

Expressed sequence tags from the Yukon ecotype of *Thellungiella* reveal that gene expression in response to cold, drought and salinity shows little overlap

C.E. Wong¹, Y. Li¹, B.R. Whitty², C. Díaz-Camino¹, S.R. Akhter¹, J.E. Brandle³, G.B. Golding², E.A. Weretilnyk², B.A. Moffatt^{1,*} and M. Griffith¹

¹Department of Biology, University of Waterloo, 200 University Avenue West, N2L 3G1, Waterloo ON, Canada; ²Department of Biology, McMaster University, 1280 Main St. West, L8S 4K1, Hamilton ON, Canada; ³Agriculture and Agri-Food Canada (AAFC) London, 1391 Sandford St, N5V 4T3, London ON, Canada; (*author for correspondence; e-mail moffatt@sciborg.uwaterloo.ca)

Received 8 January 2005; accepted in revised form 22 April 2005

Key words: abiotic stress, acclimation, *Arabidopsis*, Brassicaceae, canola, desiccation, EST, freezing, salt, tolerance

Abstract

Thellungiella salsuginea (also known as *T. halophila*) is a close relative of *Arabidopsis* that is very tolerant of drought, freezing, and salinity and may be an appropriate model to identify the molecular mechanisms underlying abiotic stress tolerance in plants. We produced 6578 ESTs, which represented 3628 unique genes (unigenes), from cDNA libraries of cold-, drought-, and salinity-stressed plants from the Yukon ecotype of *Thellungiella*. Among the unigenes, 94.1% encoded products that were most similar in amino acid sequence to *Arabidopsis* and 1.5% had no match with a member of the family Brassicaceae. Unigenes from the cold library were more similar to *Arabidopsis* sequences than either drought- or salinity-induced sequences, indicating that latter responses may be more divergent between *Thellungiella* and *Arabidopsis*. Analysis of gene ontology using the best matched *Arabidopsis* locus showed that the *Thellungiella* unigenes represented all biological processes and all cellular components, with the highest number of sequences attributed to the chloroplast and mitochondria. Only 140 of the unigenes were found in all three abiotic stress cDNA libraries. Of these common unigenes, 70% have no known function, which demonstrates that *Thellungiella* can be a rich resource of genetic information about environmental responses. Some of the ESTs in this collection have low sequence similarity with those in Genbank suggesting that they may encode functions that may contribute to *Thellungiella*'s high degree of stress tolerance when compared with *Arabidopsis*. Moreover, *Thellungiella* is a closer relative of agriculturally important *Brassica* spp. than *Arabidopsis*, which may prove valuable in transferring information to crop improvement programs.

Introduction

Plants have a remarkable ability to cope with a wide range of abiotic stresses. Nevertheless, major abiotic stresses such as drought, freezing temperatures, and salinity are primarily responsible for the discrepancy that exists between maximal and actual crop yield worldwide. The yields of most

major crop plants are reduced by more than 50% (Bray *et al.*, 2000), representing an economic hardship for farmers.

Numerous efforts have been made to understand and manipulate abiotic stress responses (for review, see Wang *et al.*, 2003). To this end, *Arabidopsis thaliana* has been the model organism of choice due to its small genome, rapid life cycle,

and availability of genetic tools. This is exemplified by identification in *Arabidopsis* of the C-repeat/dehydration-responsive element binding factor (*CBF*) gene family as key transcriptional activators and the associated downstream cold-regulated (*COR*) genes (Thomashow, 1999), as well as the salt overly sensitive (*SOS*) family, components of a signal transduction pathway that regulates ion homeostasis and salt tolerance (Zhu, 2001). However, it is becoming increasingly clear that it is difficult to study the genetics of abiotic stress tolerance using *Arabidopsis* as a model system owing to the fact that *Arabidopsis* has a limited capacity to survive saline, drought or freezing conditions (Bressan *et al.*, 2001).

Recently, *Thellungiella salsuginea* (previously classified as *T. halophila* and hereafter referred to as *Thellungiella*), another member of the family Brassicaceae (Al-Shehbaz *et al.*, 1999), has been identified as a potential model system for studies of abiotic stress tolerance (Bressan *et al.*, 2001; Inan *et al.*, 2004; Volkov *et al.*, 2004; Griffith *et al.*, unpublished). Several ecotypes of *Thellungiella* have been identified, including the Shandong ecotype from maritime habitats in China (Inan *et al.*, 2004) and the Yukon ecotype from saline meadows in subarctic Canada (Cody, 2000; Griffith *et al.*, unpublished). Not only does *Thellungiella* share many features that make *Arabidopsis* an excellent model system, but it is also an 'extremophile' that can tolerate salinity as high as 500 mM NaCl (Inan *et al.*, 2004). In addition, the Yukon ecotype of *Thellungiella* survives freezing to temperatures as low as -19°C (Griffith *et al.*, unpublished). These conditions are far more extreme than those tolerated by *Arabidopsis*.

Expressed sequence tags (ESTs) are obtained by single-pass sequencing of cDNA clones and provide information on the transcribed regions of a genome. Because cDNA libraries are typically generated from specific tissues or developmental stages or other experimental conditions and are randomly selected for sequencing, EST representations provide a dynamic view of genome content and expression. Here we present a set of EST data from an ongoing functional genomics project that is designed to identify the molecular mechanisms underlying abiotic stress tolerance in the Yukon ecotype of *Thellungiella*. In this study, we collected and analyzed 6578 ESTs from cDNA libraries

prepared from leaves cold-, drought-, and salt-stressed *Thellungiella*. Recent EST and microarray studies have focused on early time points, frequently only hours or days after imposing an abiotic stress, in order to identify genes involved in signaling pathways and transcriptional regulation of abiotic stress-induced genes in plants (Fowler and Thomashow, 2002; Kreps *et al.*, 2002; Seki *et al.*, 2002; Wang *et al.*, 2003). In contrast, our libraries included cDNAs from plants exposed for days to weeks to an abiotic stress in order to identify genes that may be mechanistically involved in acclimation or stress resistance in the steady state. Our primary goal was to survey the mRNA populations under the various stresses to determine whether the close relationship between *Arabidopsis* and *Thellungiella* could be exploited to quickly characterize responses to abiotic stress and whether the responses to different abiotic stresses were similar to each other. The genetic information obtained from this remarkable abiotic stress-tolerant crucifer is a key step in the discovery of genes involved in abiotic stress tolerance.

Materials and methods

Plant materials and stress treatments

Plants of the Yukon ecotype of *Thellungiella salsuginea* (Pall.) O.E. Schulz (Al-Shehbaz *et al.*, 1999; Cody, 2000) were grown in controlled environments with an irradiance of $250\ \mu\text{mol photons m}^{-2}\ \text{s}^{-1}$; a 21-h daylength, and a day/night temperature regime of $22/10^{\circ}\text{C}$. When the plants were 4-weeks-old, they were subjected to stress treatments as described below in order to provide material for the construction of abiotic stress-induced cDNA libraries as well as subtracted libraries.

For cold treatment, plants were shifted to a day/night temperature regime of $5/4^{\circ}\text{C}$ and leaves were sampled at 24 h, 1 week and 3 week time points. Freezing stress was conducted by transferring 1-month-old plants, which were acclimated at 5°C for 2 weeks, to a chamber with a day/night temperature regime of $5/-4^{\circ}\text{C}$ for 2 weeks. For the drought treatment, water was withheld from 1-month-old plants until they wilted (about 3 days), and for the drought plus re-watering treatment, drought-treated plants were re-watered

and allowed to regain turgor and recover for 2 days before harvest. Salt-shock treatment was imposed by watering 1-month-old plants with 0.3 M NaCl once daily and plants were sampled at 3 h, 24 h and 3 days. For acclimation of plants to salinity, plants were watered with NaCl solutions at concentrations that increased by 50 mM every 3 days until the final concentration reached 300 mM.

cDNA clones

All plants were harvested at the same time of day (8 h after the lights came on). Total RNA was extracted from only above-ground tissues as described by Danyluk and Sarhan (1990). mRNA populations were isolated by chromatography on oligo(dU) Sephadex columns (Murray *et al.*, 1981; Hondred *et al.*, 1987) and were used for constructing three stress-induced and four subtracted cDNA libraries.

For the 'cold-induced cDNA library', 5 μ g of mRNA was pooled from mRNA obtained from plants subjected to cold for 24 h, 1 week and 3 weeks. For the 'salinity-induced cDNA library', mRNA was pooled from plants subjected to 0.3 M NaCl-shock for 3 h, 24 h, and 3 days, and from plants salt-stressed gradually to 0.3 M NaCl and maintained at this level of salt for 3 days. The 'drought-induced cDNA library' was made using mRNA from plants left unwatered to the point of wilting. These three different groups of mRNA were used to synthesize cDNAs that were directionally cloned using the Superscript Plasmid System for cDNA synthesis and cloning kit (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instructions. Size-fractionated cDNA with fragments >500 bp were pooled and ligated into the vector pSPORT1 (Invitrogen) predigested with *Sal*I and *Not*I. DH10B *Escherichia coli* cells (Invitrogen) were electroporated (25 μ F, 200 Ω , 1.8 kV) with the resulting plasmids. The average titre of the libraries was 2.2×10^7 recombinants per microgram of cDNA.

For 'subtracted libraries', 2 μ g of mRNA were used for cDNA synthesis using the SMART cDNA synthesis kit (Clontech, Palo Alto, CA, USA) according to the manufacturer's protocol. The cDNA population was then enriched in cDNAs related to specific abiotic stresses by using the PCR-Select cDNA subtraction kit (Clontech).

The subtracted libraries were prepared from paired driver versus tester cDNA populations as follows: 4-week-old control versus 3-week cold-acclimated, 3-week cold-acclimated versus 3-week cold-acclimated subjected to freezing and thawing, 4-week-old control versus 3 h salt-shock, and drought versus drought plus rewatering. Each subtracted cDNA population was ligated into the pGEMT-Easy vector (Promega, Madison, WI, USA) according to the manufacturer's instructions. Each plasmid population was transformed into JM109 cells (Promega) by heat shock and plated. The resulting titres averaged 2.0×10^5 recombinants per microgram of cDNA, which were stored as glycerol stocks of individual libraries. All transformants were spread on LB agar plates containing 100 μ g ml⁻¹ ampicillin for direct picking without a library amplification step.

EST sequencing

A total of about 2500, 1800, and 1500 colonies were randomly picked from each of the cold-, drought-, and salinity-induced libraries, respectively and about 400 colonies were picked from the subtracted libraries. Plasmid DNA was prepared from these colonies and sequenced using the facilities at Agriculture and Agri-Food Canada, London, ON, Canada, and the Hospital for Sick Children, Toronto, ON, Canada. Isolated plasmid DNA was sequenced using modified SP6 (98% of the sequences) and T7 (2% of the sequences) primers that flank the cDNA insert. These sequences have been deposited in Genbank (accessions DN772677- DN779205).

Sequence analysis and annotation

Sequencing trace files were processed with the phred basecalling software [version 0.020425.c] (Ewing *et al.*, 1998; Ewing and Green, 1998) to assign base quality values and to identify and trim low quality sequence from the ends of the reads using the '-trim_alt' parameter. Vector and adaptor sequences were detected using a custom Perl script employing swat [P. Green (1993–1996) <http://www.phrap.org>], an efficient implementation of the Smith–Waterman algorithm. Poly-A/T tracts were identified by base composition using another custom Perl script. The trimmed sequences were then screened for remaining vector

sequence contamination by using BLASTN* from the WUBLAST2 package [W. Gish (1996–2004) <http://blast.wustl.edu>] with default parameters against a local version of NCBI's UniVec vector sequence database (<http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html>) and contaminated sequences were rejected.

The output from these screening steps (including raw sequences and processing statistics) was entered into a MySQL database and additional quality screening steps were performed. Trimmed sequences having a length <100 bp or an average phred quality score <15 were rejected. Sequences consisting entirely of poly A sequence or having >30 poly A/Ts at the beginning of the read such that the sequencing quality was affected were also rejected. The phred 'trace peak area ratio' measure was used to identify sequences arising from mixed templates or other sequencing errors that had otherwise passed all screening criteria. Trace files with a trace peak area ratio >0.2 were visually inspected using *trev* from the Staden Package (Staden *et al.*, 2000) and rejected in clear cases of mixed template or degenerate signal.

The TIGR Gene Indices clustering tools (TGICL) (Perteau *et al.*, 2003) software was used to classify the set of screened EST sequences into gene-oriented clusters. The clusters were manually inspected using TIGR's *clview* software (Perteau *et al.*, 2003) to ensure that there were no spurious assemblies. All clusters and singletons resulting from this clustering and assembly were considered to be the best approximation of a minimal gene set for our EST library. We have termed this set as 'unigenes', although we recognize that the method of assembly differs from that of NCBI's UniGene.

Our set of unigenes was BLASTed against local installations of TAIR's *Arabidopsis* protein [ATH1_pep_cm_20040228] and cDNA [ATH1_cdna_cm_20040228] databases (Rhee *et al.*, 2003) using BLASTX and BLASTN (WUBLAST2 build 2.0MP-WashU [16-May-2004] [linux24-i686-ILP32F64 2004-05-16T17:42:20]) from the WUBLAST2 package [W. Gish (1996–2004) <http://blast.wustl.edu>] with default parameters. Each unigene was assigned to an *At* locus based on its best sequence match to the *At* protein database (at $E \leq 10^{-5}$). In the absence of similarity to the protein database, the locus representing the best sequence match from the cDNA database [at $E \leq 10^{-5}$] was assigned.

The unigenes were further compared against a local installation of the GenBank nonredundant protein database (nr) with accompanying taxonomy information (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>) using NCBI's BLASTX 2.2.3 (Altschul *et al.*, 1997). NCBI's BLAST was used for this comparison because taxonomy information is available only with the preformatted nr database. The BLASTX comparisons were performed with all default parameters, except low complexity filtering was disabled. Best matches (by bit score) for each unigene were retained and associated with their taxonomic placement.

Phylogeny

A group of ribulose biphosphate carboxylase large subunit (*rbcL*) gene sequences was collected from *Thellungiella* and related taxa from the NCBI database (<http://www.ncbi.nlm.nih.gov>). The sequences were aligned using ClustalX (Thompson *et al.*, 1994). A Bayesian tree was inferred from 50,000 samples from 5 million generations with gamma rate variation (Huelsenbeck and Ronquist, 2001).

Results and discussion

Generation of ESTs from plants subjected to abiotic stresses

The information provided by ESTs of randomly isolated gene transcripts that have been generated under specific abiotic stress conditions provides an opportunity for gene discovery in addition to identifying the biochemical pathways involved in plant physiological responses. Here, we describe ESTs recovered from cold-, drought- and salinity-induced cDNA libraries prepared from the aerial tissue of the Yukon ecotype of *Thellungiella*. The plants were exposed to each stress for both short and long periods of time in order to obtain ESTs for both regulatory and steady-state processes. Analysis of the PCR products by agarose gel electrophoresis showed a mean insert size of 1.19 kb.

Initially, a total of 8045 cDNAs were sequenced from their 5' ends, with about equal numbers from each library. Poor quality sequences based on an average phred score of less than 15 were excluded;

vector, poly(A) tails, as well as adaptor sequences were removed. A total of 6578 EST sequences passed these criteria and were used in this study. They had an average trimmed length of 535 nt and an average phred score of 50. The 6578 ESTs were assembled using TGICL into 871 contigs and 2757 singlets, which resulted in a final annotation of 3628 unigenes. The clustering of the EST sequences by TGICL was expected to be gene-oriented with high stringency for small variations in transcript sequence (Perlea *et al.*, 2003), allowing for the discrimination of highly similar but distinct transcripts. Post-assembly analysis of our set of unigenes revealed at least one instance of a unigene containing members that were probable splice variants.

The redundancy level of EST collection was 55%, which means that continued sequencing of cDNAs selected at random from our libraries still has considerable potential to uncover novel sequences. To assess how well the 3628 genes may represent the genome, we examined the chromosomal distribution of the *Thellungiella* ESTs in the *Arabidopsis* genome by using the most similar At locus. The distribution of expressed *Thellungiella* genes shown in Table 1 was nearly the same as that for all of the predicted genes in *Arabidopsis* genome. These results show that the *Thellungiella* EST collection has a balanced representation of the genome.

To improve the likelihood of recovering rare cDNAs, three subtracted libraries were generated using plants from a single time point after applying a stress and subtracting sequences from plants grown under more optimal conditions (see Materials and methods). In order to enrich cDNAs that may be related directly to freezing tolerance, a fourth library was generated from the mRNA of cold-acclimated plants exposed to freezing and subtracting sequences from the mRNA of

cold-acclimated plants. These sequences were shorter on average (460 bp) reflecting the frequency of *RsaI* digestion of the cDNA. We sequenced 384 clones from the subtracted libraries. On average, 56% of the sequences in each of the subtracted library were redundant and there was 21% redundancy between the stress-induced libraries and the subtracted libraries. The subtracted libraries added a total of 133 unique genes to the EST collection. Two of these genes had no homologues in the public sequence databases.

Comparison of Thellungiella ESTs with Arabidopsis sequences

A comparison of the unigenes against the TAIR cDNA database (Figure 1) revealed that most of the *Thellungiella* sequences were highly similar to *Arabidopsis* sequences. Of those that were above 71% nt similarity, the main sequence discrepancies were localized in the UTRs, whereas the coding regions were almost identical. The unigenes with less than 70% nt similarity to sequences from *Arabidopsis* or other plant species may potentially be novel sequences or splice variants of similar genes and are, therefore, of obvious interest because they may be related to the extreme stress tolerance of this species. However, it is equally possible that sequences of high sequence similarity may have acquired novel functions due to subtle amino acid changes.

Another way to compare expressed sequences from *Thellungiella* with *Arabidopsis* is to determine the best BLAST match for each unigene. As shown in Table 2, 94% of the *Thellungiella* unigenes were most similar to *Arabidopsis* and a total of 98.5% were most similar to a sequence from a species within the Brassicaceae. However, the remaining

Table 1. Distribution of 3628 unique genes from *Thellungiella* across the five *Arabidopsis* chromosomes.

Source	Chromosome				
	I	II	III	IV	V
<i>Thellungiella</i> ESTs (%)	25.6	16.0	19.8	15.3	23.2
All predicted <i>Arabidopsis</i> genes (%)	25.6	16.2	19.9	15.3	23.0

Each *Thellungiella* unigene was assigned to an At locus based on its best match with the At protein database or cDNA database. The distribution of *Thellungiella* unigenes was then compared with the chromosomal distribution of all 27,117 predicted genes in *Arabidopsis* database as of April 2003 that was calculated by Hu *et al.* (2003).

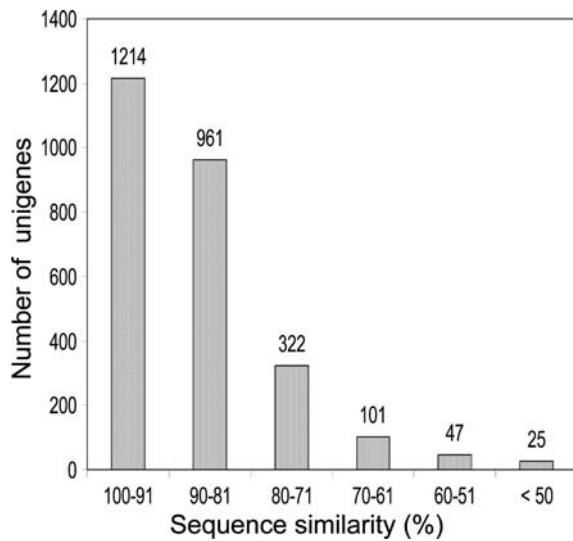


Figure 1. Comparison of nucleotide sequence similarity between *Thellungiella* unigenes and *Arabidopsis* transcripts. Sequences of all unigenes obtained from cDNA libraries from salt-, cold- and drought-stressed *Thellungiella* were compared with the TAIR cDNA database using the BLASTN algorithm. Only unigenes with a length that covered 50% or more of the corresponding *Arabidopsis* sequence are displayed. The total number of unigenes in each group is shown at the top of each bar.

1.5% of the unigenes were most similar to other plants, fungi or bacteria. Because the *Arabidopsis* genome is fully sequenced, the finding that some *Thellungiella* unigenes were not found in *Arabidopsis* suggests that *Thellungiella* contains genes

Table 2. All *Thellungiella* unigene sequences (3628) were translated and compared against the GenBank nonredundant protein sequence database using BLASTX (with default parameters except filters were disabled).

Organism	Total unigenes (%)
Brassicaceae family	98.53
<i>Arabidopsis thaliana</i>	94.14
<i>Brassica</i> sp.	3.61
<i>Thellungiella</i> sp.	0.37
<i>Sinapis</i> sp.	0.11
<i>Thlaspi</i> sp.	0.08
<i>Arabis</i> sp.	0.08
Other genera	0.14
Other Eudicotyledons	0.80
Other Liliopsida	0.48
Fungal and Bacterial	0.17

The sequence with the best match (highest bit score) for each unigene was selected and classified by organism. Final results are expressed as a percentage of total unigenes.

encoding products with functions distinct to this crucifer.

The virtual translation products of the *Thellungiella* unigenes specific to the three stresses were analyzed for similarity with the TAIR *Arabidopsis* protein database (Figure 2). In each case, the amino acid sequence similarities ranged from nearly identical (BLASTX Expect value less than 10^{-100}) to those that were only slightly similar (approximately 10^{-4}). Although the range was similar in the three libraries, the distribution of the Expect values differed. For example, the median Expect value obtained in the cold cDNA library was 1.20×10^{-79} , which was significantly different ($p < 0.0001$) from the corresponding median Expect values of 1.28×10^{-64} and 9.30×10^{-66} for the drought and salinity cDNA libraries, respectively (Figure 2). There was no significant difference in median Expect values between the drought and salinity cDNA libraries. These results indicated an overall higher conservation of the cDNA sequences associated with cold stress between *Arabidopsis* and *Thellungiella*, versus those recovered in the drought- and salt-cDNA libraries. While this suggests that the mechanism of cold tolerance of these two plants may be more similar than the mechanisms for drought or salinity, this is likely an oversimplification because modest amino acid differences may result in substantial changes in functionality. Moreover the level of expression of specific sequences may also contribute to tolerance, which is not reflected in this analysis.

Classification of *Thellungiella* ESTs by biological process and cellular component

As shown in Figure 3, all unigenes obtained from cold, drought and salinity libraries were classified according to terms developed by the Gene Ontology Consortium (Berardini *et al.*, 2004) by using the TAIR database. However, many of the transcripts identified in *Arabidopsis* (30%) have not yet been assigned to a specific functional category due to a lack of information about their gene products. This limited our ability to assign a role to 55.2% of the *Thellungiella* sequences (i.e., the sequences were assigned to the following GO categories: biological processes unknown, other cellular, other metabolic, other physiological and other biological processes). We successfully

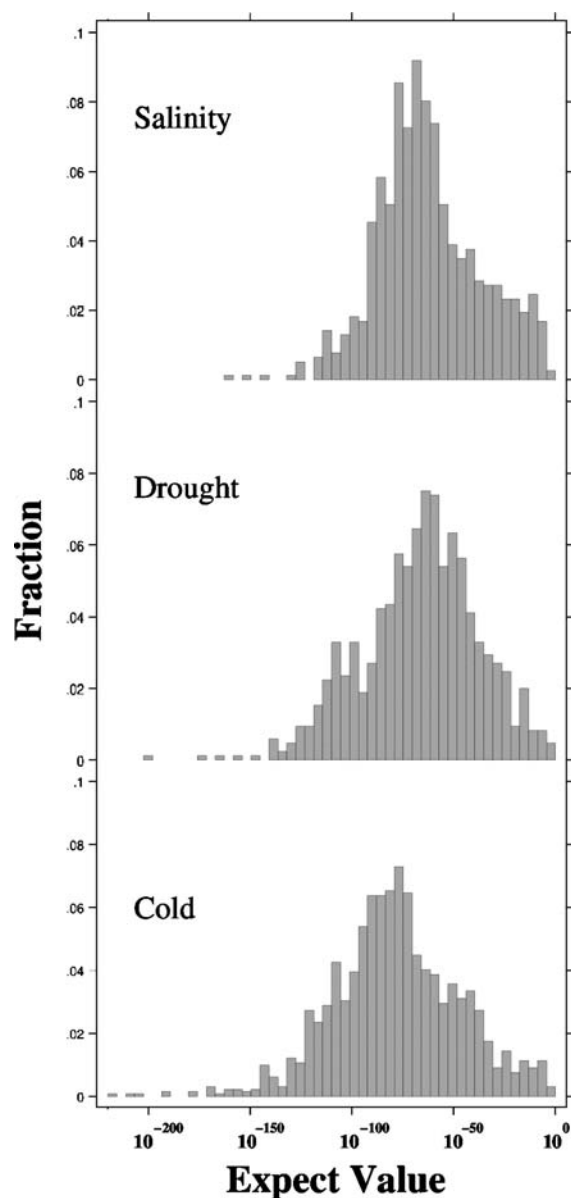


Figure 2. Frequency of Expect values obtained using the BLASTX algorithm to compare translated *Thellungiella* unigenes obtained from all three abiotic stress cDNA libraries with amino acid sequences of *Arabidopsis* proteins from TAIR protein database.

classified the remaining unigenes to roles in transcription and signal transduction, protein, DNA or RNA metabolism, energy pathways, development or response to stress (Figure 3A). Of the categories with defined functions, the largest number of unigenes was involved with protein metabolism (11.5%) and transport (9.3%). The

former consists of proteins involved in moving, modifying, storing and degrading proteins. Approximately half of the genes in this category are involved in proteolysis or function as protease inhibitors. Proteases, including cysteine proteinases, CIP protease and ubiquitin-conjugating enzyme, are thought to be required for protein turnover and recycling of amino acids.

There are several potential strategies that a plant could use to optimize function under abiotic stress. One is to reprogram activities underway in existing leaves by expressing different isozymes to regain normal function. In this scenario, the proteases may selectively degrade proteins that are inactivated by the stress or perform suboptimally in response to stress. A second strategy is to discard fully developed leaves and produce new leaves with an improved capacity to function under different environmental conditions. In this case, increases in protein metabolism could be related to nutrient recovery and transport to developing leaves. *Thellungiella* may employ both strategies. For example, we have observed that *Thellungiella* plants transferred to cold temperature produce new leaves to replace the mature leaves that were developed at warm temperature, and that multiple copies of proteases are found in the cold-induced cDNA library (data not shown). A similar mechanism has been reported in *Arabidopsis*, where salinity induces programmed cell death in primary roots and the plants produce secondary roots that are better able to cope with the stress (Huh *et al.*, 2002). Drought stress has also been shown to accelerate leaf senescence, which is characterized by many subcellular changes, including an increase in protease activities (Thomas and Stoddart, 1980).

Aquaporins and sugar transporters are among the gene products that are found in the transporter category. Collectively, these products transport water and sugars through plasma membranes and the tonoplast and can help cells adjust to changes in osmotic pressure that cells experience under stress conditions. Also in the same category is a potassium transporter that may be involved in the acquisition of K⁺, which is an essential cofactor for many enzymes (Hasegawa *et al.*, 2000); or it may serve to balance K⁺ and Na⁺ uptake, which can be an important determinant of salinity tolerance (Bray *et al.*, 2000). In fact, Volkov *et al.* (2003) have shown that *Thellungiella* is more

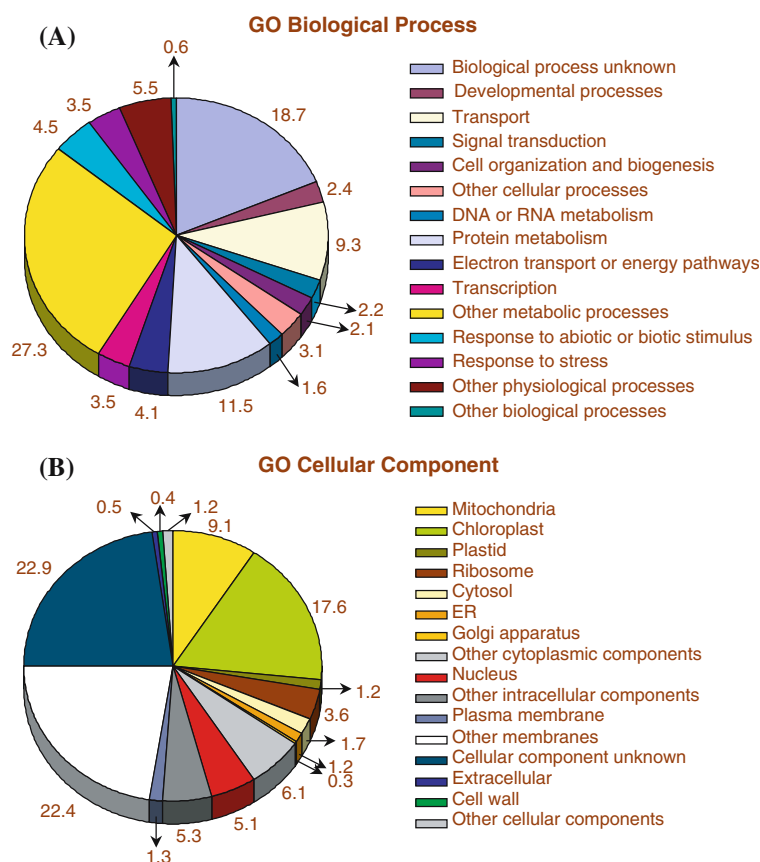


Figure 3. Categorization of *Thellungiella* unigenes by Gene Ontology. *Thellungiella* unigenes were assigned an *At* locus and then categorized using TAIR automatic system. Note that a gene may be assigned to more than one biological process in the GO classification system.

selective for K^+ and accumulates less Na^+ after an application of $NaCl$ than *Arabidopsis*.

When the GO cellular annotation of the unigenes was examined, we found that products associated with all subcellular locations were represented, with the highest number of sequences being attributed to chloroplasts and mitochondria (Figure 3B). Again, analysis of the data was limited by the fact that 57.9% of the sequences could not be assigned to a cellular location (i.e., the sequences were assigned to other cytoplasmic components, other intracellular components, other membranes, cellular component unknown, and other cellular components). We examined the cellular component data by individual abiotic stress but found that the transcriptional response was similar for cold, drought and salinity: i.e., all cellular compartments were represented and most of the predicted gene products were localized in

chloroplasts and mitochondria as well (data not shown).

Relationships among ESTs in response to cold, drought and high salinity

It has recently been proposed that plants use common signaling pathways and components to respond to different abiotic stresses (Pastori and Foyer, 2002; Chinnusamy *et al.*, 2004; Kacperska, 2004). We used a Venn diagram to illustrate the relationships to cold, drought, and high salinity among the 3628 unigenes obtained in this study (Figure 4A). We identified 200 (5.5%) common transcripts between cold and drought libraries, 136 (3.7%) between salinity and cold, and 93 (2.6%) between drought and salinity. Surprisingly, only 140 unigenes (3.9%) were present in all three libraries. We had expected a much

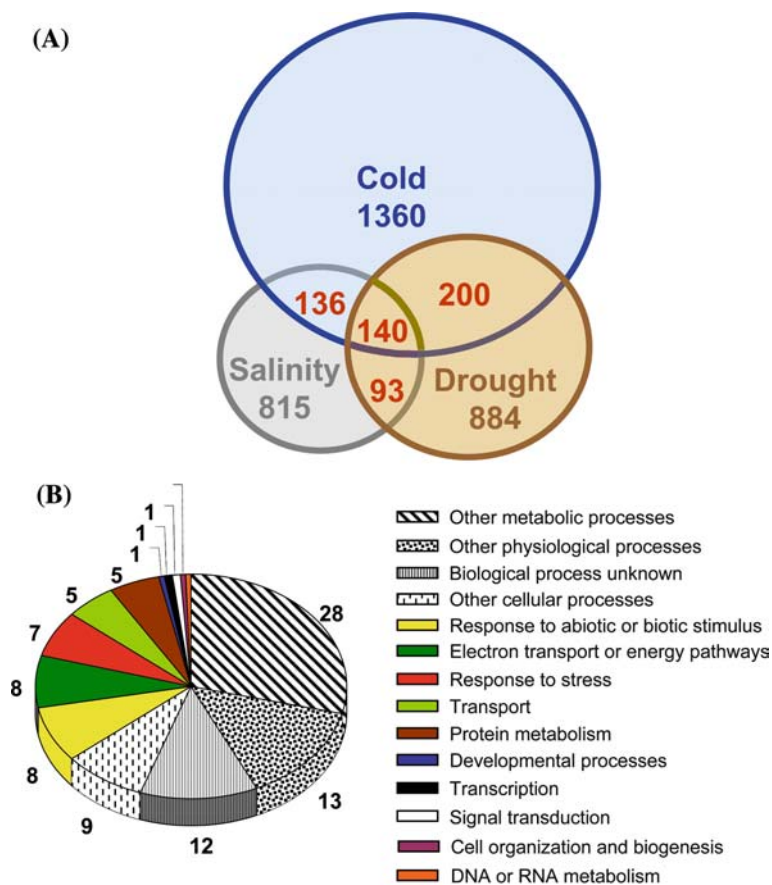


Figure 4. Functions of *Thellungiella* unigenes present in cDNA libraries from all three abiotic stresses. (A) Venn diagram depicting the number of unigenes recovered from cDNA libraries of one or more abiotic stresses. (B) Categorization by biological process of the 140 unigenes common to all three abiotic stresses, annotated using the TAIR automatic system. Categories with unknown functions are all shown in black and white.

greater degree of overlap between the three cDNA libraries because equivalent numbers of clones were chosen at random from each library and each set could have contained a high number of constitutively expressed genes. Moreover, all three stresses could have caused similar physiological problems such as dehydration and photoinhibition, which could have elicited similar patterns of gene expression from the plants. The lower than expected degree of overlap could be attributable, in part, to the fact that a mixture of mRNAs from both early and late time points was used for the construction of the cold and salinity libraries but not for drought library. However, there was a greater overlap of genes found in cold and drought libraries (200) when compared with genes found in both the drought and salinity libraries (93). This implies that similar protective

mechanisms may be triggered under cold and drought stress.

A putative functional assignment, inferred from significant similarity scores from BLASTN and BLASTX reports, was given to each EST in the group that was common to all three abiotic stresses (Figure 4B). This analysis revealed that 70% of the genes that were expressed in all three libraries are of unknown function. Only 8% of the genes have already been classified in *Arabidopsis* as responsive to either biotic or abiotic stimuli. These include genes encoding COR47, dehydrin (ERD14), plant defensin protein (PDF1.2a), cysteine proteinase RD19a, pathogenesis-related thaumatin family protein, chitinase, and, intriguingly, proteins regulating the intracellular level of H₂O₂: catalase 3, peroxidase, glutathione peroxidase, and superoxide dismutase. The presence of

multiple copies of these antioxidative proteins across the different libraries infers that there is a heightened level of reactive oxygen species in plants under abiotic stress and these proteins are needed to maintain the redox homeostasis. In fact, transgenic plants with suppressed H₂O₂-scavenging enzymes are hypersensitive to abiotic stress conditions and pathogen attack (Orvar and Ellis, 1997; Willekens *et al.*, 1997; Mittler *et al.*, 1999). In addition, overexpression of H₂O₂-scavenging enzymes was found to increase the tolerance of plants to abiotic stress conditions (Roxas *et al.*, 2000; Yan *et al.*, 2003).

Interestingly, 1% of the genes that fall in the category of transcription or signal transduction elements are common to all libraries (Figure 2B). A *myb* family transcription factor that bears similarity to *MybSt1* is one of the genes that falls in the transcription category. *MybSt1* was first isolated from potato and reported to be a novel class of myb factor that can act as a transcriptional activator (Baranowskij *et al.*, 1994). The presence of this myb factor in all three stress-induced libraries implies that it may play a role in regulating abiotic stress responses.

Comparison of the expression profiles of genes represented by stress-induced ESTs in two ecotypes of Thellungiella

A recent study reported the generation, sequencing and analysis of a NaCl-treated cDNA library from the Shandong ecotype of *T. salsuginea* (also known as *T. halophila*) (Wang *et al.*, 2004). The Shandong ecotype was reported to be cold-, salt- and oxidative stress-tolerant, but not drought-tolerant (Inan *et al.*, 2004). We compared the expression profile of our EST collection with the Shandong ecotype by tabulating ESTs that were present at a frequency of three or more copies (Table 2). The number of ESTs for a given gene should reflect the mRNA level of gene expression and has been referred to as digital northern analysis (Audic and Claverie, 1997). Although there were fewer unigenes available for the Shandong ecotype (813), we identified a homologous sequence in our EST collection for most of the putative genes that were reported in the Shandong library and we did find some differences in transcript abundance that could be interpreted as differences between the ecotypes in their physiological responses to stress.

One observation was that multiple copies of the gene coding for mannitol dehydrogenase were present in our EST collection but the gene was not found among the Shandong ESTs. Mannitol has been reported to function as a 'compatible solute' in plants that exhibit increased tolerance to salinity and drought stress (reviewed by Stoop *et al.*, 1996). However, mannitol dehydrogenase is involved in the catabolism of mannitol, which would be inconsistent with a role in osmoprotection where mannitol accumulation would be expected to occur. An increased capacity for mannitol turnover raises the possibility that mannitol is an alternative source of fixed carbon under stress conditions in *Thellungiella*. The functional significance of possible changes in mannitol metabolism requires further investigation given its role as an osmoprotectant in some plants.

Thellungiella leaves are described as glaucous and contain about 13 times more epicuticular wax than *Arabidopsis thaliana* (Teusink *et al.*, 2002). The EST that appeared most frequently in response to salinity stress in the collection from the Shandong ecotype was *LTP4*, which encodes a secreted lipid transfer protein thought to be involved in the production of epicuticular waxes (Table 2, Inan *et al.*, 2004). In contrast, *LTP4* was not present in the salinity-induced library of the Yukon ecotype but did appear in the drought library. This result may indicate that *LTP4* is regulated differently in the Yukon ecotype.

Use of Thellungiella salsuginea as an Arabidopsis-related model system for abiotic stress responses

Due to their common use in plant evolutionary studies, gene sequences from *rbcL* were used to illustrate the phylogenetic history of *Thellungiella*. The resulting phylogeny, shown in Figure 5, was well supported and was rooted on the branch leading to *Chlamydomonas*. These results confirmed an earlier phylogeny of *Thellungiella* that was based on an analysis of sequences from nuclear genes encoding arginine decarboxylase (Galloway *et al.*, 1998). Both studies showed that *Thellungiella* is the member of the Brassicaceae that is located nearest to the branch point between the clade containing *Arabidopsis* and the clade containing *Brassica* spp. Therefore, *Thellungiella* may be an important intermediary between

Table 3. Comparison of abiotic-stress related ESTs between the Yukon and Shandong ecotypes of *Thellungiella*.

	<i>A. thaliana</i> locus	Yukon ecotype (<i>n</i>)			Shandong ecotype (<i>n</i>)
		Cold	Drought	Salinity	Salinity
<i>Antioxidant enzymes</i>					
Metallothionein-like protein	At3g09390	4	6	1	29
Catalase (SEN2)	At1g20620	14	4	16	16
Peroxidase 42	At4g21960	8	12	8	2
Glutathione S-transferase	At2g30860	4	2	0	4
Aldehyde dehydrogenase	At1g54100	0	6	2	0
HesB-like domain-containing protein	At1g10500	1	4	0	0
Superoxide dismutase [Cu-Zn]	At1g08830	0	4	1	0
Glutathione S-transferase ζ-1	At2g02390	0	3	0	0
<i>Development/cell differentiation</i>					
Thioglucoside glucohydrolase	At5g25980	6	6	7	1
Spermidine synthase 1 (SPDSYN1)	At1g23820	4	0	0	0
Adenosylmethionine decarboxylase	At3g02470	5	2	1	0
Hydrolase, alpha/beta fold protein family	At4g37470	4	0	1	0
Hydroxyproline-rich glycoprotein	At3g25690	0	3	0	0
LEA	At3g17520	0	3	0	0
Lipid transfer protein 4	At5g59310	0	14	0	26
<i>Metabolism</i>					
Glyceraldehyde-3-phosphate dehydrogenase-related	At3g04120	13	11	5	7
Glucose-1-phosphate adenyllyltransferase	At5g48300	3	7	2	1
Ferritin 1 (FER1)	At5g01600	1	7	1	0
<i>Osmolyte biosynthesis</i>					
Delta-1-pyrroline-5-carboxylase synthase A	At2g39800	1	4	0	3
<i>Plant hormonal regulation</i>					
Chalcone synthase	At5g13930	6	6	7	0
Acetyl-CoA C-acyltransferase	At2g33150	1	2	4	6
Jacaline lectin family protein	At3g16470	0	0	3	2
2-oxoglutarate-dependent dioxygenase	At1g04350	7	2	1	3
Phenylalanine ammonia-lyase 1	At2g37040	3	0	0	0
1-aminocyclopropane-1-carboxylate oxidase	At1g62380	1	4	1	3
<i>Signaling components</i>					
Polyubiquitin (SEN3)	At4g05320	13	11	5	6
Protein kinase, putative (MRK1)	At3g63260	0	0	3	0
Ras-related GTP-binding nuclear protein	At5g20010	0	0	3	4
Elongation factor 1-α	At1g07920	17	4	8	1
Peptidyl-prolyl cis-trans isomerase	At3g62030	4	0	0	0
Proline-rich family protein	At4g19200	3	1	0	3
<i>Stress proteins</i>					
Dehydrin (RAB18)	At5g66400	4	21	0	1
Dehydrin (COR47)	At1g20440	9	5	3	1
Chitinase	At2g43570	2	3	3	3
Osmotin-like protein	At4g11650	7	1	2	1
Early light inducible protein	At3g22840	5	1	1	0
Luminal binding protein 2	At5g42020	1	0	0	3
Mannitol dehydrogenase	At4g37990	0	4	2	0
Dehydration-responsive protein	At5g25610	0	3	0	0
<i>Transcriptional regulators</i>					
Heat shock cognate 70 kDa protein	At5g02500	8	6	3	1
Zinc finger (C3HC4-type)	At3g18773	0	3	0	0

Table 3. Continued

	<i>A. thaliana</i> locus	Yukon ecotype (<i>n</i>)			Shandong ecotype (<i>n</i>)
		Cold	Drought	Salinity	Salinity
<i>Transmembrane transport</i>					
Plasma membrane proton ATPase	At2g18960	3	4	2	9
Bile acid:sodium symporter family	At2g26900	0	0	3	0
ATPase 1, plasma membrane-type	At2g18960	3	4	2	4
Water channel protein-like	At4g00430	1	2	4	1

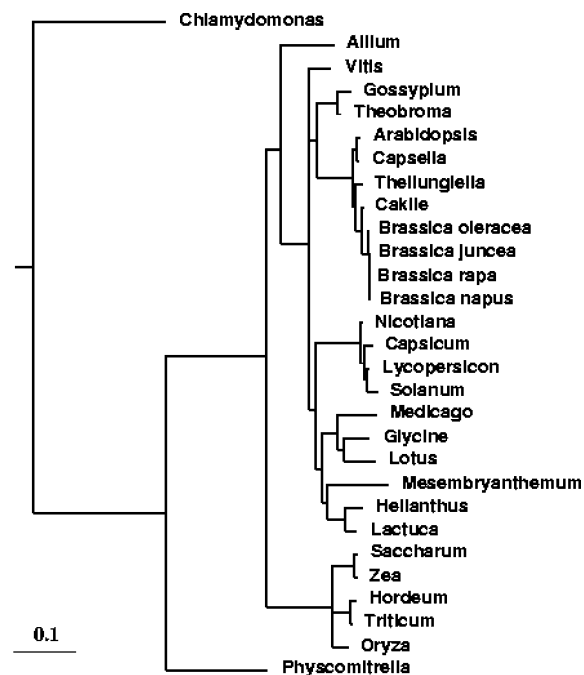


Figure 5. A Bayesian tree of RBCL protein sequences of the indicated genera inferred from 50,000 samples from 5 million generations. All visually discernable branch lengths have posterior support values greater than 0.85. The only exception is the common ancestral branch leading to *Glycine* and *Lotus*, which has a posterior support of 0.67. The scale bar represents 0.1 amino acid substitutions per site. The tree has been rooted on the branch leading to *Chlamydomonas*.

the genetic model plant *Arabidopsis* and the agriculturally and commercially important *Brassica* spp. that include both oilseed crops (rapeseed, canola) and vegetable crops (broccoli, cabbage, cauliflower, kale).

Conclusions

We believe that there are many compelling reasons to choose *Thellungiella* as a model plant to use for

further genetic and mechanistic studies of abiotic stress resistance in plants. First of all, *Thellungiella* grows in areas that require a much higher level of tolerance to abiotic stresses such as cold, freezing, drought, and salinity, than *Arabidopsis* and *Brassica* spp. Secondly, *Thellungiella* is intermediate between *Arabidopsis*, whose genome is fully sequenced, and *Brassica* spp that are commercially important. Third, *Thellungiella* has many of the characteristics of a model plant species in that it is small and self-fertile with a rapid life cycle. Moreover, unlike *Arabidopsis*, *Thellungiella* is perennial and can be maintained for longer periods, which means that it can provide more tissue for examination and many more seeds over multiple flowering cycles compared with *Arabidopsis*. Fourth, our analyses of stress-induced ESTs from *Thellungiella* show that abiotic stress tolerance could arise from both the presence of genes and changes in gene expression in response to environmental factors that are distinct from *Arabidopsis*. Fifth, there are several ecotypes of *Thellungiella* that could be exploited as a source of genetic variability for both mapping and physiological studies. Finally, our results also show that many of the genes that are expressed in response to abiotic stress *Thellungiella* encode proteins of unknown function, thus demonstrating that this plant is a rich resource for future genetic studies. The development of genetic tools, such as our EST collection from *Thellungiella* plants subjected to abiotic stress, is an important step toward realizing this potential.

Acknowledgements

Our work was supported by grants from the Natural Science and Engineering Research Council of Canada, Agriculture and Agri-Food

Canada, the Canola Council of Canada, the Food System Biotechnology Centre at the University of Guelph, and Performance Plants Inc. We thank Bruce Bennett from Yukon Wildlife, Whitehorse, YT, and David Guevara, McMaster University, and Lynn Hoyles, University of Waterloo, for assistance in obtaining seeds and growing plants.

References

- Al-Shehbaz, I.A., O'Kane, S.L. Jr. and Price, R.A. 1999. Generic placement of species excluded from *Arabidopsis* (Brassicaceae) *Novon* 9: 296–307.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs *Nucleic Acids Res.* 25: 3389–3402.
- Audic, S. and Claverie, J.M. 1997. The significance of digital gene expression profiles *Genome Res.* 7: 986–995.
- Baranowskij, N., Frohberg, C., Prat, S. and Willmitzer, L. 1994. A novel DNA binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator *EMBO J.* 13: 5383–5392.
- Berardini, T.Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L.A., Yoon, J., Doyle, A., Lander, G., Moseyko, N., Yoo, D., Xu, I., Zoeckler, B., Montoya, M., Miller, N., Weems, D. and Rhee, S.Y. 2004. Functional annotation of the *Arabidopsis* genome using controlled vocabularies *Plant Physiol.* 135: 745–755.
- Bray, E.A., Bailey-Serres, J. and Weretilnyk, E. 2000. Responses to abiotic stress. In: W. Gruissem, B. Buchanan and R. Jones (Eds.), *Biochemistry and Molecular Biology of Plants*, American Society of Plant Physiologists, Rockville, MD, pp. 1158–1249.
- Bressan, R.A., Zhang, C., Zhang, H., Hasegawa, P., Bohnert, H. and Zhu, J.K. 2001. Learning from the *Arabidopsis* experience. The next gene search paradigm *Plant Physiol.* 127: 1354–1360.
- Chinnusamy, V., Schumaker, K. and Zhu, J.K. 2004. Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants *J. Exp. Bot.* 55: 225–236.
- Cody, W.J. 2000. *Flora of the Yukon Territory*, 2nd ed., NRC Research Press, Ottawa, Canada, pp. 669.
- Danyluk, J. and Sarhan, F. 1990. Differential messenger-RNA transcription during the induction of freezing tolerance in spring and winter wheat *Plant Cell Physiol.* 31: 609–619.
- Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities *Genome Res.* 8: 186–194.
- Ewing, B., Hillier, L., Wendl, M.C. and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment *Genome Res.* 8: 175–185.
- Fowler, S. and Thomashow, M.F. 2002. *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway *Plant Cell* 14: 1675–1690.
- Galloway, G.L., Malmberg, R.L. and Price, R.A. 1998. Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae *Mol. Biol. Evol.* 15: 1312–1320.
- Hasegawa, P.M., Bressan, R.A., Zhu, J.K. and Bohnert, H.J. 2000. Plant cellular and molecular responses to high salinity *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51: 463–499.
- Hondred, D., Wadle, D-W., Titus, D.E. and Becker, W.M. 1987. Light-stimulated accumulation of the proximal enzymes hydroxypyruvate reductase and serine:glyoxylate aminotransferase and their translatable mRNAs in cotyledons of cucumber seedlings *Plant Mol. Biol.* 9: 259–275.
- Hu, W., Wang, Y., Bowers, C. and Ma, H. 2003. Isolation, sequence analysis, and expression studies of florally expressed cDNAs in *Arabidopsis* *Plant Mol. Biol.* 53: 545–563.
- Huelsenbeck, J.P. and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees *Bioinformatics* 17: 754–755.
- Huh, G.H., Damez, B., Matsumoto, T.K., Reddy, M.P., Rus, A.M., Ibeas, J.I., Narasimhan, M.L., Bressan, R.A. and Hasegawa, P.M. 2002. Salt causes ion disequilibrium-induced programmed cell death in yeast and plants *Plant J.* 29: 649–659.
- Inan, G., Zhang, Q., Li, P.H., Wang, Z.L., Cao, Z.Y., Zhang, H., Zhang, C.Q., Quist, T.M., Goodwin, S.M., Zhu, J.H., Shi, H.H., Damsz, B., Charbaji, T., Gong, Q.Q., Ma, S.S., Fredricksen, M., Galbraith, D.W., Jenks, M.A., Rhodes, D., Hasegawa, P.M., Bohnert, H.J., Joly, R.J., Bressan, R.A. and Zhu, J.K. 2004. Salt cress. A halophyte and cryophyte *Arabidopsis* relative model system and its applicability to molecular genetic analyses of growth and development of extremophiles *Plant Physiol.* 135(3): 1718–1737.
- Kacperska, A. 2004. Sensor types in signal transduction pathways in plant cells responding to abiotic stressors: do they depend on stress intensity? *Physiol. Plant.* 122: 159–168.
- Kreps, J.A., Wu, Y., Chang, H.S., Zhu, T., Wang, X. and Harper, J. 2002. Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress *Plant Physiol.* 130(4): 2129–2141.
- Mittler, R., Hallak-Herr, E., Orvar, B.L., Camp, W.Van, Willekens, H., Inze, D. and Ellis, B. 1999. Transgenic tobacco plants with reduced capability to detoxify reactive oxygen intermediates are hyper-responsive to pathogen infection *Proc. Natl. Acad. Sci. USA* 96: 14165–14170.
- Murray, M.G., Peters, D.L. and Thompson, W.F. 1981. Ancient repeated sequences in the pea and mung bean genomes and implications for genome evolution *J. Mol. Evol.* 17: 31–42.
- Orvar, B.L. and Ellis, B.E. 1997. Transgenic tobacco plants expressing antisense RNA for cytosolic ascorbate peroxidase show increased susceptibility to ozone injury *Plant J.* 11: 1297–1305.
- Pastori, G.M. and Foyer, C.H. 2002. Common components, networks, and pathways of cross-tolerance to stress. The central role of “redox” and abscisic acid-mediated controls *Plant Physiol.* 129: 460–468.
- Perteau, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J. and Quackenbush, J. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets *Bioinformatics* 19: 651–652.
- Rhee, S.Y., Beavis, W., Berardini, T.Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M., Miller, N., Mueller, L.A., Mundodi, S., Reiser, L., Tacklind, J., Weems, D.C., Wu, Y., Xu, I., Yoo, D., Yoon, J. and Zhang, P. 2003. The *Arabidopsis*

- Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community Nucleic Acids Res. 31: 224–228.
- Roxas, V.P., Lodhi, S.A., Garrett, D.K., Mahan, J.R. and Allen, R.D. 2000. Stress tolerance in transgenic tobacco seedlings that overexpress glutathione S-transferase/glutathione peroxidase Plant Cell Physiol. 41: 1229–1234.
- Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K. 2002. Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high salinity stresses using a full-length cDNA microarray Plant J. 31: 279–292.
- Staden, R., Beal, K.F. and Bonfield, J.K. 2000. The Staden package, 1998 Methods Mol. Biol. 132: 115–130.
- Stoop, J.M.H., Wichers, H.J. and Mooibroek, H. 1996. Mannitol metabolism in salt stressed *Agaricus bisporus* Acta Bot. Neerl. 45: 572–572.
- Teusink, R.S., Rahman, M., Bressan, R.A. and Jenks, M.A. 2002. Cuticular waxes on *Arabidopsis thaliana* close relatives *Thellungiella halophila* and *Thellungiella parvula* Internat. J. Plant Sci. 163: 309–315.
- Thomas, H. and Stoddart, J.L. 1980. Leaf senescence Annu. Rev. Plant Physiol. 31: 83–111.
- Thomashow, M.F. 1999. Plant cold acclimation: freezing tolerance genes and cold acclimation Physiol. Plant Mol. Biol. 50: 571–599.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice Nucleic Acids Res. 22: 4673–4680.
- Volkov, V., Wang, B., Dominy, P.J., Fricke, W. and Amtmann, A. 2004. *Thellungiella halophila*, a salt-tolerant relative of *Arabidopsis thaliana*, possesses effective mechanisms to discriminate between potassium and sodium Plant Cell Environ. 27: 1–14.
- Wang, Z.I., Li, P.H., Fredricksen, M., Gong, Z.H., Kim, C.S., Zhang, C.Q., Bohnert, H.J., Zhu, J.K., Bressan, R.A., Hasegawa, P.M., Zhao, Y.X. and Zhang, H. 2004. Expressed sequence tags from *Thellungiella halophila*, a new model to study plant salt-tolerance Plant Sci. 166: 609–616.
- Wang, W., Vinocur, B. and Altman, A. 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance Planta 218: 1–14.
- Willekens, H., Chamnongpol, S., Davey, M., Schraudner, M., Langebartels, C., Montagu, M.Van, Inze, D. and Camp, W.Van 1997. Catalase is a sink for H₂O₂ and is indispensable for stress defence in C-3 plants EMBO J. 16: 4806–4816.
- Yan Wang, J., Tissue, D., Holaday, A.S., Allen, R. and Zhang, H. 2003. Photosynthesis and seed production under water-deficit conditions in transgenic tobacco plants that overexpress an *Arabidopsis* ascorbate peroxidase gene Crop Sci. 43: 1477–1483.
- Zhu, J.K. 2001. Plant salt tolerance Trends Plant Sci. 6: 66–71.