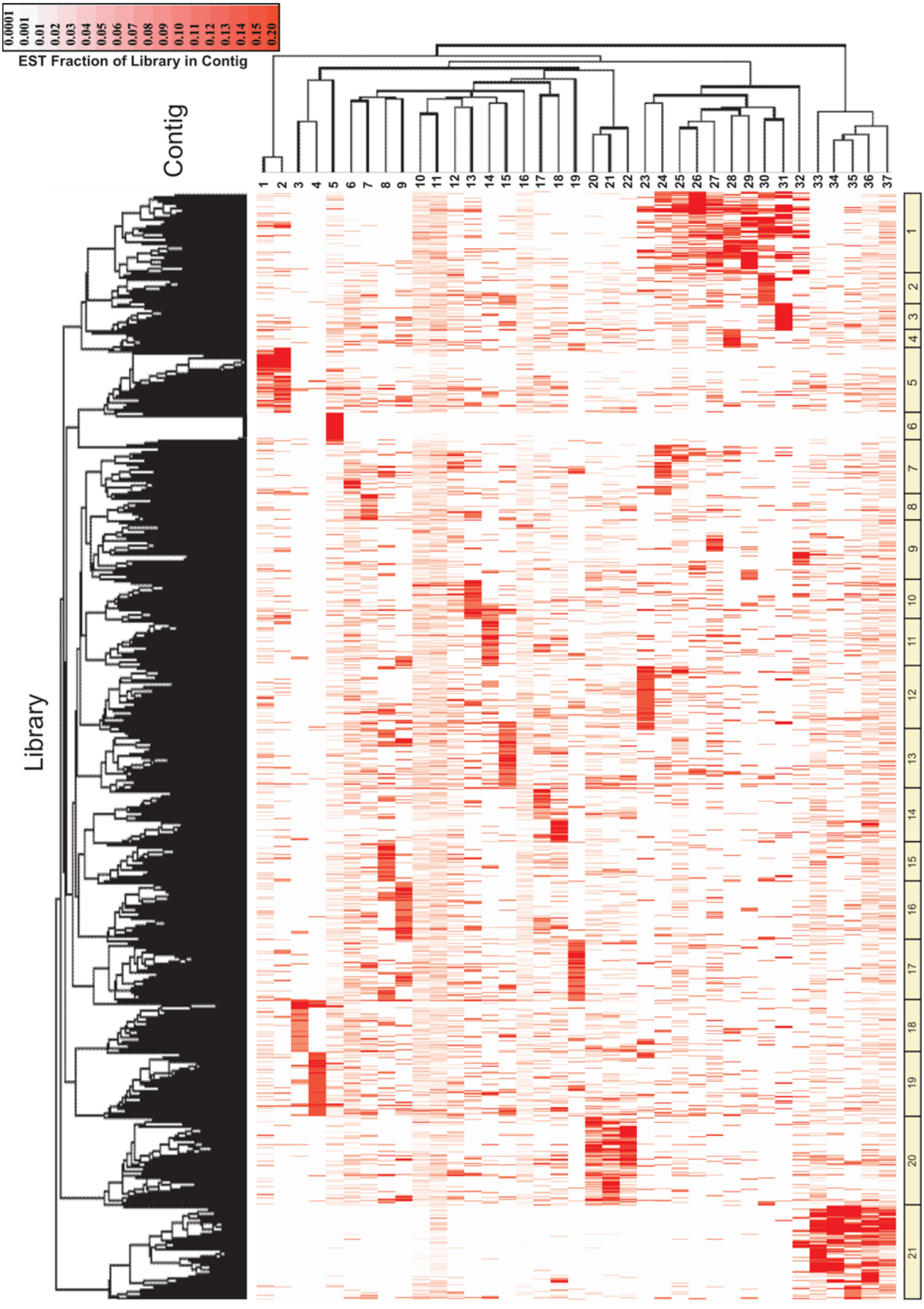


Fig. 3. The 2-dimensional graphic display based on the results from the cluster analysis of both libraries and contigs. Twenty-one major clusters shown on the side bars were formed according to the degree of gene expression in different tissues. The color scale indicates the fraction of ESTs in each cDNA library contributing to each contig.

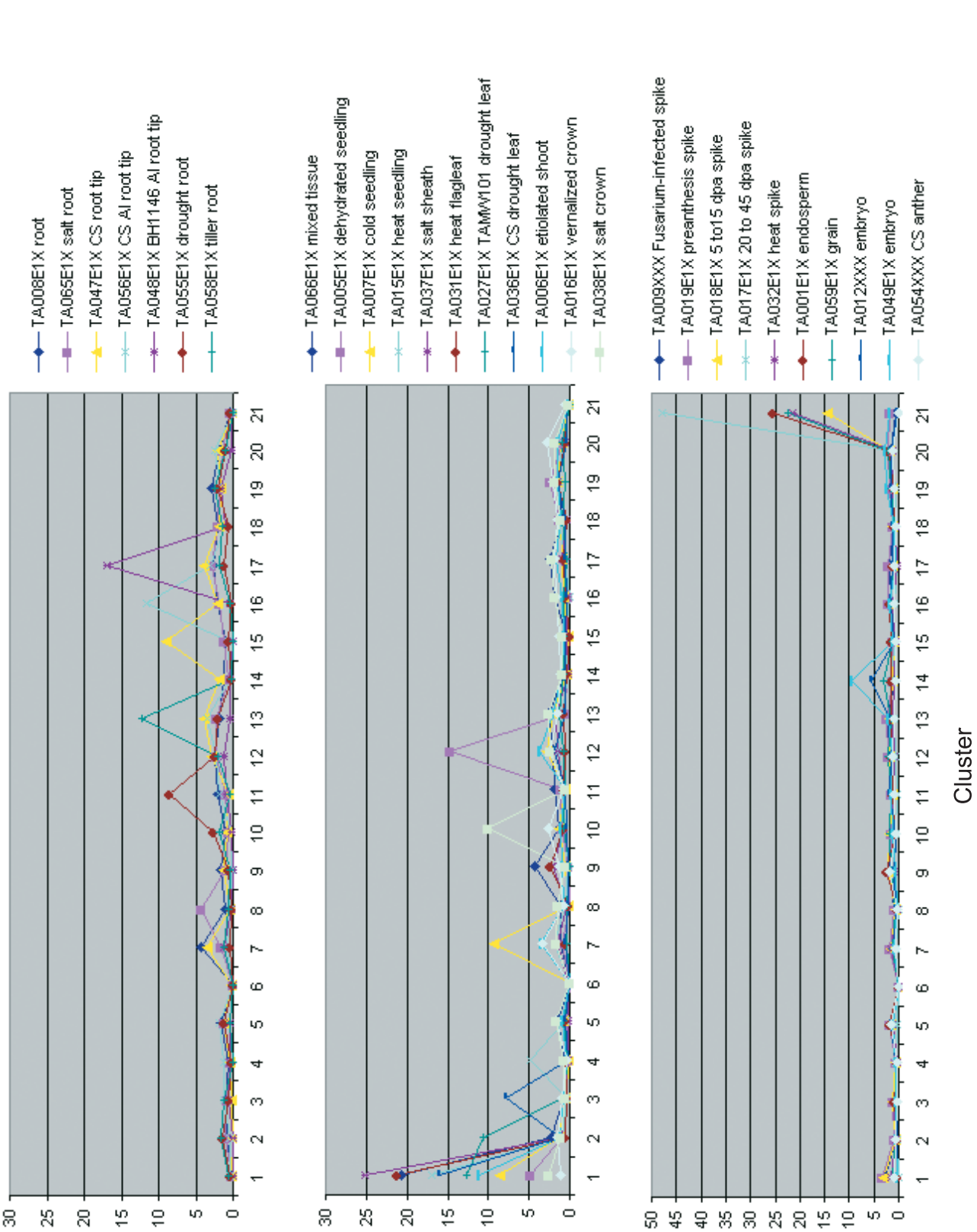


Fig. 4. Wheat gene expression patterns in 21 clusters and their corresponding tissues. (A) The expression patterns for all root and root-tip tissues. (B) The expression patterns for all green tissues. (C) The expression patterns for spike, grain, embryo, and anther tissues. Levels of gene expression in each cluster were measured as the percentage of EST in each library calculated by summing EST members from all contigs in that cluster and divided by the total number of 5′ ESTs produced in that library.

Table 3. Probability ($p < 0.05$) of gene upregulation in root and shoot tissues under various stresses, and gene annotations (cutoff 10^{-5}).

(*a*) Pair-wise comparison between TA008E1X, seedling root and TA065E1X, salt-stressed root, cluster 8		
Contig no.	Probability	Gene annotation (BLASTX best hit, cutoff 10^{-5})
16895	1.53×10^{-4}	Pathogenesis-related protein 1.2
18582	1.53×10^{-4}	Ubiquinol-cytochrome *c* reductase iron–sulfur subunit, mitochondrial precursor
18147	4.59×10^{-4}	Putative eukaryotic translation initiation factor 3 large subunit
18300	1.24×10^{-2}	Similar to NADP-dependent malic enzyme, chloroplast precursor
18380	1.24×10^{-2}	Metallothionein-like protein
18206	1.97×10^{-2}	Putative phospho-2-dehydro-3-deoxyheptonate aldolase 1, chloroplast precursor
18458	3.71×10^{-2}	Unknown protein
17903	3.71×10^{-2}	Unknown protein
18365	4.53×10^{-2}	*S*-adenosylmethionine synthetase 1
18485	4.53×10^{-2}	Putative cytochrome c1 precursor
18406	4.54×10^{-2}	Actin-depolymerizing factor 3
(*b*) Pair-wise comparison between TA006E1X, seedling shoot and TA038E1X, salt-stressed crown, cluster 10		
Contig no.	Probability	Gene annotation (BLASTX best hit, cutoff 10^{-5})
18722	6.62×10^{-4}	Lipid transfer protein homolog
18852	3.88×10^{-3}	Heat shock protein 70
17280	7.60×10^{-3}	Lipid transfer protein homolog
18650	7.60×10^{-3}	Lipid transfer protein
18842	7.60×10^{-3}	Metallothionein-like protein
18795	1.01×10^{-2}	Lipid transfer protein
18680	1.01×10^{-2}	*S*-adenosylmethionine decarboxylase
16886	2.57×10^{-2}	Unknown protein
17006	2.57×10^{-2}	Glutathione *S*-transferase
18840	2.57×10^{-2}	No hit
17548	2.57×10^{-2}	At5g26751.1, shaggy-like kinase α
18360	2.57×10^{-2}	*S*-adenosylmethionine decarboxylase
17875	2.57×10^{-2}	Unknown protein
18091	2.57×10^{-2}	At3g17020.1, unknown
18154	2.57×10^{-2}	At1g10840.1, putative translation initiation factor
18368	2.57×10^{-2}	Eukaryotic initiation factor 4A
17480	2.57×10^{-2}	No hit
18018	2.57×10^{-2}	Na^+/H^+ antiporter
17525	2.57×10^{-2}	Phosphoethanolamine methyltransferase
18761	2.69×10^{-2}	No hit
(*c*) Pair-wise comparison between TA006E1X, seedling shoot and TA005E1X, drought-stressed seedling, cluster 12		
Contig no.	Probability	Gene annotation (BLASTX best hit, cutoff 10^{-5})
18686	4.56×10^{-3}	Catalase isozyme 2
18760	4.56×10^{-3}	Adenosylhomocysteinase
17913	1.76×10^{-2}	Proteasome subunit α type 5
18327	1.76×10^{-2}	Glutathione *S*-transferase 2
18700	1.76×10^{-2}	Enolase
17710	1.76×10^{-2}	Late embryogenesis abundant protein, group 3
18490	1.76×10^{-2}	No hit
18767	1.76×10^{-2}	Plasma membrane H^+-ATPase
18844	1.76×10^{-2}	40S ribosomal protein S4
18068	1.81×10^{-2}	At4g31290.1, unknown protein
(*d*) Pair-wise comparison between TA006E1X, seedling shoot and TA015E1X, heat-stressed seedling, cluster 4		
Contig no.	Probability	Gene annotation (BLASTX best hit, cutoff 10^{-5})
17859	5.09×10^{-3}	Probable peptidylprolyl isomerase
18536	5.09×10^{-3}	Tubulin β chain
18746	5.09×10^{-3}	Unknown protein
18817	1.73×10^{-2}	Tubulin α-3 chain
17790	1.91×10^{-2}	Ribosomal protein L17
17423	1.91×10^{-2}	Translation initiation factor 5A
18175	1.91×10^{-2}	Transmembrane protein
18527	1.91×10^{-2}	Putative 14-3-3 protein

Table 3 (*concluded*).

(*e*) Pair-wise comparison between TA006E1X, seedling shoot and TA007E1X, cold-stressed seedling, cluster 7		
Contig no.	Probability	Gene annotation (BLASTX best hit, cutoff 10^{-5})
18821	1.65×10^{-2}	Glycine-rich RNA-binding protein, low-temperature-responsive
18838	9.97×10^{-3}	Glycine-rich RNA-binding protein
18596	2.54×10^{-2}	No hit
18696	2.54×10^{-2}	40S ribosomal protein S3A
17180	2.64×10^{-2}	Unknown protein
18445	2.86×10^{-2}	Glycine-rich RNA-binding protein, low-temperature-responsive
18691	2.86×10^{-2}	Glycine-rich RNA-binding protein, low-temperature-responsive

ters 2, 3, 4, 7, 10, and 12 (Fig. 4B) correspond to 'TAMW101' drought-stressed leaf, CS drought-stressed leaf, heat-stressed seedling, cold-stressed seedling, salt-stressed crown, and dehydrated seedling shoot. Most of these tissues have been treated with various abiotic stresses. To determine which gene expression has been upregulated in which tissue in response to specific stresses, a pairwise comparison method developed by Audic and Claverie (1997) was applied. This method allowed the likelihood of changes of gene expression to be estimated statistically by comparing the levels of gene expression through EST count between normal tissues and tissues with stress treatment. Table 3 summarizes a list of gene candidates showing significant changes of gene expression ($p < 0.05$) in root and seedling tissues under salt, drought, heat, and cold stresses.

Under salt stress, the results show that salt-responsive genes in root and in crown tissue share no detectable overlap. This suggests that the induction of salt-responsive genes may be regulated in a tissue-specific manner. Similar findings based on microarray studies have been reported in *Arabidopsis* (Kreps et al. 2002) and barley (*Hordeum vulgare* L.) (Ozturk et al. 2002). In salt-stressed roots, in addition to stress-responsive genes (such as pathogenesis-related protein and metallothionein protein), 2 genes encoding for chloroplast proteins and 1 for mitochrondrial protein were also induced, which may be used to sustain plant growth under salt-stress conditions. In contrast, genes that are highly induced in crown tissue under salt stress are lipid transfer proteins. Additional genes that are salt responsive in crown tissue include heat-shock protein 70, a molecular chaperone; shaggy-like kinase known to respond to osmotic change; and a Na^+/H^+ antiporter known to be involved in ionic homeostasis during salt stress. These genes have been reported to be activated in signal transduction pathways under salt stress (Zhu 2002).

Table 3 summarizes genes that are upregulated in seedlings under salt, drought, heat, and cold stresses, which have no apparent overlap in gene transcript among all 4 different stress responses. Previous studies have suggested gene overlaps in signal transduction pathways under different stress responses (Cooper et al. 2003). In microarray studies, it has been found that some transcripts were shared, but a majority of the expression changes were stress-specific (Kreps et al. 2002; Ozturk et al. 2002). In drought-stressed seedling tissue, the genes found to be associated with drought stress included catalase, LEA (late embryogenesis abundant) protein, glutathione-*S*-transferase, and plasma membrane H^+-ATPase (Table 3). Under cold stress, genes found to be highly induced were low-temperature-responsive glycine-rich RNA-binding proteins, indicating their importance in cold acclimation in wheat. Under heat stress, in addition to expression levels of genes involving protein synthesis and cell growth, gene expression for the transmembrane protein and the 14-3-3 protein involved in signal transduction pathways was affected.

To demonstrate how genetic differences influenced the level of gene expression for stress tolerance, pairwise comparisons between 2 genotypes of wheat under the same stress treatment were carried out for drought-stressed leaf and aluminum-stressed root tip (Table 4). For drought-stress response, no genes with significant upregulation were found in CS, but several transcripts were highly induced in the drought-tolerant cultivar 'TAMW101'. Similar genes have also been found to be associated with drought-stress response in barley (Ozturk et al. 2002). For aluminum-stress response, gene expression levels were compared between CS and an aluminum-tolerant cultivar 'BH1146'. One transcript was found with significantly higher expression in CS, whereas several transcripts were responsive to aluminum stress and found to be highly expressed in 'BH1146'. Genes encoding for oxalate oxidase and peroxidase have previously been found to be associated with aluminum toxicity but not to the cultivar's tolerance to aluminum stress (Hamel et al. 1998). In this study, 2 genes were identified as being highly expressed in 'BH1146' ($p < 0.001$) but with no known functions. Further analyses are necessary in order to confirm their role in aluminum tolerance.

From this analysis, genes that are responsive to stress growth conditions were predicted. Additional experiments are needed, but the current analyses are useful in helping to identify possible genes responsible for stress tolerance from the large amounts of EST data in wheat. Stress tolerance is often a quantitative trait and is difficult to map to a precise location on chromosomes. ESTs found with elevated levels of gene expression under specific stress conditions in this analysis may provide a source of DNA markers that can be used to find an association with quantitative trait loci controlling stress tolerance in wheat.

In this study the gene expression patterns for ESTs originating from different homoeologous chromosomes were analyzed separately. The gene annotations were often found to be shared among multiple contigs, as shown in Tables 2, 3, and 4. It is possible that the ESTs formed in separate contigs could correspond to the same gene; EST overlaps, if provided, should link them together. It is also likely that the contigs may contain ESTs derived from different homoeologous chromosomes. Further experiments are required to differentiate between these 2 possibilities and to shed some

Table 4. Effect of wheat genotypes on stress responses.

(*a*) Drought-stressed leaf – CS (TA036E1X) vs. TAMW101 (TA027E1X); TA036E1X, CS, cluster 3		
Contig No.	Probability	Gene annotation (BLASTX best hit, cutoff 10^{-5})
none	—	—
(*b*) Drought-stressed leaf – CS (TA036E1X) vs. TAMW101 (TA027E1X); TA027E1X, TAMW101, cluster 2		
Contig No.	Probability	Gene annotation (BLASTX best hit, cutoff 10^{-5})
18687	4.83×10^{-3}	No hit
18350	8.99×10^{-3}	Putative acid phosphatase
18547	1.76×10^{-2}	r40g3 protein, ABA, salt-stressed responsive cDNA
17488	3.45×10^{-2}	Rab15B protein, dehydrin-like protein
18864	3.45×10^{-2}	*S*-adenosylmethionine decarboxylase precursor
18499	3.79×10^{-2}	Unknown protein
18480	3.98×10^{-2}	Lipid transfer protein precursor
18587	6.94×10^{-2}	Glutathione transferase
(*c*) Aluminum-stressed root tip – CS (TA056E1X) vs. BH1146 (TA048E1X); TA056E1X, CS, cluster 16		
Contig No.	Probability	Gene annotation (BLASTX best hit, cutoff 10^{-5})
17087	1.76×10^{-2}	Oxalate oxidase GF-2.8 precursor
(*d*) Aluminum-stressed root tip – CS (TA056E1X) vs. BH1146 (TA048E1X); TA048E1X, BH1146, cluster 17		
Contig No.	Probability	Gene annotation (BLASTX best hit, cutoff 10^{-5})
18533	7.44×10^{-2}	Unknown protein
18598	6.33×10^{-2}	At2g42210.1, unknown protein
17340	2.80×10^{-3}	At1g76680.1, 12-oxophytodienoate reductase
17729	2.80×10^{-3}	Peroxidase 2 precursor
18359	2.80×10^{-3}	Delta-type tonoplast intrinsic protein
18453	2.80×10^{-3}	Peroxidase 2 precursor
18824	1.38×10^{-2}	Similar to caffeoyl-CoA 3-*O*-methyltransferase

Note: Probability ($p < 0.05$) of gene upregulation in a comparison of 2 wheat cultivars for drought-stressed leaves and aluminum-stressed root-tips, and gene annotations (cutoff 10^{-5}).

light on whether the gene expression for homoeologous ESTs would differ spatially and temporally.

The gene expression profile analyses presented here have predicted the possible functionalities of 4% of the total wheat unigenes. Because of the need to have sufficient ESTs in each contig to give statistically meaningful results, only 4% of the unigenes were considered in this analysis. The remaining 96% of the unigenes expressed in moderate to low levels await more advanced technologies and analyses, such as microarrays, to explore their functional roles and gene expression patterns. Nonetheless, it is clear that the EST data generated in this project provide a diverse and rich source for gene discovery in wheat.

Acknowledgements

This paper is based upon research supported by the National Science Foundation Cooperative Agreement no. DBI-9975989. The authors thank Dr. Roger Thilmony at USDA-ARS, Albany, Calif., for suggestions and assistance in using the TreeView software and in preparing Fig. 3.

References

Adams, M.D., Kelley, J. M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., et al. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science (Washington, D.C.) **252**: 1651–1656.

Akhunov, E.D., Akhunov, A.R., Linkiewicz, A.M., Dubcovsky, J, Hummel, D., Lazo, G.R., et al. 2003*a*. Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. Proc. Acad. Natl. Sci. U.S.A. **100**: 10836–10841.

Akhunov, E.D., Goodyear, J.A., Geng, S., Qi, L.L., Echalier, B., Gill, B.S., et al. 2003*b*. The organization and rate of evolution of the wheat genomes are correlated with recombination rates along chromosome arms. Genome Res. **13**: 753–763.

Anderson, O.D., and Greene, F.C. 1989. The characterization and comparative analysis of high-molecular-weight glutenin genes from genomes A and B of a hexaploid bread wheat. Theor. Appl. Genet. **77**(5): 689–700.

Audic, S., and Claverie, J.-M. 1997. The significance of digital gene expression profiles. Genome Res. **7**: 986–995.

Clarke, B.C., Hobbs, M., Skylas, D., and Appels, R. 2000. Genes active in developing wheat endosperm. Funct. Integr. Genomics, **1**: 44–55.

Cooper, B., Clarke, J.D., Budworth, P., Kreps, J., Hutchison, D., Park, S. et al. 2003. A network of rice genes associated with stress response and seed development. Proc. Natl. Acad. Sci. U.S.A. **100**: 4945–4950.

Dvorak, J., and Akhunov, E.D. 2005. Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the Aegilops-Triticum alliance. Genetics, **171**: 323–332.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Bostein, D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. **95**: 14863–14868.

Ewing, R.M., Kahla, A.B., Poirot, O., Lopez, F., Audic, S., and Claverie, J.-M. 1999. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Res. **9**: 950–959.

Gustafson, J.P., McGuire, P.E., and Qualset, C.O. 2004. Genomic resources: moving wheat into the mainstream in plant genomics. Genetics, **168**: 583–584.

Hamel, F., Breton, C., and Houde, M. 1998. Isolation and characterization of wheat aluminum-regulated genes: possible involvement of aluminum as a pathogen response elicitor. Planta, **205**: 531–538.

Kreps, J.A., Wu, Y., Chang, H.-S., Zhu, T., Wang, X., and Harper, J.F. 2002. Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. Plant Physiol. **130**: 2129–2141.

Kruger, W.M., Pritsch, C., Chao, S., and Muehlbauer, G.J. 2002. Functional and comparative bioinformatics analysis of expressed genes from wheat spikes infected with *Fusarium graminearum*. Mol. Plant–Microbe Interact. **15**: 454–455.

Lazo, G.R., Chao, S., Hummel, D.D., Edwards, H., Crossman, C.C., Lui, N., et al. 2004. Development of an expressed sequence tag (EST) resource for wheat (*Triticum aestivum* L.): EST generation, unigene analysis, probe selection, and bioinformatics for a 16,000-locus bin-delineated map. Genetics, **168**: 585–593.

Ogihara, Y., Mochida, K., Nemoto, Y., Murai, K., Yamazaki, Y., Shin-I, T., and 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.

Ohlrogge, J., and Benning, C. 2000. Unraveling plant metabolism by EST analysis. Curr. Opin. Plant Biol. **3**: 224–228.

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y., and Matsubara, K. 1992. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nat. Genet. **2**: 173–179.

Ozturk, Z.N., Talame, V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., Tuberosa, R., and Bohnert, H.J. 2002. Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol. Biol. **48**: 551–573.

Qi, L.L., Echalier, B., Chao, S., Lazo, G.R., Butler, G.E., Anderson, O.D., et al. 2004. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics, **168**: 701–712.

Ronning, C.M., Stegalkina, S.S., Ascenzi, R.A., Bougri, O., Hart, A.L., Utterbach, T.R., et al. 2003. Comparative analyses of potato expressed sequence tag libraries. Plant Physiol. **131**: 419–429.

Rubin, G.M., Hong, L., Brokstein, P., Evans-Holm, M., Frise, E., Stapleton, M., and Harvey, D.A. 2000. A *Drosophila* complementary DNA resource. Science (Washington, D.C.), **287**: 2222–2224.

Sorrells, M.E., La Rota, M., Bermudez-Kandianis, C.E., Greene, R.A., Kantety, R., Munkvold, J.D., et al. 2003. Comparative DNA sequence analysis of wheat and rice genomes. Genome Res. **13**: 1818–1827.

Whetten, R., Sun, Y.-H., Zhang, Y., and Sederoff, R. 2001. Functional genomics and cell wall biosynthesis in loblolly pine. Plant Mol. Biol. **47**: 275–291.

Zhang, D., Choi, D.W., Wanamaker, S., Fenton, R.D., Chin, A., Malatrasi, M., et al. 2004. Construction and evaluation of cDNA libraries for large-scale expressed sequence tag sequencing in wheat (*Triticum aestivum*, L.). Genetics, **168**: 595–608.

Zhu, J.-K. 2002. Salt and drought stress signal transduction in plants. Annu. Rev. Plant Biol. **53**: 247–273.