

Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. *indica* cv. Nagina 22)

Arjula R. Reddy, W. Ramakrishna, A. Chandra Sekhar, Nagabhushana Ithal, P. Ravindra Babu, M.F. Bonaldo, M.B. Soares, and Jeffrey L. Bennetzen

Abstract: We have utilized an efficient method to enrich cDNA libraries for novel genes and genes responsive to drought stress in rice (*Oryza sativa* L. subsp. *indica*). We separately constructed standard and normalized cDNA libraries from leaf tissue of rice seedlings grown under controlled drought stress. Sequencing from the 3' end was performed on 1000 clones from the normalized leaf cDNA library and 200 clones from the standard leaf cDNA library. For the first 200 clones, the clone redundancy in the non-normalized library was about 10%, compared with 3.5% in the normalized cDNA library. Comparison of these cDNAs with the sequences in public databases revealed that 28.2% of the expressed sequence tags (ESTs) from the normalized library were novel. Clones from the standard and normalized leaf libraries and a root library uncovered numerous cDNAs that are highly homologous to known drought-responsive genes including those that encode metallothioneins, late embryonic abundant (LEA) proteins, heat-shock proteins, cytochrome P450 enzymes, catalases, peroxidases, kinases, phosphatases, and transcription factors.

Key words: *Oryza sativa* L., drought tolerance, normalization, ESTs, redundancy, cDNA library.

Résumé : Les auteurs ont employé une méthode efficace pour enrichir des banques d'ADNc en gènes non encore décrits et en gènes dont l'expression est affectée par un stress hydrique chez le riz (*Oryza sativa* L. subsp. *indica*). Les auteurs ont préparé des banques d'ADNc standard et normalisée à partir de tissus foliaires de plantules de riz cultivées en conditions de déficit hydrique. Le séquençage de l'extrémité 3' a été réalisé sur 1000 clones de la banque normalisée et sur 200 clones de la banque standard. Pour les 200 premiers clones, la redondance était d'environ 10 % chez la banque standard et de 3,5 % pour la banque normalisée. Une comparaison de ces ADNc avec les séquences disponibles dans les bases de données publiques a révélé que 28,2 % des étiquettes de séquences exprimées (EST) de la banque normalisée n'avaient pas encore été décrits. Des clones des banques normalisée et standard ainsi que des clones d'une banque racinaire ont montré une forte homologie avec des gènes connus comme étant régulés par le stress hydrique dont des gènes codant pour des métallothionéines, des protéines abondantes dans l'embryon mature (LEA), des protéines de choc thermique, des cytochromes P450, des catalases, des peroxydases, des kinases, des phosphatases et des facteurs de transcription.

Mots clés : *Oryza sativa* L., tolérance à la sécheresse, normalisation, EST, redondance, banque d'ADNc.

[Traduit par la Rédaction]

Introduction

Rapid advances in genomic technologies are leading to an increased understanding of global gene expression in plants.

Numerous projects are aimed at identifying and characterizing the full set of transcribed genes (transcriptome) in target organisms. Large-scale cDNA sequencing projects have

Received June 8, 2001. Accepted September 5, 2001. Published on the NRC Research Press Web site at <http://genome.nrc.ca> on January 14, 2002.

Corresponding Editor: P.B. Moens.

A.R. Reddy,^{1,2} A.C. Sekhar, N. Ithal, and P.R. Babu. Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad 500 046 (A.P.), India.

W. Ramakrishna¹ and J.L. Bennetzen. Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, U.S.A.
M.F. Bonaldo and M.B. Soares. Department of Pediatrics, University of Iowa, Iowa City, IA 52242, U.S.A.

¹These authors contributed equally to this work.

²Corresponding author (e-mail: arjulsl@uohyd.ernet.in).

identified expressed sequence tags (ESTs) and provide an effective means of identifying expressed genes in organisms across all kingdoms. Ideally, ESTs generated from a total cDNA library should represent all the expressed genes in the tissue from which the library was constructed. However, the expression patterns of different genes in a given tissue yield mRNAs that differ in abundance, making it difficult to capture rare mRNAs from cDNA libraries. This problem also leads to redundant sequencing of clones representing the same expressed genes, thereby affecting the efficiency and cost effectiveness of the EST approach (Bonaldo et al. 1996). To overcome the redundancy problem in large-scale cDNA sequencing projects, several laboratories have used cDNA library normalization. In theory, a normalized cDNA library approach will generate uniform abundances of cDNA classes within the library. Different methods for construction of normalized cDNA libraries from a variety of tissues and organisms have been reported (Weissman 1987; Patanjali et al. 1991; Soares et al. 1994). Normalized cDNA libraries developed from different human tissues and organs have proven effective in representing rare and low-abundance mRNAs (Bonaldo et al. 1996). Recently, normalized cDNA libraries have been constructed for *Arabidopsis* (Asamizu et al. 2000a), *Lotus japonicus* (Asamizu et al. 2000b), and *Triticum aestivum* (Ali et al. 2000).

Rice (*Oryza sativa* L.), the most important world food crop, has now emerged as a model crop plant for genome analysis. To date, over 75 000 rice ESTs are available in the dbEST database (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). These ESTs were mostly from *O. sativa* L. subsp. *japonica*. Large-scale EST data are not available for *O. sativa* L. subsp. *indica*, which is grown in a wide range of agroclimatic conditions ranging from deep water to dry land environments. In a typical year, abiotic stresses decrease rice yields by about 15% in Asia, more than twice the damage caused by biotic stresses (Dey and Upadhyaya 1996). In particular, drought stress is the major cause of yield instability in rice production across diverse crop ecosystems. Its multigenic, incompletely penetrant, quantitative nature makes it difficult to breed for drought tolerance. Furthermore, the drought stress response in plants involves an array of different pathways associated with stress perception, signal transduction, gene expression, and synthesis of a number of compounds. Given this complexity of the drought response, a large-scale EST approach would assist gene discovery and genetic engineering of rice for drought tolerance. Several drought stress responsive genes were identified from ESTs generated from abscisic acid (ABA) treated and desiccated moss cDNA libraries (Machuka et al. 1999; Wood et al. 1999). Recently a full-length cDNA microarray has been used to study drought- and cold-responsive genes in *Arabidopsis* (Seki et al. 2001). To date, a large-scale EST approach has not been utilized for gene discovery in drought-stressed rice plants.

We report here, for the first time, large-scale EST development from cDNA libraries constructed from drought-stressed leaf and root tissues of an upland *O. sativa* subsp. *indica* cultivar, Nagina 22. We demonstrate the efficiency and cost-effectiveness of the normalization method in enriching the leaf cDNA library for novel ESTs and identifying stress responsive genes in rice.

Materials and methods

Plant material and water stress treatment

Nagina 22 is an early maturing, deep-rooted, drought-tolerant cultivar adapted to upland conditions. Rice seedlings were grown in pots (3 seedlings/pot) with top vertisol soil maintained at defined field capacity (FC) in a Conviron growth chamber (Conviron, Winnipeg, Man.) simulating upland growth conditions. Field capacity is defined as the amount of water held in the soil after excess water has drained away and the rate of downward movement of water has perceptibly decreased. The control plants were grown at 100% FC, which is the maximum amount of water retained by the soil at saturation. FC of the soil was calculated as in Singh and Vittal (1997). The seedlings were maintained at $32 \pm 1^\circ\text{C}$ during the day and $20 \pm 1^\circ\text{C}$ during the night in 60% relative humidity. A photoperiod of 11 h light : 13 h dark was used throughout this experiment. One-month-old seedlings grown at 70% FC were subjected to drought stress by regulating water supply to gradually reach 50% FC. Leaf and root samples were harvested at 50% FC between 11:00 and 13:00. At each sampling point, leaf relative water content (RWC) was measured from the mid-portion of the leaf.

Construction of cDNA libraries

Total RNA was isolated from drought-stressed leaf and root samples using Trizol reagent (Life Technologies, Rockville, Md.). Poly(A⁺) RNA was purified from total RNA using Oligotex suspension (Qiagen GmbH, Hilden, Germany). First-strand cDNA was synthesized by priming poly(A⁺) RNA with the *NotI*-(dT)₁₈ (Amersham Pharmacia Biotech, Uppsala, Sweden) oligonucleotide primer (5'-TGT-TACCAATCTGAAGTGGGAGCGGCCGACAA(T)₁₈-3') using Superscript reverse transcriptase (Life Technologies); second-strand cDNA synthesis and blunt ends were made as described by Soares and Bonaldo (1997). *EcoRI* adapters (Amersham Pharmacia Biotech) were added to double-stranded cDNAs that had been size selected and blunt ended. The phosphorylated cDNAs were digested with *NotI* and cloned into *EcoRI*- and *NotI*-digested phagemid vector pT7T3-Pac (Bonaldo et al. 1996). The pT7T3-Pac vector is essentially the same as pT7T318D (Amersham Pharmacia Biotech) except that it has a slightly different polylinker. The M13 reverse primer (5'-AGCGGATAACAATTTCCACACAGGA-3') homology is located 50 bp upstream from the *SfiI* site and the M13 universal-sequencing primer (5'-GTT-TTCCCAGTCACGAC-3') homology is located 53 bp downstream from the *HindIII* site. Furthermore, the T7 promoter primer (5'-TAATACGACTCACTATAGGGA-3') homology is 4 bp upstream from the *SfiI* site and the T3 promoter primer (5'-TCCCTTTAGTGAGGGTTAAT-3') homology is only 2 bp downstream from the *HindIII* site. The cDNA library constructed in pT7T3-Pac was amplified in *Escherichia coli* DH10B (Life Technologies) electrocompetent cells using Gene Pulser (Bio-Rad, Hercules, Calif.). The library was size selected on an agarose gel for an insert size of above 300 bp and extracted from the gel using the Qiagen gel extraction kit.

Normalization

The normalization procedure was essentially according to

Bonaldo et al. (1996) and Soares and Bonaldo (1997) with some modifications. Single-stranded DNA (ssDNA) was prepared from the directionally cloned cDNA library using helper bacteriophage M13K07 (Amersham Pharmacia Biotech). ssDNA was incubated with *Pvu*II (New England Biolabs, Beverly, Mass.) and purified using the Bio-Gel HTP hydroxyapatite (HAP) column (Bio-Rad). The driver DNA was prepared by PCR amplification of ssDNA using T7 and T3 primers and *Taq* DNA polymerase (Qiagen). The PCR products were purified using Centrisep spin columns (Princeton Separations, Adelphia, N.J.). Hybridization of driver with ssDNA tracer was performed in the presence of a 5'-blocking oligo, a 3'-blocking oligo, and a tail-blocking oligo (poly(A)) (Amersham Pharmacia Biotech) in the presence of 50% formamide. The reaction mixture was heated to 80°C for 3 min and reassociation was carried out in the presence of 120 mM NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 1% SDS at 30°C. The reaction was allowed to proceed for 22 h to achieve the calculated C_0t value of 5 s-mol/L. The fraction that remains ssDNA after reassociation represents the unamplified, normalized library, which was subsequently purified from the reassociated molecules by chromatography on HAP columns. The reassociation reaction mixture was adjusted to 0.11 M sodium phosphate (pH 6.8), 10 mM EDTA, and 1% SDS. This mixture was loaded onto a HAP column (glass jacketed; Bio-Rad) that had been prewarmed to 40°C. The column was subsequently heated and maintained at 60°C. The eluate was collected and reloaded onto the column for the second time. Using radiolabelled ssDNA, the ssDNA fraction was found to elute at 0.1 M SP buffer (0.1 M sodium phosphate (pH 6.8), 10 mM EDTA, and 1% SDS) and dsDNA was found to elute at 0.4 M SP (0.4 M sodium phosphate (pH 6.8), 10 mM EDTA, and 1% SDS). The column was washed twice with 3 mL of 0.1 M SP buffer to elute the ssDNA. ssDNA eluted from the HAP column was concentrated by butanol extraction and ethanol precipitation. ssDNA was converted into partial duplexes by controlled primer extension using the bacteriophage M13 forward primer 5'-GTAAAACGACGGCCAGT-3' and Sequenase (Amersham Pharmacia Biotech). Partial duplexes were then transformed into *E. coli* DH10B by electroporation. The plasmid DNA prepared from these bacteria represents the amplified normalized cDNA library.

Sequencing and analysis of ESTs

Randomly selected clones from all libraries were partially sequenced from the 3' end on an ABI 3700 (Perkin Elmer, Foster City, Calif.). Base calling was done using phred (Ewing and Green 1998) and sequences were assembled using phrap to estimate redundancy and viewed using consed (Gordon et al. 1998). The ESTs were further analyzed using BLASTX and BLASTN. ESTs were defined as redundant when they exhibited more than 95% identity over aligned regions or to the same database accession.

Results and discussion

Experimental strategy

Molecular approaches towards understanding the drought response in rice are largely confined to studies on individual

drought-responsive genes and gene products. Genomic approaches to analyze drought-responsive genes in rice have not been reported. To study global gene expression under drought stress, we constructed cDNA libraries from leaf and root tissues of drought-stressed rice seedlings. The physiological status of the leaf material at 50% FC was primarily assessed by RWC and only the samples exhibiting RWC of 50–60% were selected for the cDNA library preparation. Control plants at 100% FC showed 95% RWC. The seedlings at 50% FC exhibited clear symptoms of drought stress such as leaf rolling and basal leaf senescence. The rationale of growing plants at 70% FC and then gradually reducing the moisture level to 50% FC is to enrich the library with genes involved in stress response and adaptation rather than shock as the result of severe stress. The standard leaf cDNA library made from poly(A⁺) RNA of these samples contained about 200 000 recombinant clones with insert sizes ranging from 300 to 1500 bp, where the median was about 800 bp. The root standard library, after size fractionation, contained about 5000 clones with an average insert size of about 800 bp.

Normalization and generation of ESTs from the stressed leaf cDNA library

The normalization of a standard leaf cDNA library using the modified procedure of Bonaldo et al. (1996) yielded about 25 000 clones. Randomly picked clones from this normalized cDNA library were partially sequenced from the 3' end. About 1000 high quality (phred score > 20) ESTs were generated and analyzed using the BLAST algorithm (Altschul et al. 1997) applied to the current GenBank database. Similarly, 200 ESTs generated from the standard leaf cDNA library were also analyzed. Phrap analysis revealed a redundancy of about 3.5% in the normalized leaf library compared with 10% in the standard library for the first 200 clones analyzed (Table 1). The frequency of redundant clones among ESTs from the normalized leaf cDNA library also was calculated for all 1000 ESTs (Fig. 1). ESTs were defined as redundant when they gave a BLASTX or BLASTN hit to the same accession number or when they exhibited more than 95% identity over aligned regions and were assembled in a single contig. Of the 1000 normalized leaf ESTs, 735 were unique, 62 were represented twice, 16 were found three times, 17 were found four times, and 5 appeared five times. Novel genes were defined as ESTs not showing a match to any other nucleotide sequences in the database by a cut off *E* value of 10^{-6} . Novel ESTs constituted only 5% of the ESTs of the non-normalized library as compared with 28.2% of the normalized library. Beyond this greater than fivefold increase in novel ESTs, normalization also yielded a greatly reduced redundancy of cDNAs in the library. Hence, we feel that this is a very cost-effective procedure for large-scale EST generation and gene discovery.

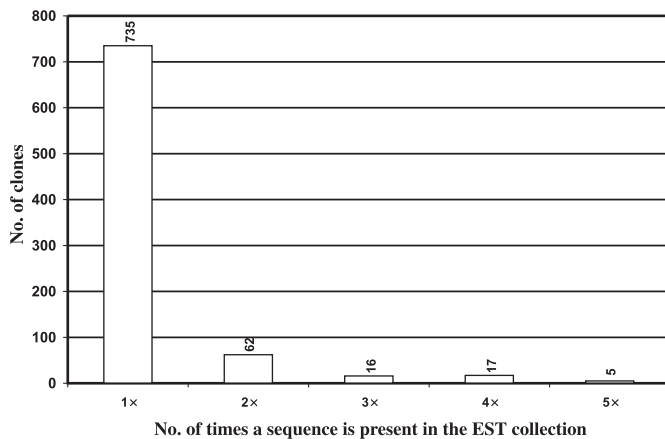
Putative genes identified from ESTs of leaf and root cDNA libraries

ESTs generated from leaf and root cDNA libraries were screened against the current GenBank database using the BLAST algorithm. Putative functions of the ESTs were assigned after applying a stringency level of *E* value of 10^{-6} . A total of 718 out of 1000 ESTs from the normalized leaf

Table 1. Comparison of standard and normalized cDNA libraries from drought-stressed leaves.

	Leaf	
	Non-normalized	Normalized
Average insert size (bp)	800	800
Average read length (bp)	500	500
No. of ESTs generated	200	1000
% redundancy	10*	3.5*
% novel genes	5.1	28.2

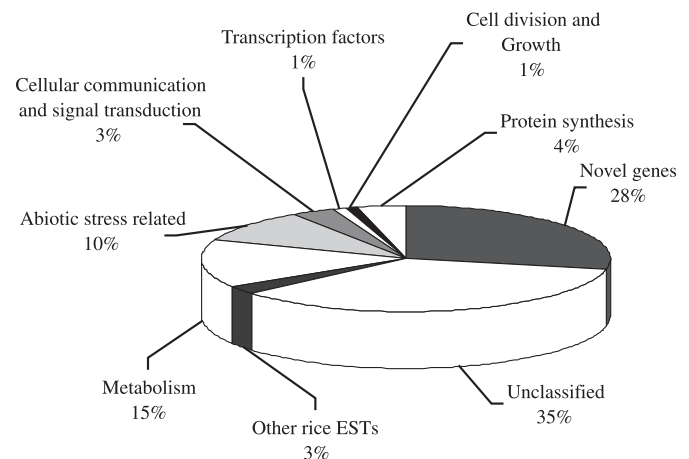
*Percentages for the first 200 clones analyzed.

Fig. 1. Frequency of redundant clones among ESTs from the normalized leaf cDNA library.

cDNA library showed significant similarity to known sequences in the database. The remaining 282 ESTs did not show significant homology to any known sequences in the databases and were deemed novel. ESTs with matches in the database were classified based on their putative function (Fig. 2). Genes involved in metabolism constitute the most abundant class among ESTs. Genes related to drought-stress response are highly represented among ESTs from stressed seedlings.

Metallothionein-like genes turned out to be the most abundant class in the normalized leaf library. In an earlier study of transcript profiling of rice seedlings using serial analysis of gene expression (SAGE), metallothionein-like sequences were found to be an abundant class, suggesting that they might perform essential functions of plant growth besides metal detoxification (Matsumura et al. 1999). Such predicted functions include cell wall lignification, cell elongation (Omann et al. 1994; Yu et al. 1998), and reducing the concentration of free metal ions in the cell to prevent the increase of reactive oxygen species under water stress conditions (Batt et al. 1998). They are also reported to be upregulated upon salt stress in rice (Kawasaki et al. 2001). Several ESTs showed significant sequence similarity to genes that had already been shown to be affected by ABA, drought, and other environmental stresses in different plants (Table 2). Examples include glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Velasco et al. 1994), aldolase (Michelis and Gepstein 2000), *rd22* (Yamaguchi-Shinozaki and Shinozaki 1993), late embryonic abundant (LEA) proteins (Baker et al.

Fig. 2. Expression analysis of 1000 ESTs from the normalized leaf cDNA library from drought-stressed rice seedlings. cDNA clones with BLASTX scores 10^{-5} were classified according to their putative functions. Novel genes, no hits to known amino acid sequences; unclassified, sequences with known homologs but not included in any of the above classes; other rice ESTs, matching to unannotated rice ESTs.



1988; Close et al. 1989; Xu et al. 1996), and heat-shock proteins (Vierling 1991). Genes for a variety of transcription factors that contain typical DNA binding motifs, such as dehydration responsive element binding (DREB1A) protein, MYB, MYC, AP2, and zinc fingers that have been reported as stress inducible (Shinozaki and Yamaguchi-Shinozaki 1999), were also identified among these ESTs, as were many classes of protein kinases, such as mitogen activated protein (MAP) kinases, calcium-dependent protein kinase (CDPK), and protein phosphatases that were demonstrated to be either stress inducible or upregulated by dehydration (Mizoguchi et al. 1996; Shinozaki and Yamaguchi-Shinozaki 1997; Xu et al. 1998). Other genes include those implicated in oxidative stress response such as catalases, peroxidases, and glutathione *S*-transferase. In fact, a significant number of these ESTs are found to be up-regulated by salt stress in rice roots as revealed by microarray data (Kawasaki et al. 2001). These genes include homologues of glutathione *S*-transferase, LEA, *S*-adenosylmethionine decarboxylase, *S*-adenosylmethionine synthetase, water channel proteins, CDPK, peroxidase, calmodulin, ascorbate peroxidase, and *asr1* (ABA- and stress-responsive protein). Similarly, a comparison with gene expression profiles of *Arabidopsis* under drought and cold stress also revealed a number of common genes such as those for LEA, ascorbate peroxidase, enolase, glycine-rich protein, thioredoxin, catalase, and ethylene-responsive element binding protein (Seki et al. 2001). Our EST collection therefore represents a rich source of drought-responsive genes and will be useful in expression analysis in rice and other grasses.

Mining for SSRs (simple-sequence repeats) among the ESTs from our leaf library resulted in the identification of 101 SSRs represented by 8 dinucleotide, 53 trinucleotide, 34 tetranucleotide, and 6 pentanucleotide repeats (data not shown) that can be useful in marker development and molecular mapping strategies.

Table 2. Putative functions assigned to ESTs from normalized library prepared from leaves of drought-stressed rice seedlings.

EST accession No. ^a	Putative function	Identical registration ^b	Organism	E value
Stress related				
BI305515	Putative RAD23	AC021640	<i>Arabidopsis thaliana</i>	4×10 ⁻⁵⁵
BI305617	Metallothionein-like proteins	U77294	<i>Oryza sativa</i>	1×10 ⁻¹⁰
BI305618	GrepE protein	AJ010819	<i>Arabidopsis thaliana</i>	2×10 ⁻²⁶
BI305676	Glycine-rich protein	AL010579	<i>Oryza sativa</i>	6×10 ⁻⁰⁶
BI305699	Remorin gene	A005244	<i>Arabidopsis thaliana</i>	4×10 ⁻¹⁵
BI305703	Metallothionein-like protein	U18404	<i>Oryza sativa</i>	9×10 ⁻¹⁷
BI305736	Transaldolase	U95923	<i>Solanum tuberosum</i>	1×10 ⁻⁵¹
BI305766	Hypersensitivity-related gene	X95343	<i>Nicotiana tabacum</i>	1×10 ⁻²⁰
BI305776	Ethylene-responsive element binding factor 3	AB037183	<i>Oryza sativa</i>	8×10 ⁻²⁸
BI305797	Enolase	U17973	<i>Zea mays</i>	1×10 ⁻⁵⁰
BI306042	Glutamine synthetase	D14577	<i>Zea mays</i>	6×10 ⁻²²
BI305726	Low molecular weight heat-shock protein precursor	AF035460	<i>Zea mays</i>	1×10 ⁻³⁷
BI305910	Putative leucine-rich repeat protein	AC004165	<i>Arabidopsis thaliana</i>	9×10 ⁻¹⁰
BI305897	Cytochrome <i>c</i> oxidase subunit 5c	AB027123	<i>Oryza sativa</i>	4×10 ⁻⁶
BI305977	Abscisic acid inducible	X59138	<i>Zea mays</i>	2×10 ⁻³⁹
BI305990	Peroxidase	X98322	<i>Arabidopsis thaliana</i>	4×10 ⁻⁴³
BI306006	Group 4 LEA protein	M88321	<i>Gossypium hirsutum</i>	7×10 ⁻³¹
BI306013	Xyloglucan endotransglycosylase (XET)	X93175	<i>Hordeum vulgare</i>	2×10 ⁻³⁴
BI306026	Aldolase C-1	D50307	<i>Oryza sativa</i>	1×10 ⁻²⁹
BI306095	Metallothionein-like protein	AF001396	<i>Oryza sativa</i>	3×10 ⁻²³
BI306121	Dehydrogenase-like protein	AB025639	<i>Arabidopsis thaliana</i>	2×10 ⁻⁴⁸
BI306125	Shaggy-like protein	Y13437	<i>Oryza sativa</i>	1×10 ⁻¹⁴⁸
BI306129	rd22	AP000364	<i>Oryza sativa</i>	2×10 ⁻³²
BI306144	Glyceraldehyde-3-phosphate dehydrogenase	U31676	<i>Oryza sativa</i>	5×10 ⁻¹¹
BI306214	Chaperonin 21 precursor	AF233745	<i>Lycopersicon esculentum</i>	3×10 ⁻⁵¹
BI306235	EF-hand Ca ²⁺ -binding protein CCD1	AF181661	<i>Arabidopsis thaliana</i>	6×10 ⁻²⁴
BI306254	Glycine-rich protein	AF011331	<i>Oryza sativa</i>	2×10 ⁻⁴²
BI306233	Peroxidase BP1	M73234	<i>Hordeum vulgare</i>	1×10 ⁻³¹
BI306328	Wound induced protein	X59882	<i>Lycopersicon esculentum</i>	1×10 ⁻⁶
BI306369	Peroxiredoxin Q	AB037598	<i>Sedum lineare</i>	8×10 ⁻⁶⁶
BI306376	Open reading frame; able to induce HR-like lesions	U66269	<i>Nicotiana tabacum</i>	5×10 ⁻³⁷
BI306388	Jasmonate-induced protein	X98124	<i>Hordeum vulgare</i>	1×10 ⁻²¹
BI306411	Glutathione <i>S</i> -transferase II	AF062403	<i>Oryza sativa</i>	1×10 ⁻³⁷
BI306443	Catalase	D26484	<i>Oryza sativa</i>	1×10 ⁻⁵⁶
BI306457	Glyceraldehyde-3-phosphate dehydrogenase	M36650	<i>Hordeum vulgare</i>	1×10 ⁻²⁹
BI306437	rd22	D10703	<i>Arabidopsis thaliana</i>	5×10 ⁻³⁸
BI306029	Hsp82	ZI5018	<i>Oryza sativa</i>	2×10 ⁻⁶
BI305808	Cytochrome P450	M32885	<i>Persea americana</i>	4×10 ⁻⁴³
BI305704	<i>S</i> -Adenosyl methionine synthetase	Z26867	<i>Oryza sativa</i>	1×10 ⁻³⁸
BI305631	Cytochrome <i>b</i> ₅	X75670	<i>Oryza sativa</i>	3×10 ⁻¹⁷
BI306153	Cytochrome <i>c</i> oxidase subunit	AB027123	<i>Oryza sativa</i>	1×10 ⁻³⁰
Transcription factors				
BI305518	Similar to <i>Methanobacterium thermoautotrophicum</i> transcriptional regulator	AF118223	<i>Arabidopsis thaliana</i>	3×10 ⁻⁶⁰
BI305544	Myb DNA binding protein-like	AL3555775	<i>Arabidopsis thaliana</i>	2×10 ⁻¹³
BI305625	Zinc finger protein, putativae, 5' partial	AC069474	<i>Arabidopsis thaliana</i>	5×10 ⁻³⁰
BI305705	Zinc finger protein	AB028132	<i>Oryza sativa</i>	1×10 ⁻¹⁵
BI305762	Transcription factor IIA small subunit	AJ223634	<i>Arabidopsis thaliana</i>	2×10 ⁻¹⁴
BI306059	MybHv5	X70876	<i>Hordeum vulgare</i>	3×10 ⁻¹⁶
BI306078	Similar to RING-H2 finger protein RHA1a	AP000616	<i>Oryza sativa</i>	2×10 ⁻⁴⁸
BI306209	Similar to <i>Lotus japonicus</i> gene encoding RING finger protein	AP002521	<i>Oryza sativa</i>	6×10 ⁻⁴³
BI306249	Putative transcription factor BTF3	AC010556	<i>Arabidopsis thaliana</i>	2×10 ⁻⁵⁶
BI306273	Similar to RING-H2 finger protein RHA1a	AP000616	<i>Oryza sativa</i>	1×10 ⁻⁵⁶
BI306362	Putative zinc finger protein	AC079281	<i>Arabidopsis thaliana</i>	1×10 ⁻²⁵

Table 2 (concluded).

EST accession No. ^a	Putative function	Identical registration ^b	Organism	E value
BI306120	Putative AP2 domain transcription factor	AC016529	<i>Arabidopsis thaliana</i>	1×10 ⁻¹²
BI305874	AP2 domain containing protein	AF071893	<i>Prunus armeniaca</i>	4×10 ⁻¹⁵
BI305899	VIP2 protein	AJ251051	<i>Avena fatua</i>	2×10 ⁻⁴³
BI306107	Similar to mRNA for DREB1A	AP001168	<i>Oryza sativa</i>	2×10 ⁻³⁹
Cellular communication and signal transduction				
BI305642	Putative phytochrome-associated protein	AP002743	<i>Oryza sativa</i>	4×10 ⁻³¹
BI305653	ADP-rybosylation factor 1	AF012896	<i>Oryza sativa</i>	1×10 ⁻⁷³
BI305663	10-kDa phosphoprotein	X12695	<i>Oryza sativa</i>	7×10 ⁻¹⁸
BI305778	Photoreceptor-interacting protein-like	AB013389	<i>Arabidopsis thaliana</i>	2×10 ⁻¹²
BI305795	ADP ribosylation factor	AF108891	<i>Capsicum annuum</i>	1×10 ⁻³²
BI305802	Small GTP binding protein	AF112964	<i>Triticum aestivum</i>	7×10 ⁻⁶⁷
BI305837	Small GTP-binding protein OsRac3	AB029510	<i>Oryza sativa</i>	3×10 ⁻⁵³
BI306130	Protein kinase, putative	AC027135	<i>Arabidopsis thaliana</i>	2×10 ⁻⁶⁹
BI306475	Calmodulin	AF042839	<i>Oryza sativa</i>	4×10 ⁻⁵⁸
BI305849	Serine (threonine) protein phosphatase	AC002411	<i>Arabidopsis thaliana</i>	1×10 ⁻²³
BI306067	OsCDPK7	AB042550	<i>Oryza sativa</i>	9×10 ⁻⁸⁷
BI306284	Serine (threonine) protein kinase	AF159691	<i>Myxococcus xanthus</i>	2×10 ⁻³⁴

^aGenBank accession Nos. of ESTs from the present study.

^bGenBank accession Nos. of registered sequences that are identical to our ESTs.

Table 3. Putative functions assigned to ESTs from the cDNA library prepared from roots of drought-stressed rice seedlings.

EST accession No. ^a	Putative function	Identical registration ^b	Organism	E value
Stress related				
BI305193	Respiratory burst oxidase protein	AB016886	<i>Arabidopsis thaliana</i>	2×10 ⁻⁴⁷
BI305199	Ascorbate peroxidase	D45423	<i>Oryza sativa</i>	3×10 ⁻⁶⁴
BI305212	Respiratory burst oxidase protein D	AF05357	<i>Arabidopsis thaliana</i>	1×10 ⁻¹¹⁴
BI305213	Heat-stress transcription factor A3		<i>Lycopersicon peruvianum</i>	6×10 ⁻⁶⁵
BI305274	RAB21 protein	Y00842	<i>Oryza sativa</i>	7×10 ⁻³⁰
BI305310	Glyceraldehyde-3-phosphate dehydrogenase	U31676	<i>Oryza sativa</i>	1×10 ⁻¹³⁹
BI305323	Ethylene-responsive binding factor 3	AB037183	<i>Oryza sativa</i>	4×10 ⁻²⁴
BI305325	Lipoxygenase	L23968	<i>Arabidopsis thaliana</i>	9×10 ⁻⁸⁵
BI305339	Plastidic aldolase NPALDP1	AB027001	<i>Nicotiana paniculata</i>	1×10 ⁻¹¹⁰
BI305378	Heat-stress transcription factor	AF208544	<i>Lycopersicon peruvianum</i>	1×10 ⁻⁶⁵
BI305298	Cytochrome P450 monooxygenase	AJ004810	<i>Zea mays</i>	6×10 ⁻³⁸
BI305352	S-Adenosylmethionine decarboxylase 2	AJ251899	<i>Oryza sativa</i>	1×10 ⁻¹⁰²
Cellular communication and signal transduction				
BI305186	MAP 3 kinase	AF076275	<i>Arabidopsis thaliana</i>	3×10 ⁻¹¹
BI305216	Putative protein kinase	AC005623	<i>Arabidopsis thaliana</i>	7×10 ⁻²⁹
BI305220	Calcium-dependent protein kinase	AL133248	<i>Arabidopsis thaliana</i>	3×10 ⁻¹¹³
BI305267	Protein-kinase-like protein	AL356014	<i>Arabidopsis thaliana</i>	2×10 ⁻²⁴
BI305315	Serine (threonine) protein kinase	Y12465	<i>Sorghum bicolor</i>	1×10 ⁻⁶⁹
BI305307	Signal peptidase I	AE004511	<i>Pseudomonas aeruginosa</i>	4×10 ⁻⁶⁸
BI305358	MAP3k beta 1 protein kinase	AJ010093	<i>Brassica napus</i>	2×10 ⁻⁸³

^aGenBank accession Nos. of ESTs from the present study.

^bGenBank accession Nos. of registered genes that are identical to our ESTs.

The limited data from 300 ESTs of our root library show the expression of many classes of kinases, perhaps in accordance with the likelihood that roots are the first organs of drought stress perception and signal transduction (Table 3). Although three different GAPDH clones were identified in the root cDNA library, only one of them was found among

1000 ESTs from the leaf cDNA library. Interestingly, the most redundant classes of genes in the root library are cytochrome P-450 oxidases and serine (threonine) protein kinases. With these limited data, it is not possible to interpret the role of cytochrome P-450 enzymes in the drought stress response in rice. In summary, our results highlight a simple

and cost-effective method to enrich the cDNA libraries for novel genes and drought stress responsive genes in rice.

Conclusions and perspectives

Plant ESTs have been increasingly used in analyzing global gene expression and function, and as markers and probes in genome mapping (Ewing et al. 1999). Several large-scale rice EST projects are underway (Sasaki et al. 1994; Umeda et al. 1994; Yamamoto and Sasaki 1997); however, of more than 75 000 rice entries in dbEST, only a few are from stressed whole plants (Kawasaki et al. 2001). No entries represent ESTs from drought-stressed rice plants. Large-scale EST projects are often compromised by high redundancy and thus an increased cost of novel gene discovery. In this project, we report a cost-effective way of reducing the redundancy and increasing the novel gene component using normalized cDNA libraries. We are generating more ESTs from our leaf and root cDNA libraries and will use these in high-throughput analyses of gene expression patterns. Our ongoing experiments in EST generation, mapping, and expression analysis will help in identification of novel genes involved in drought tolerance in rice.

Acknowledgements

This project was funded by the Rockefeller Foundation in the form of a Research Grant and Career Biotechnology Fellowship to A.R.R. and by Purdue University support to J.L.B. We acknowledge the help of Dr. Callum J. Bell, National Center for Genome Resources, New Mexico City, New Mexico U.S.A. We thank Dr. N.P. Saxena for facilitating the plant growth experiments at the International Crop Research Institute for the Semi Arid Tropics, Hyderabad, India.

References

- Ali, S., Holloway, B., and Taylor, W.C. 2000. Normalization of cereal endosperm EST libraries for structural and functional genomic analysis. *Plant Mol. Biol. Rep.* **18**: 123–132.
- Altschul, Stephen, F., Madden, T.L., Schäffer, 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.
- Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S. 2000a. A large-scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12 028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. *DNA Res.* **7**: 175–180.
- Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S. 2000b. Generation of 7137 non-redundant expressed sequence tags from a legume, *Lotus japonicus*. *DNA Res.* **7**: 127–130.
- Baker, J., Steele, C., and Dure, L., III. 1988. Sequence and characterization of six LEA proteins and their genes from cotton. *Plant Mol. Biol.* **11**: 277–291.
- Batt, A., Mousley, G., and Morris, K. 1998. Differential expression of senescence-enhanced metallothionein gene in *Arabidopsis* in response to isolates of *Peronospora parasitica* and *Pseudomonas syringae*. *Plant J.* **16**: 209–221.
- Bonaldo, M.F., Lennon, G., and Soares, M.B. 1996. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res.* **6**: 791–806.
- Close, T.J., Kortt, A.A., and Chandler, P.M. 1989. A cDNA based comparison of dehydration induced proteins (dehydrins) in barley and corn. *Plant Mol. Biol.* **13**: 95–108.
- Dey, M.M., and Upadhaya, H.K. 1996. Yield loss due to drought, cold, and submergence in Asia. In *Rice research in Asia, progress and priorities*. Edited by R.E. Evenson, R.W. Herdt, and M. Hossain. Oxford University Press, Cary, N.C. pp. 231–242.
- Ewing, B., and Green, P. 1998. Base calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**: 186–194.
- 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.
- Gordon, D., Abajian, C., and Green, P. 1998. consed: a graphical tool for sequence finishing. *Genome Res.* **8**: 195–202.
- Kawasaki, S., Borchert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., Galbraith, D., and Hans J. Bohnert, H.J. 2001. Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell*, **13**: 889–906.
- Machuka, J., Bashiardes, S., Ruben, E., Spooner, K., Cuming, A., Knight, C., and Cove, D. 1999. Sequence analysis of expressed sequence tags from an ABA-treated cDNA library identifies stress response genes in the moss *Physcomitrella patens*. *Plant Cell Physiol.* **40**: 378–387.
- Matsumura, H., Nirasawa, S., and Terauchi, R. 1999. Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *Plant J.* **20**: 719–726.
- Michelis, R., and Gepstein, S. 2000. Identification and characterization of a heat-induced isoform of aldolase in oat chloroplast. *Plant Mol. Biol.* **44**: 487–498.
- Mizoguchi, T., Irie, K., Hirayama, T., Hayashida, N., Yamaguchi-Shinozaki, K., Matsumoto, K., and Shinozaki, K. 1996. A gene encoding a MAP kinase is induced simultaneously with genes for a MAP kinase and an S6 kinase by touch, cold and water stress in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* **93**: 765–769.
- Omman, F., Beaulieu, N., and Tyson, H. 1994. cDNA sequence and tissue-specific expression of an anionic flax peroxidase. *Genome*, **37**: 137–147.
- Patanjali, S.R., Parimoo, S., and Weissman, S.M. 1991. Construction of a uniform-abundance (normalized) cDNA library. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 1943–1947.
- Sasaki, T., Song, J., Kong-Ban, Y., Matsui, E., Fang, F., Higo, H., Nagasaki, H., Hori, M., Miya, M., Murayama-Kayano, E., Takiguchi, T., Takasuga, A., Niki, T., Ishimaru, K., Ikeda, H., Yamamoto, Y., Mukai, Y., Ohta, I., Miyadera, N., Havukkala, I., and Minobe, Y. 1994. Towards cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. *Plant Journal*, **6**: 615–624.
- Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y., and Shinozaki, K. 2001. Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell*, **13**: 61–72.
- Shinozaki, K., and Yamaguchi-Shinozaki, K. 1997. Gene expression and signal transduction in water stress response. *Plant Physiol.* **115**: 327–334.
- Shinozaki, K., and Yamaguchi-Shinozaki, K. 1999. Molecular responses to drought stress. In *Molecular responses to cold, drought, heat and salt stress in rice*. Edited by K. Shinozaki and K. Yamaguchi-Shinozaki. R.G. Landes Company, Austin, Tex. pp. 11–28.

- Singh, P., and Vittal, K.P.R. 1997. Soil moisture measurement. *In* Measuring soil processes in agricultural research. Technical manual No. 3. Edited by K.B. Laryea, P. Pathak, and J.C. Katyal. International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India. pp. 34–49.
- Soares, M.B., and Bonaldo, M.F. 1997. Constructing and screening normalized cDNA libraries. *Genome analysis: a laboratory manual*. Vol. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. pp. 49–157.
- Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L., and Efstratiadis, A. 1994. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 9228–9232.
- Umeda, M., Hara, C., Matsubayashi, Y., Li, H., Liu, Q., Tadokoro, F., Aotsuka, S., and Uchimiya, H. 1994. Expressed sequence tags from cultured cells of rice (*Oryza sativa* L.) under stressed conditions: analysis of transcripts of genes engaged in ATP-generating pathways. *Plant Mol. Biol.* **25**: 469–478.
- Velasco, R., Salamini, F., and Bartels, D. 1994. Dehydration and ABA increase mRNA levels and enzyme activity of cytosolic GAPDH in the resurrection plant *Craterostigma plantagineum*. *Plant Mol. Biol.* **26**: 541–546.
- Vierling, E. 1991. The role of heat-shock proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42**: 579–620.
- Weissman, S.M. 1987. Molecular genetic techniques for mapping the human genome. *Mol. Biol. Med.* **4**: 133–143.
- Wood, A.J., Duff, R.J., and Oliver, M.J. 1999. Expressed sequence tags (ESTs) from desiccated *Tortula ruralis* identify a large number of novel plant genes. *Plant Cell Physiol.* **40**: 361–368.
- Xu, D., Duan, X., Wang, B., Ho, T.-H.D., and Wu, R. 1996. Expression of late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol.* **110**: 249–257.
- Xu, Q., Fu, H.-H., Gupta, R., and Luan, S. 1998. Molecular characterization of a tyrosine-specific protein phosphatase encoded by a stress-responsive gene in *Arabidopsis*. *Plant Cell*, **10**: 849–857.
- Yamaguchi-Shinozaki, K., and Shinozaki, K. 1993. The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of *rd22*, a gene responsive to dehydration stress in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **238**: 17–25.
- Yamamoto, K., and Sasaki, T. 1997. Large-scale EST sequencing in rice. *Plant Mol. Biol.* **35**: 135–144.
- Yu, L.H., Umeda, M., Liu, J.Y., Zhao, N.M., and Uchimiya, H. 1998. A novel MT gene of rice plants is strongly expressed in the node portion of the stem. *Gene*, **206**: 29–35.