Cereal Research Communications 39(4), pp. 475–486 (2011)
DOI: 10.1556/CRC.39.2011.4.2

Functional Characterization of Three cDNA Libraries from the Diploid Wheat *Triticum monoccocum* (A^mA^m) with Different Growth Habits

M.L. DÍAZ[1]*, I. GARBUS[2,3]*, B. STAMOVA[4] and V.C. ECHENIQUE[2,5]**

[1]Departamento de Biología, Bioquímica y Farmacia, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina
[2]Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS) – CONICET, Bahía Blanca, Argentina
[3]Departamento de Ciencias de la Salud, UNS, Bahía Blanca, Argentina
[4]Medical Investigation of Neurodevelopmental Disorders Institute, UC Davis Medical Center, Sacramento, CA 95817, USA
[5]Departamento de Agronomía, UNS, Bahía Blanca, Argentina

(Received May 12 2010; accepted 30 March 2011;
Communicated by H. Bürstmayr)

The diploid wheat *Triticum monococcum* (2n = 2x = 14, A^mA^m) constitutes an excellent and simplified model to study wheat genome expression since it has lower genome complexity compared to polyploid wheat species.

The analysis of a collection of near 10,000 ESTs obtained from two libraries of *T. monococcum* (TM011XXX and TM043E1X) and one from *T. monococcum* ssp. *aegilopoides* (TM046E1X), lead to 2246, 1843 and 2405 unigenes, respectively. The unigenes from each library were individually analyzed through the Blast2Go interface revealing higher identity compared to EST databases. However, a low percentage of sequences showed significant similarity with *Triticum* databases, reflecting that several novel *Triticum* expressed genes have been identified in these libraries. The sequences annotation and classification under GO categories showed similar distribution for the three libraries, but differences were identified in some subcategories.

When a virtual unique library was constructed including the sequences from the three libraries, seventeen genes were found to be over-expressed between reproductive and vegetative tissues and thirty five showed genotype-specific expression. These differentially expressed genes included several regulatory proteins and some unreported sequences. Some of these differentially expressed genes were validated by qRT-PCR. The exploration for specific polymorphisms allowed the identification of several SNPs.

These EST libraries constitute a valuable tool for the assessment of active metabolic pathways in vegetative and reproductive tissues in diploid wheat species. A more detailed assessment of expression profiles of genotypes with spring and winter growth habits would allow the identification of the transcriptomic strategies developed under the pressure of these antagonistic growth conditions.

Keywords: diploid wheat, cDNA libraries, ESTs, Digital Differential Expression, qRT-PCR

* Díaz and Garbus contributed equally to this manuscript
** Corresponding author; Phone: 54-291-4861666; Fax 54-291-4862882; E-mail: echeniq@criba.edu.ar

Introduction

Large scale EST sequencing projects have emerged as a valuable tool to the conduction of studies aiming to the identification of genes involved in several biological processes as well as evolutive and phenotypic analysis (Lazo et al. 2004; Houde et al. 2006) and the identification of genetic markers with direct application in mapping projects (Eujayl et al. 2002; Gadaleta et al. 2009).

The availability of EST sequences has grown exponentially and the access to them has been increased during the past decade such that, as in January 2010, there were >64 millions public EST sequence accessions in the National Center for Biotechnology Information (NCBI) dbEST database, representing nearly the 75% of the sequences deposited in GenBank (www.ncbi.nlm.nih.gov/dbEST/index.html). More than one million ESTs belong to the genus *Triticum*.

Hexaploid wheat (2n = 6x = 42, AABBDD) contains three nuclear genomes with nearly 16,700 Mb (Bennett and Leitch 1995), with ~80% of repetitive non-coding DNA (McCarthy et al. 2002). Since the A genome found in the diploid wheat species *Triticum monococcum* (2n = 2x = 14, A^mA^m) and in other modern wheat species with different ploidy level share a common ancestor (Huang et al. 2002; Chantret et al. 2005), its lower genome complexity positions this diploid species as a simplified model to study wheat genome expression.

Here, we reported three cDNA libraries from *T. monococcum*, two from different tissues of cultivated *T. monococcum* (DV92) and the other one from the wild *T. monococcum* ssp. *aegilopoides* (G3116). The primary goals of this research were (a) the characterization of the libraries obtained from vegetative and reproductive apexes from winter and spring varieties of *T. monococcum,* b) the annotation and classification of the ESTs from each library according to GO categories; c) the identification of genes differentially expressed among genotypes and/or tissues; and, d) the detection of genotype-specific gene markers. The low percentage of sequences with significant similarity with *Triticum* databases compared to that of model species reflects that novel *Triticum* expressed genes have been identified in these libraries, constituting an important addition to publicly available resources, especially in relation to the study of species with different growth habits.

Materials and Methods

Plant materials

Diploid wheat, *T. monoccocum* (A^mA^m), genotypes DV92 (cultivated) and G3116 (wild), were the primary source of tissues for the cDNA libraries construction. Plants were grown in the greenhouse at 25°C under a 16/8 (light/dark) photoperiod cycle. For the construction of cDNA library TM011XXX, vegetative apexes of one month old *T. monococcum* genotype DV92 were collected. The same genotype was used for the cDNA library TM043E1X, which was constructed from early reproductive differentiated apexes of six-week-old plants (double-ridge stage to the terminal-spikelet stage). The other cDNA library, TM046E1X was constructed from *T. monoccocum* genotype G3116. Apexes were

collected from plants under different stages of the vernalization process starting with 4-week-old plants and after 2, 4 and 6 weeks of growing at 8°C, under a 16/8 (light/dark) photoperiod cycle. For the three libraries, the material was collected under the stereomicroscope and stored immediately into liquid nitrogen.

cDNA libraries construction and quality evaluation

General details of libraries construction were reported in Zhang et al. (2004). Briefly, total RNA was prepared using the TRIZOL (GIBCO BRL, Grand Island, NY) method. Poly(A) mRNA was purified using the PolyATract mRNA Isolation System (Promega, Madison, WI). cDNAs with *Eco*RI on the 5'-end and *Xho*I on the 3'-end were synthesized using the ZAP-cDNA synthesis kit (Stratagene). cDNAs >0.5 kb were selected using size fractionation by gel filtration and then directionally cloned into the Uni-ZAPXR (Stragene) (TM046E1X and TM043E1X) or into the ZAP-Express vector (Stratagene) (TM011XXX). Recombinants were packaged using GigaPack III Gold packaging extract (Stratagene). A mass excision was prepared *in vivo* from the unamplified libraries using the host strain XL1-Blue-MRF' and the helper phage ExAssist (Stratagene) to produce pBluescript phagemid populations. For the TM011XXX library, the phagemids were transferred to three different types of cells to evaluate them for plasmid production. *E. coli* strains XLOLR, DH10B and DH5α were compared for growth and plasmid yield, showing DH5α cells to be the better choice. For libraries TM043E1X and TM046E1X, the phagemids were transferred to XLOLR cells. Mass excision was performed at 37°C for 3 hr with a high ratio of recipient cells to primary λ-phage and a high multiplicity of infection of helper phage to the same host cells. Cultures were pelleted and supernatant was heated at 70°C for 20 min to create a "low amplification" phagemid population.

Quality libraries evaluation was performed as described in Zhang et al. (2004). There was a strong positive relationship between the size of the primary λ-Hlibrary and the frequency of clones that carried fragments of the *E. coli* or phage λ-Hgenome, which were assumed to be low level contaminants inherent in all the cloning systems that were used.

Sequence analyses

EST sequencing was performed in O. Anderson's laboratory (USDA, Albany, CA, USA). The original EST sequence data were processed and assembled using the HarvEST:Wheat pipeline (Close et al. 2004; http://harvest.ucr.edu/). Sequence and quality files were generated applying Phred (version 0.020425.c, Ewing and Green 1998; http://www.phrap.org/). Masking of cloning system sequences were performed through the use of cross_match (version 0.990329, http://www.phrap.org/). Then, removal of low quality regions outside of a sliding window with an average Phred quality value of 17, reduction of poly(T) or poly(A) ends to a maximum of 17 consecutives T or A nucleotides, and removal of residual cloning system sequences was performed with an in-house script "qvtrim". Sequences smaller than 100 bp or corresponding to rRNA, mitochondrial or chloroplast DNA were excluded from analysis. Poor quality sequencing were filtered based on frequency of four-nucleotide repeats. EST sequences were deposited in GenBank (TM011XXX: accessions BE491974 to BE493750 and BF199502 to

BF200506; TM043E1X: accessions CD490311 to CD492271 and BG314023 to BG314454; TM046E1X: accessions BQ800633 to BQ803995). BLASTN (Altschul et al 1997; http://www.ncbi.nlm.nih.gov/BLAST/) searches confirmed the quality of the sequences and discarded contamination.

ESTs were assembled using CAP3 software (Huang and Madan 1999) with stringent settings (p = 95, d = 60, f = 100, h=50). All individual ESTs and all consensus sequences were compared using BLASTX to the NCBI translated nonredundant (nr) database, the translated rice genome from TIGR, the TIGR translated *Arabidopsis* genome and with SwissProt (SwPt) database.

The functional annotation inferred by sequence similarity was referred to the vocabulary proposed by Gene Ontology (GO; Ashburner et al. 2000), that includes the three classical ontological descriptions, i.e., molecular function, biological process and cellular component.

Validation of differential expression

Differential genes expression was validated though the quantitative real time PCR technique. For this, RNA of genotypes DV92 and G3116 was extracted from vegetative and reproductive apexes of seedlings growing under the same experimental conditions used for libraries construction, using a commercial available kit (SV Total RNA Isolation System, Promega).

The cDNA synthesis reaction was performed with 200 µg of total RNA, according to manufacturer protocol (First Strand cDNA Synthesis Kit, Fermentas). Specific primers for quantitative real time PCR were designed through the Integrated DNA Technologies (IDT) webpage tools (http://www.idtdna.com/Scitools/Applications/RealTimePCR/). The unigenes that were selected for the differential expression validation, summarized together with the primer sequence in Supplementary[1] Table 1 (Table S1), followed simultaneously the criteria of having introns in the correspondent DNA sequence and high e-value. A fragment from the *T. monococcum TRANSLATION ELONGATION FACTOR 1*

Table 1. Source materials and main characteristics of the diploid wheat libraries

Library designation	TM011XXX	TM043E1X	TM046E1X
T. monoccocum accession	DV92	DV92	G3116
Tissue	Vegetative apex and shoot	Early reproductive apex and shoot	Vernalized reproductive apexes
Titer of the primary library (pfu/ul)	$1.6\ 10^6$	$2.9\ 10^5$	$9.3\ 10^6$
Titer of the amplified library (pfu/ul)	$1.7\ 10^6$	$6.7\ 10^6$	$1.5\ 10^{10}$
Average EST length (bp)	432 (100–765)	449 (100–772)	589 (100–843)

[1] Supplementary material can be found at www.cerzos-conicet.gob.ar/crc_diaz_garbus_al_2011/

Table 2. Individual wheat libraries characterization

Library	TM011XXX	TM043E1X	TM046E1X
Sequenced ESTs	2990	2633	3357
Non overlapping ESTs	1859	1497	1962
ESTs in contigs	1131	1136	1395
Number of contigs	387	346	443
Redundancy	1.33	1.43	1.39
Unigenes	2246	1843	2405
No. of genes identified	1691	1364	1884

(*TEF1*) gene (Table S1) was taken as an endogenous control for gene expression (Distelfeld and Dubcovsky 2010).

Quantitative real time PCR was performed using a Rotor-Gene 6000 (Corbett Research), in a final reaction volume of 20 µl including 50 pmoles of forward and reverse primers, 1 µl of cDNA dilution and 10 ul of Real Mix (Biodynamics). The amount of cDNA used for each primer combination was determined by performing the amplification with several cDNA dilutions (from 1/10 to 1/300) in the sample in which the expression was expected to be the highest. Amplification protocol consisted of 2 min at 94 °C followed by 40 cycles of a three-step loop: 10 s at 94°C, 15 s at 54°C and 30 s at 72°C. The $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001) was used to normalize and calibrate transcript values relative to *TEF1* expression.

Detection of gene molecular markers: EST-derived simple nucleotide polymorphisms (EST-SNPs)

Differentially expressed genes were analyzed in order to detect SNPs using the SNP Discovery program (Barker et al. 2003), where the confidence for each SNP was calculated considering the redundancy of the polymorphism at a SNP locus, and the co-segregation of the candidate SNP with other SNPs in the alignment. A minimum of 10 was considered for the redundancy parameter and five polymorphisms for co-segregation parameter.

Results

cDNA libraries characterization

Table 1 summarized the plant materials used for TM011XXX, TM043E1X and TM046E1X cDNA libraries construction, as well as their initial characterization. The average insert size of the 8980 ESTs generated is shown in Table 1. Contig assembly was performed individually for each library using the CAP3 software. Near 60% of the sequences from each library remained as non overlapping ESTs (Table 2). The other sequences were grouped in contigs, leading to values of redundancy, i.e., number of sequenced ESTs/number of unigenes that ranged between 1.33 and 1.43 (Table 2).

Sequence comparison with other plant species

The similarities among sequences were compared with those reported for wheat and other two plant species, *A. thaliana* and rice, through the BLASTN algorithm (nucleotide collection nr/nt) (Fig 1). Comparison with rice lead to the highest percentage of sequences with significant identity along the range of cutoff e-values analyzed (Fig. 1). However, it is interesting to remark that, as expected, at highest e-values cutoffs there is an increase in the relative relevance of the identity with previously reported wheat sequences.

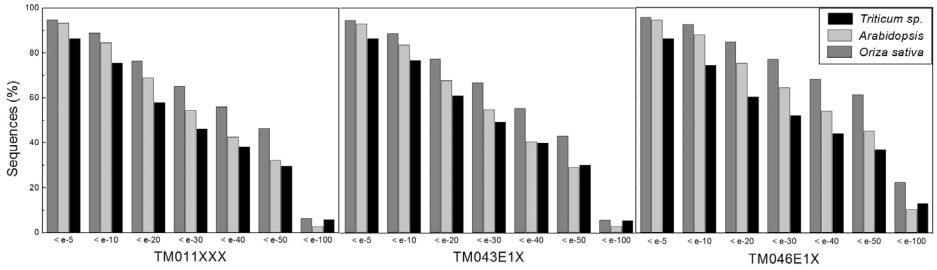

Figure 1. Sequence similarities. Relative percentages of transcripts sequences from *T. monoccocum* similar to sequences in the *Arabidopsis thaliana,* rice and *Triticum* public databases

Then, the sequences were compared to protein databases. Using the algorithm BLASTX through the interface Blast2Go, the percentages of sequences showing identity with database proteins were ~72% and ~56% at the e-values e^{-5} and e^{-25}, respectively, for the three libraries. The comparison with protein sequences from other species revealed a high percentage of unigenes showing identity with rice (~85%) and *Arabidopsis* (~75%). The highest percentage observed with rice is given by the more closely relation wheat-rice compared to wheat-*A. thaliana.* Near 35% showed similarity with *Triticum* spp. protein databases, probably due to the fact that this species has not been explored as intensively as the mentioned model species.

Functional annotation of T. monococcum *unigenes*

Unigenes were annotated based on the existing sequence information for the proteome of others species, in which functions were categorized according to the GO Consortium. Near 25% of the sequences were designated "No gene ontology" with a cutoff e-value of 10^{-3}, whereas 6.5% were classified as "Hypothetical protein", i.e. open reading frames predicted from *Arabidopsis* or rice genomic DNA. As an average, three quarters of the unigenes whose products showed homology with public database sequences significantly matched with the categorized proteins, and putative functions were assigned.

According to molecular function, twelve different GO categories including 41, 40 and 30% of the annotated sequences were identified for the libraries TM011XXX, TM043EIX and TM046E1X, respectively (Fig. 2a). For the three libraries, catalytic activity, nucleic acid binding and ion binding were the subcategories that grouped the most sequences.

Following the criterion of biological process, seventeen different GO categories were identified including 38, 37 and 27% of the annotated sequences for libraries TM011XXX, TM043EIX and TM046E1X, respectively (Fig. 2b). Such categorization resulted in the majority of the sequences falling into the metabolic subcategories of nucleic acid, biopolymers, carbohydrates and proteins.

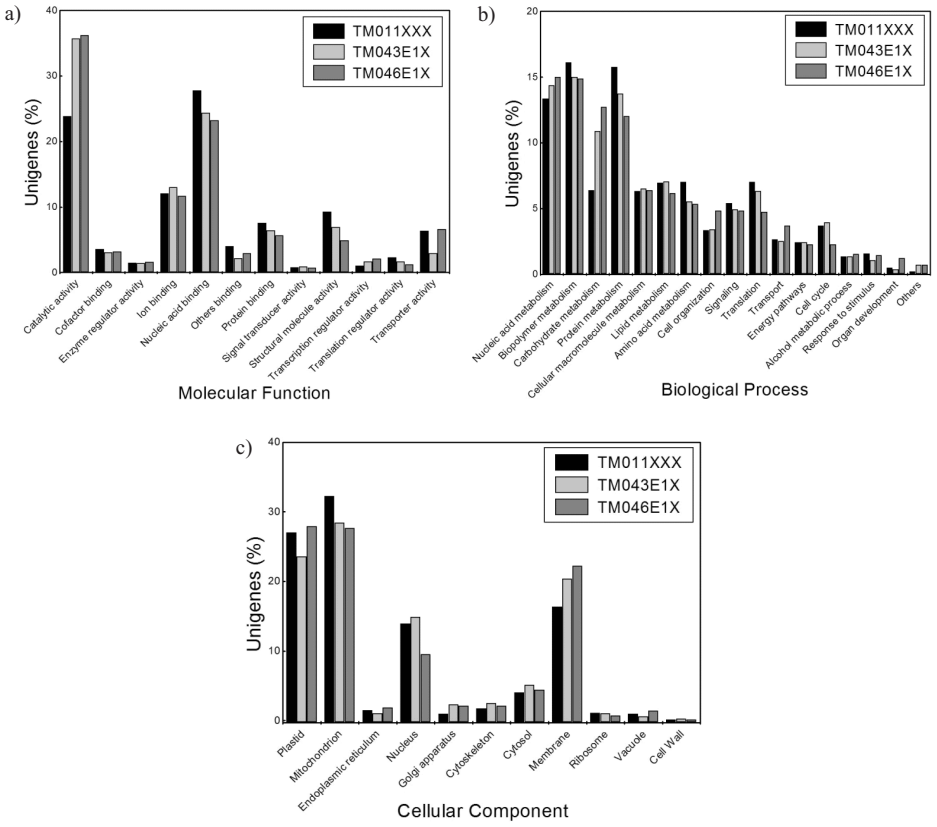

Figure 2. Functional classification of *T. monoccocum* unigenes according to Gene Ontology. a) Molecular function; b) Biological process; c) Cellular component

The categorization according to cellular component resulted in eleven different GO categories that included 44, 46 and 36% of the annotated sequences for libraries TM011XXX, TM043EIX and TM046E1X, respectively (Fig. 2c), with plastid, mitochondrion, membrane and nucleus being the most represented subcategories.

Differentially expressed genes

An *in silico* cDNA library was generated with sequences from the three wheat libraries using the CAP3 software, resulting in 5351 unigenes (1345 contigs and 4005 singletons) in order to identify genes that are differentially expressed among them through the Audic and Claverie (1997) test in pair-wise comparisons ($P < 0.03$). Each library was contrasted against the other two in a pairwise manner and differential expression was considered to occur when the differential criteria was met in both comparison pairs.

The first comparison looked at the sequence contribution in each contig from the library TM011XXX (vegetative apex) versus TM043E1X and TM046E1X (early reproductive apex), in order to identify differences between vegetative and reproductive tissues. Seventeen genes showed differential expression (Table S2 and S3). A group of five genes were over-expressed in the libraries TM043E1X and TM046E1X (Table S2) whereas the remaining sequences were found to be over-expressed in library TM011XXX, suggesting their relationship with the vegetative tissue (Table S3).

The comparison between library TM046E1X (genotype G3116) and libraries TM011XXX and TM043E1X (genotype DV92) lead to the identification of thirty five genes that show differential gene expression between genotypes. A group of twenty four genes were found to be differentially expressed in the winter genotype G3116 (Table S4) whereas the other ones were found to be specific of the spring genotype (Table S5).

Validation of the differential expressed genes

Real time PCR reactions were carried out onto samples from vegetative and reproductive apexes of the two genotypes, growing at same conditions used when the RNA was extracted for the libraries construction. Thus, three cDNA samples were obtained, #1, #2 and #3, representatives of TM011XXX, TM043E1X and TM046 E1X libraries, respectively.

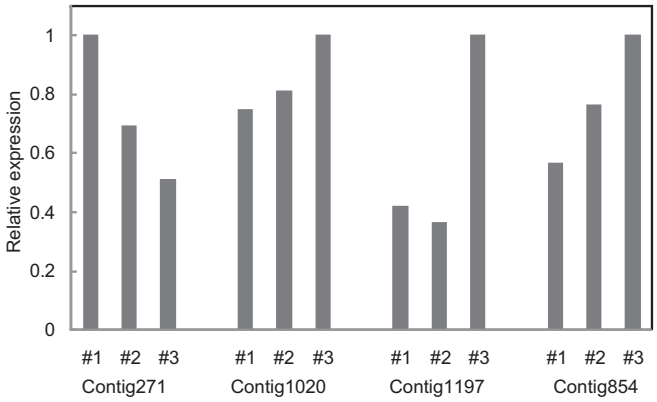

Figure 3. qRT-PCR analysis of four expressed genes in the libraries. The Y-axis shows the diminution in the gene expression with respect to the sample in which the expression was expected to be the highest. Normalization was performed in relation to endogenous control gene expression in each sample. Samples #1, #2 and #3 reproduced the conditions of libraries TM011XXX, TM043E1X and TM046E1X, respectively

From the genes that showed differential expression among the libraries in the *in silico* analysis of the data (Tables S2–5), five were selected to be validated by qRT-PCR technique. Three of them, the transcripts related to sequences grouped in Contig854, Contig1020 and Contig1197 were demonstrated to be differentially expressed between genotypes, being over-expressed in G3116 (sample #3) compared to DV92 (samples #1 and #2) (Fig. 3). The transcripts related to Contig271, showed an increase expression in the vegetative tissue (sample #1) with respect to the reproductive one (samples #2 and #3) (Fig. 3). Results were confirmed by duplicates (data not shown).

Development and detection of gene molecular markers

The occurrence of differences in the sequences contributing to each assembled contig (SNPs, INDELs) was interrogated through the software SNP discovery.

In Contig955, the ribosomal protein S3a (e^{-120}), five SNPs were detected in five of 13 assembled ESTs sequences, over expressed in floral tissue. Although these SNPs do not involve changes in the predicted protein, they lead to three cleavage amplification polymorphic sequences (CAPS), that showed to be genotypic specific. The restriction sites correspond to *Hind*III (genotype DV92) and *Eco*0109I and *Alw*I (genotype G3116).

In Contig773, a metallothioneine type 2 ($2e^{-8}$), a SNP was detected among sequences derived from the genotype DV92, leading to the presence of Valine or Alanine in position 5 of the predicted protein. This polymorphism may be indicating the presence of two allelic forms since there are 5 ESTs of genotype DV92 with the codon GTC (V) and 12 with the codon GCC (A). Moreover, the valine codon could be associated to a restriction site for the enzyme *Bsu*36I, and thus is plausible for the development of an allele specific CAPS. This polymorphism co-segregates with other five SNPs located in the coding region but were not associated with changes in the predicted protein sequence.

Discussion

Three cDNA libraries of *T. monoccocum* (genotypes DV92 and G3116) were constructed and 8989 high quality ESTs were obtained and deposited in public databases. The sequences were assembled into unigenes that were further annotated and classified according to the Gene Ontology (GO) Consortium categories. In agreement with other wheat (Zhang et al. 2004) and non related species (Yu et al. 2006) libraries, categorization according to molecular function, lead to high percentage of sequences in the subcategory catalytic activity. This percentage is lower in library TM011XXX, suggesting that reproductive tissues are metabolically more active than the vegetative ones (Fig. 2a). The contributions to the four main categories of biological process of the unigenes from the three libraries were indistinguishable, except for the library TM011XXX which contribution to carbohydrate metabolism was near the half of the one observed in the other two libraries (Fig. 2b). A reasonable explanation of such observation is that the reproductive tissues accumulate reserves, mainly the complex carbohydrate starch to ensure the mature grain energy supply. The similar contribution to the functional categories for the unigenes from the three libraries, could be taken as another evidence of their equal quality.

Sequence comparison through BLASTN and BLASTX algorithms revealed that the highest identity percentages are observed with rice, followed by *Arabidopsis*. Thus, the ESTs described here constitute novel information about protein-coding regions of diploid wheat genome. The fact that databases contain a less extent of wheat sequences compared to the other two species justifies the lowest similarity observed with *Triticum* spp.

A statistical analysis was performed in order to compute the significance of the gene expression profiles among the 1345 contigs from the virtual cDNA library. Although the redundancies of the libraries indicate that new genes could be discovered by adding ESTs, the equivalent quality of the libraries justifies the validity of the comparison of their expression profiles. Among the genes that were identified as expressed in a tissue-specific manner, five of them have not been previously identified as directly involved in the flowering process, so this study position them to be linked to metabolic pathways that become active in the flowering stage of wheat. Contig1308 (assembled from 11 ESTs), over-expressed in reproductive tissue, showed no identity with reported proteins, suggesting that novel information about tissue-specific proteins could be revealed from the analysis of the predicted proteins of such sequences. The expression of four of these *in silico* identified genes were demonstrated to be differentially expressed by a quantitive RT-PCR study.

It is interesting to remark that in the comparison of expression profiles between genotypes, several regulatory proteins such as two translation initiation factors, a transcription factor and Zinc fingers domains containing proteins were identified. Other differentially expressed genes could be related to cold acclimation given the vernalization conditions required by the winter genotype to flower. Among these genes, the most relevant are the hydrophobic protein LTI6B and alcohol dehydrogenase enzyme (ADH1A), associated with low-temperature stress in maize and rice (Christie et al. 1991; Peters and Frenkel 2004) (Table S5). Finally, there were three sequences that showed no identity with reported proteins, suggesting that their analysis could provide novel information about sequences specifically related to this genotype. Among the genes expressed specifically in the spring genotype, there were some proteins related to gene expression such as three different histones and Elongation factor 1-delta. Several evidences demonstrate the involvement of histone 1 in regulation of gene expression. Some variants of the gene *His 1* (histone 1) have been associated with drought resistance (Wei and O'Connell 1996; Ascenzi and Gantt 1999).

The ESTs libraries reported in the present article provide novel sequence information for the wheat genome A. They constitute a valuable tool for the assessment of active metabolic pathways in vegetative and reproductive tissues in a diploid wheat species.

Acknowledgements

The authors express their gratitude to Dr. J. Dubcovsky for providing his laboratory facilities for libraries construction. This work was funded by the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, PICT 1011), Argentina.

References

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic. Acids Res. **25**:3389–3402.

Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., Davis, A., Dolinski, K., Dwight, S., Eppig, J., Harris, M., Hill, D., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J., Richardson, J., Ringwald, M., Rubin, G., Sherlock, G. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. **25**:25–29.

Ascenzi, R., Gantt, J. 1999. Molecular genetic analysis of the drought-inducible linker histone variant in *Arabidopsis thaliana*. Plant. Mol. Biol. **41**:159–169.

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

Barker, G., Batley, J., O'Sullivan, H., Edwards, K., Edwards, D. 2003. Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. Bioinformatics **19**:421–422.

Bennett, M., Leitch, I. 1995. Nuclear DNA amounts in angiosperms. Ann. Bot. **73**:113–176.

Chantret, N., Salse, J., Sabot, F., Rahman, S., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdille, P., Joudrier, P., Gautier, M.F., Cattolico, L.,Beckert, M., Aubourg, S., Weissenbach, J., Caboche, M., Bernard, M., Leroy, P., Chalhoubb, B. 2005. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). The Plant Cell **17**:1033–1045.

Christie, P., Hahn, M., Walbot, V. 1991. Low-Temperature accumulation of alcohol dehydrogenase-1 mRNA and protein activity in maize and rice seedlings. Plant Physiol. **95**:699–706.

Close, T., Wanamaker, S., Caldo, R., Turner, S., Ashlock, D., Dickerson, J., Wing, R., Muehlbauer, G., Kleinhofs, A., Wise, R. 2004. A new resource for cereal genomics: 22K barley GeneChip comes of age. Plant Physiol. **134:**960–968.

Distelfeld, A., Dubcovsky, J. 2010. Characterization of the *maintened vegetative phase* deletions from diploid wheat and their effect on *VRN2* and *FT* transcript levels. Mol. Genet. Genomics **238**:223–232.

Eujayl, I., Sorrells, M., Baum, M., Wolters, P., Powell, W. 2002. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor. Appl. Genet. **104**:399–407.

Ewing, B., Green, P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8**:186–194.

Gadaleta, A., Giancaspro, A., Giove, S., Zacheo, S., Mangini, G., Simeone, R., Signorile, A., Blanco, A. 2009. Genetic and physical mapping of new EST-derived SSRs on the A and B genome chromosomes of wheat .Theor. Appl. Genet. **118**:1015–1025.

Houde, M., Belcaid, M., Oullet, F., Danyluk, J., Monroy, A., Dryanova, A., Gulick, P., Bergeron, A., Laroche, A., Links, M., MacCarthy, L., Crosby, W., Sarhan, F. 2006. Wheat EST resources for functional genomics of abiotic stress. BMC Genomics **7**:149.

Huang, X., Madan, A. 1999. CAP3: A DNA sequence assembly program. Genome Res. **9**:868–877.

Huang, S., Sirikhachornkit, A., Su, X.J., Faris, J., Gill, B., Haselkorn, R., and Gornicki, P. 2002. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. Proc. Natl. Acad. Sci. USA **99**:8133–8138.

Lazo, G., Chao, S., Hummel, D., Edwards, H., Crossman, C., Lui, N., Matthews, D., Carollo, V., Hane, D., You, F., Butler, G., Miller, R., Close, T., Peng, J., Lapitan, N., Gustafson, J., Qi, L., Echalier, B., Gill, B., Dilbirligi, M., Randhawa, H., Gill, K., Greene, R., Sorrells, M., Akhunov, E., Dvorak, J., Linkiewicz, A., Dubcovsky, J., Hossain, K., Kalavacharla, V., Kianian, S., Mahmoud, A., Miftahudin, Ma, X., Conley, E., Anderson, J., Pathan, M., Nguyen, H., McGuire, P., Qualset, C., Anderson, O. 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.

Livak, K.J., Schmittgen, T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods **25**:402–408.

McCarthy, E.M., Liu, J., Gao, L., McDonald, J.F. 2002. Long terminal repeat retrotransposons of *Oryza sativa*. Genome Biol. **3**:1–11.

Peters, J., Frenkel, C. 2004. Relationship between alcohol dehydrogenase activity and low-temperature in two maize genotypes, Silverado F_1 and *Adh1*$^{-}$*Adh2*$^{-}$ doubly null. Plant Physiol. and Biochem. **42**:841–846.

Wei, T., O'Connell, M. 1996. Structure and characterization of a putative drought-inducible H1 histone gene. Plant Mol. Biol. **30**:255–68.

Yu, J., Sun, Q., La Rota, M., Edwards, H., Tefera, H., Sorrells, M. 2006. Expressed sequence tag análisis in tef (*Eragrostis tef* (Zucc) Trotter). Genome **49**:365–372.

Zhang, D., Choi, D., Wanamaker, S., Fenton, R., Chin, A., Malatrasi, M., Turuspekov, Y., Walia, H., Akhunov, E., Kianian, P., Otto, C., Simons, K., Deal, K., Echenique, V., Stamova, B., Ross, K., Butler, G., Strader, L., Verhey, S., Johnson, R., Altenbach, S., Kothari, K., Tanaka, C., Shah, M., Laudencia-Chingcuanco, D., Han, P., Miller, R., Crossman, C., Chao, S., Lazo, G., Klueva, N., Gustafson, J., Kianian, S., Dubcovsky, J., Walker-Simmons, M., Gill, K., Dvorak, J., Anderson, O., Sorrells, M., McGuire, P., Qualset, C., Nguyen, H., Close, T. 2004. Construction and evaluation of cDNA libraries for large-scale Expressed Sequence Tag sequencing in Wheat (*Triticum aestivum* L.). Genetics **168**:595–608.