*Journal of Experimental Botany, Vol. 58, No. 2, pp. 253–265, 2007*
*Integrated Approaches to Sustain and Improve Plant Production under Drought Stress Special Issue*
doi:10.1093/jxb/erl213 Advance Access publication 28 November, 2006

Journal of Experimental Botany

SPECIAL ISSUE PAPER

# Identification of stress-responsive genes in an *indica* rice (*Oryza sativa* L.) using ESTs generated from drought-stressed seedlings

**Markandeya Gorantla[1], P. R. Babu[1], V. B. Reddy Lachagari[1], A. M. M. Reddy[1], Ramakrishna Wusirika[2,*], Jeffrey L. Bennetzen[3] and Arjula R. Reddy[1,†]**

[1] *Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad-500046, AP, India*
[2] *Department of Biological Sciences, Purdue University, West Lafayette, IN 47906, USA*
[3] *Department of Genetics, University of Georgia, Athens, GA 30602, USA*

Received 25 November 2005; Accepted 22 September 2006

## Abstract

**The impacts of drought on plant growth and development limit cereal crop production worldwide. Rice (*Oryza sativa*) productivity and production is severely affected due to recurrent droughts in almost all agroecological zones. With the advent of molecular and genomic technologies, emphasis is now placed on understanding the mechanisms of genetic control of the drought-stress response. In order to identify genes associated with water-stress response in rice, ESTs generated from a normalized cDNA library, constructed from drought-stressed leaf tissue of an *indica* cultivar, Nagina 22 were used. Analysis of 7794 cDNA sequences led to the identification of 5815 rice ESTs. Of these, 334 exhibited no significant sequence homology with any rice ESTs or full-length cDNAs in public databases, indicating that these transcripts are enriched during drought stress. Analysis of these 5815 ESTs led to the identification of 1677 unique sequences. To characterize this drought transcriptome further and to identify candidate genes associated with the drought-stress response, the rice data were compared with those for abiotic stress-induced sequences obtained from expression profiling studies in *Arabidopsis*, barley, maize, and rice. This comparative analysis identified 589 putative stress-responsive genes (SRGs) that are shared by these diverse plant species. Further, the identified leaf SRGs were compared to expression profiles for a drought-stressed rice panicle library to identify common sequences. Significantly, 125 genes were found to be expressed under drought stress in both tissues. The functional classification of these 125 genes showed that a majority of them are associated with cellular metabolism, signal transduction, and transcriptional regulation.**

Key words: Abiotic stress, candidate genes, drought, stress-responsive genes, transcriptome.

## Introduction

Rice, the world's most important cereal crop, is the primary source of food and calories for about half of mankind (Khush, 2005). In Asia, rice provides as much as 80% of the dietary calories in countries such as Bangladesh and Indonesia. Rice-growing areas span the tropics, subtropics, semi-arid tropics, and temperate regions of the world. The predominantly rice-growing areas in Asia (~130 million hectares) are often threatened by severe abiotic stresses, the most common being drought. These areas include irrigated and rainfed lowlands, which together account for more than 85% of total world rice production. Drought has become the most significant constraint to realizing the yield potential of rice across all agro-climatic zones. In some years, abiotic stresses cause crop losses by as much as 50% (Boyer, 1982; Bray *et al*., 2000) and drought alone may cause yield losses of as

* Present address: Department of Biological Sciences, Michigan Technological University, Houghton, MI 49931, USA.
† To whom correspondence should be addressed. E-mail: arjulsl@uohyd.ernet.in

much as 15% (Dey *et al*., 1996). Drought spells across Asia have become more frequent and severe, leading to irregular and insufficient irrigation of the crop and depletion of groundwater resources leading to 100% yield losses in certain areas. Some genetic improvement of rice for water-limited environments has been achieved by crop breeding and improved crop husbandry. At least part of the reason for the slow progress in improving the genetic foundation of drought tolerance in rice has been a lack of sufficient genetic information about genes that govern this complex trait and its component secondary traits.

Insufficient water availability leads to a host of biochemical, physiological, and metabolic changes in rice. These changes, many apparently adaptive, include a host of biochemical pathways associated with signal perception, transduction, and regulation of gene expression in a temporal and spatial pattern. A significant number of genes, gene products, and pathways associated with drought response have been identified in rice using a variety of experimental approaches (Rabbani *et al*., 2003; Kawasaki *et al*., 2001; Matsumura *et al*., 1999; Gibbings *et al*., 2003; Gowda *et al*., 2004). Numerous laboratory water-stress experiments investigating dehydration-induced changes in rice gene expression have revealed several candidate genes that may be associated with drought tolerance. Molecular genetic analysis of drought tolerance through phenotyping and marker assisted selection (MAS) has identified several genomic regions, quantitative trait loci (QTLs), associated with drought tolerance.

With the near-completion of the rice genome sequence (Goff *et al*., 2002; Yu *et al*., 2002; IRGSP, 2005) and rapidly growing databases, complex traits like drought tolerance are now amenable to a detailed molecular analysis using genomic tools. The rice genome was variously estimated to have 37 000–60 000 genes (Goff *et al*., 2002; Yu *et al*., 2002, 2005; IRGSP, 2005). One reason for this variation in gene number estimation is a lack of supporting evidence from deep Expressed Sequence Tag (EST) coverage. Many ESTs have been generated for rice, and these have been valuable in confirming and cataloguing genes (Sasaki *et al*., 1994; Uchimiya *et al*., 1992; Umeda *et al*., 1994; Yamamoto and Sasaki, 1997; Reddy *et al*., 2002*a*; Markandeya *et al*., 2003; Zhang *et al*., 2005) and in deciphering the role of transcriptionally regulated genes in different tissues (Ewing *et al*., 1999). However, only a few studies have focused on the analysis of transcriptome profiles of rice seedlings subjected to abiotic stress (Umeda *et al*., 1994; Matsumura *et al*., 1999; Kawasaki *et al*., 2001) or drought (Babu *et al*., 2002; Markandeya *et al*., 2005). ESTs provide a direct approach for discovering genes associated with a stress response. This has been demonstrated in several plant systems (Michalek *et al*., 2002; Fernandes *et al*., 2002; Echenique *et al*., 2002; Reddy *et al*., 2002*a*; Markandeya *et al*., 2005). Rice has good EST coverage, in general, and a relatively large collection of ESTs generated from drought-stressed plants has been reported (Reddy *et al*., 2002*a*, b; Markandeya *et al*., 2003). These resources can be valuable for further expression studies using microarrays and in single nucleotide polymorphism (SNP) analysis for discovering specific alleles of target genes associated with the drought-stress response. Numerous putative drought-responsive genes have been uncovered by genome wide expression studies in rice (Rabbani *et al*., 2003; Gibbings *et al*., 2003; Gowda *et al*., 2004; Kawasaki *et al*., 2001; Matsumura *et al*., 1999, 2003). Most of these are dehydration-associated expression profiles of rice conducted under laboratory conditions, and therefore may not mimic true field drought responses. However, because the experiments were rigorously controlled and environmental variables kept at a minimum, these expression analyses provided uniquely valuable information.

The EST approach has been taken to identify genes associated with drought-stress response and tolerance in rice. A normalized cDNA library has been constructed from drought-stressed seedlings of *indica* rice cultivar, N22, and large-scale EST data sets have been generated (Reddy *et al*., 2002*a*) that have been deposited in GenBank (Reddy *et al*., 2002*b*; Markandeya *et al*., 2003). In the present study, more than 6000 additional ESTs are generated, allowing construction of an N22 unigene set of 1677 sequences. This unigene set was used in a comparative analysis of rice, *Arabidopsis*, maize, and barley for discovering shared genes for the plant drought response. The identification of 589 candidate shared genes for the plant drought response is reported here. Their predicted molecular functions and potential utility are discussed.

## Materials and methods

### *Drought-stress treatment and cDNA library construction*

A drought tolerant, deep-rooted *indica* rice genotype Nagina 22 was used for drought-stress induction under defined field capacity. Rice seedlings were grown in pots and maintained in a growth chamber simulating upland growth conditions. The seedlings were maintained at 32±1 ° C during the day and 20±1 ° C during the night in 60% relative humidity. The control plants were grown at 100% FC and 1-month-old plants grown at 70% FC were gradually subjected to drought stress in order to reach 50% FC by regulating the water supply. The physiological condition of plants at 50% FC was monitored by RWC and leaf samples were collected from the plants exhibiting 50–60% RWC. The drought-stress symptoms such as leaf rolling and basal leaf senescence were apparent at this stage in stress-induced plants, while control plants growing at 100% FC were observed to grow well showing 95% RWC. Total RNA isolation, cDNA synthesis, normalization, and the cDNA library construction technique were elaborately discussed in our previous report (Reddy *et al*., 2002*a*) wherein 1200 ESTs had been generated and deposited at NCBI (Reddy *et al*., 2002*b*).

### *cDNA cloning and EST sequencing*

In the present study, this normalized cDNA library was used for further EST generation. Chemically competent *E. coli* (DH5α′) cells

were transformed with this library and individual colonies were selected randomly. Cultures from individual colonies were grown overnight and used in plasmid DNA preparation for sequencing, after column purification (Qiagen). The quality and concentration of the template plasmid DNA was checked on 1% agarose (USB Biochemicals) gels. Acceptable quality plasmid DNAs were used directly for sequencing. ESTs were generated from 3′ end single-pass sequencing of 6144 cDNA clones using $M_{13}$ (−40) reverse primer 5′-CGCCGAGGTTTTCCCAGTCACGAC-3′ or $M_{13}$ (−20) reverse primer 5′-GTAAAACGACGGCCAGTG-3′, on an automated capillary genetic analysis system (MegaBACE 500). DYEnamic ET terminator chemistry (Amersham Biosciences) was used for sequencing reaction set-up. Post-sequencing reaction clean-up and loading of samples onto the MegaBACE 500 were according to the manufacturer's instructions with adjustments to suit our conditions. The average run time was about 180 min at 6–7 kV.

### *Sequence repositories and software resources used in EST analysis*

The EST sequences generated in the present study, as well as those reported earlier from the same library (Reddy *et al*., 2002*a*, *b*) and IR64 drought-stressed panicle ESTs (Bennett *et al*., 2002) were the primary data sources for the analyses performed. Standard sequence processing tools, Phred (Ewing and Green, 1998), Phrap and cross_match (Smith and Waterman 1981; Gotoh, 1982) were used with Codoncode InterPhace (http://www.codoncode.com). Homology searches in the NCBI database were carried out using network client software with the DNATools interface (http://www.crc.dk/dnatools).

Genome sequence data for *O. sativa* subsp. *japonica* cv. Nipponbare collected from TIGR (http://www.tigr.org/tdb/e2k1/osa1/) and draft sequence for *O. sativa* subsp. *indica* cv. 93–11 of the Beijing Genomics Institute (http://btn.genomics.org.cn/rice/) were used in the analysis. In addition, full-length cDNA sequences from Nipponbare (The Rice Full-Length cDNA Consortium, 2003) and full-length cDNA sequences of putative candidate genes derived from *Arabidopsis* expression profiling studies from The *Arabidopsis* Information Resource (www.arabidopsis.org) database were also employed. The nucleotide, protein, and EST databases at NCBI (http://www.ncbi.nlm.nih.gov) were utilized for homology searches using the BLAST program (Altschul *et al*., 1997).

### *Sequence processing and analysis*

The low quality regions present at the beginning and end of each sequence were trimmed using a Phred 20 cutoff value. Vector screening was performed using the cross_match program with Codoncode InterPhace software. Sequences were edited for the removal of oligodT tracks and other contaminants. A batch file of ESTs having greater than 100 bp length of sequence reads were submitted to the NCBI dbEST division of GenBank. After the rice genome sequence was largely completed (IRGSP, 2005), all ESTs from this project were compared to the genomic sequence. All sequences that did not exhibit excellent nucleotide homology with the Nipponbare genomic sequence were removed from GenBank, with the assumption that they were most likely to be derived from microbial contaminants. Phrap and CAP3 (Huang and Madan, 1999) assembly algorithms were used to assemble the individual ESTs into clusters of sequences derived from the same transcript as tentative consensus sequences (TCs) and singletons representing unique transcripts.

### *Annotation*

Homology searches were performed against non-redundant (nr) nucleotide and protein sequence databases using BLASTN 2.2.2 and BLASTX 2.2.2 versions of the BLAST programs (Altschul *et al*., 1997) through BLAST 2.0 network client software with the DNAtools interface (http://www.crc.dk/dnatools). The BLASTN program was used to identify rice EST hits and rice BAC/PAC clones in the non-redundant (nr) nucleotide sequence database, High Throughput Genomic Sequences (HTGS) division of GenBank and the Beijing WGS (whole genome shotgun contigs) draft sequence of the *indica* rice genome (Yu *et al*., 2002) in the NCBI database.

### *Identification of ESTs consistently associated with abiotic stress*

The ESTs associated with stress responses were identified from multiple sources, based on the compiled list of stress-regulated genes documented in more than one plant species (http://stress-genomics.org/stress.fls/expression/expression.html). In addition, data from microarray expression profiles of possible candidate gene sequences comprising 650 from *Arabidopsis* (Seki *et al*., 2001, 2002*a*, *b*; Kreps *et al*., 2002), 150 from barley (Ozturk *et al*., 2002), and 100 from rice (Matsumura *et al*., 1999; Kawasaki *et al*., 2001; Rabbani *et al*., 2003) have been used. All stress-responsive gene sequences were retrieved from the above studies and a local database was constructed and utilized for BLAST analysis. These were compared to the EST data set using TBLASTX with *E*-value >$1e^{-20}$.

## Results

### *Expressed sequence tag generation and analysis*

A total of 7794 cDNA clones were sequenced from the 3′ end; of these 6694 readable sequences were obtained with a high quality index (Phred score >20). The sequencing strategy proved to be very efficient, with a success rate of ~85%. Our optimized sequencing efforts, through preparation of high-quality, uniform concentrations of sequencing templates and reduced dye chemistries, drastically reduced the costs of single-pass sequencing. The high-quality readable sequences were screened for vector contamination, highly redundant ribosomal RNA sequences, *E. coli* DNA contamination, and these clones were eliminated from further analysis. Low-quality sequence regions were trimmed and sequences less than 100 bp in length were excluded. The resulting 5815 sequences were submitted to the dbEST division of NCBI (GenBank accession numbers: BI305180 to BI306756; BU672765 to BU673915; and CB964418 to CB967504). Of these, 390 were found to have no homologues in the nearly-completed Nipponbare rice genome sequence (IRGSP, 2005). Although it is possible that some of these are from the few rice genes that have not yet been sequenced from Nipponbare, or even very rare genes that might be found in *indica* cultivar Nagina 22 and not in *japonica* cultivar Nipponbare (Bennetzen *et al*., 2004; Ma and Bennetzen, 2004), but it is probable that most or all of these are ESTs from microbial contaminants in our field-grown rice seedlings. For instance, 380 ESTs were removed prior to the submission of the 5815 sequences because it was clear they were viral sequences from Adenoviral type 2 encoding minor capsid protein VI (Table 1). Microbial

contamination is an unavoidable outcome of EST analysis on field-grown plants, but they can easily be excluded from data analysis, now that a full rice genome sequence is available (IRGSP, 2005). A summary of the EST data is provided in Table 1.

### *Construction and functional classification of a unigene set from EST data*

Clustering of the 5815 ESTs allowed construction of a unigene set of 2067 unique gene expression products from our drought-stressed rice library. The assembly of sequences produced 1239 singletons and the remaining 4576 sequences were grouped into 828 contiguous sequences (contigs). Of these 2067, 390 were removed as microbial contaminants, leading to the identification of the 1677 N22 unigene set.

### *Sequence analysis of the N22 unigene set*

The assembled N22 unigene set comprising 1677 unigenes have been annotated and functionally classified based on the GO database (Gene Ontology Consortium, 2001). Annotation of the assembled unigene set, through homology searches in the NCBI nr nucleotide and protein databases, revealed that 57% of the unigene set has hits with known putative functions, the remaining 43% of the unigene set comprised hits with no functional characterization and include expressed proteins, unknown proteins, hypothetical proteins, putative proteins, and predicted proteins. Among the functionally classified unigenes, the transcription factor class constitutes the third highest category of functionally classified unigenes, the first two being that of cellular metabolism and protein synthesis (Table 2). Among the ESTs identified, 334 did not show any homology to rice dbEST or rice cDNAs, but were localized onto the rice genome sequence (IRGSP, 2005). These constitute 19% of the total N22 unigene set. These novel ESTs provide expression evidence for the *in silico* predicted genes and will assist in their intron and exon annotation. As the ESTs in this study were from a cDNA library constructed from drought stress, these novel ESTs may mainly represent genes involved in the drought-stress response. The N22 unigene set was mapped onto rice genomic sequences, and the number of unigenes mapped onto each chromosome is given in Table 3.

### *Identification of putative abiotic stress-responsive genes*

This additional coverage of the rice transcriptome with the drought-stressed leaf library resulted in the identification of potential stress-related genes. As these are from a normalized library constructed from drought-stressed seedling tissue, the profiles may provide clues in the identification of drought-stress responsive genes. The highly represented transcripts were further verified by annotation and comparison with those described in previous studies on the abiotic stress response in several plant species. Accordingly, the redundancy of the stress-responsive genes were considered for *in silico* northern analysis and expression profiles of these highly expressed genes are listed

**Table 1.** *Summary of EST generation and analysis*

| | |
|---|---|
| Total number of readable sequences obtained | 6694 |
| Vector sequences | 354 |
| Viral contaminants (Adenovirus type 2) | 380 |
| Highly redundant ribosomal RNA sequence | 224 |
| Sequences between 50–75 bp | 142 |
| Mean average read length (bp) | 483 |
| Number of high quality sequences deposited in GenBank | 5815 |
| Unigenes identified by CAP3 assembly | 2067 |
| Number of unigenes found with no significant homology to the finished rice genome sequence (library contaminants) | 390 |
| Number of rice unigenes | 1677 |
| Number of unigenes which have no expressional evidence in rice (novel unigenes) | 334 |

**Table 2.** *Functional classification of N22 unigene sequences*

| Category | Number of sequences (%) | Number of novel sequences(%) |
|---|---|---|
| Cellular metabolism | 229 (13.7) | 25 (7.5) |
| Cell structure | 51 (3.0) | 6 (1.8) |
| Detoxification | 56 (3.3) | 8 (2.4) |
| Hormone response | 17 (1.0) | 4 (1.2) |
| Heat shock proteins | 26 (1.5) | 1 (0.3) |
| Osmotic protectants | 38 (2.3) | 4 (1.2) |
| Protein kinases and phosphatases | 62 (3.7) | 8 (2.4) |
| Pathogen response | 31(1.9) | 3 (0.9) |
| Photosynthesis | 65 (3.9) | 10 (3.0) |
| Protein synthesis | 142 (8.5) | 20 (6.0) |
| Signal transduction | 49 (2.9) | 9 (2.7) |
| Transcription factors | 95 (5.7) | 15 (4.5) |
| Transport | 52 (3.1) | 3 (0.9) |
| Protein degradation | 40 (2.4) | 5 (1.5) |
| Secondary metabolism | 12 (0.78) | 1 (0.3) |
| Unknown and unclassified | 712 (42.5) | 212 (63.5) |
| Total | 1677 | 334 |

**Table 3.** *Distribution of unigene sequences in the rice genome*

| Chromosome | Number of contigs | Percentage |
|---|---|---|
| 1 | 258 | 15.4 |
| 2 | 209 | 12.5 |
| 3 | 233 | 13.9 |
| 4 | 147 | 8.8 |
| 5 | 138 | 8.2 |
| 6 | 138 | 8.2 |
| 7 | 118 | 7.0 |
| 8 | 112 | 6.7 |
| 9 | 84 | 5.0 |
| 10 | 78 | 4.7 |
| 11 | 74 | 4.4 |
| 12 | 88 | 5.3 |
| Total | 1677 | 100.0 |

in Table 4. Those ESTs that exhibit an abundance of 10 or more are considered here. Comparative *in silico* analysis of paralogues from multiple sources of rice (Matsumura *et al*., 1999; Kawasaki *et al*., 2001; Rabbani *et al*., 2003) and orthologues from other plants (Seki *et al*., 2001, 2002*a*; Kreps *et al*., 2002; Ozturk *et al*., 2002) led to the identification of 589 putative stress-responsive genes (SRGs). These are classified into 15 functional groups

**Table 4.** *Abundantly expressed stress-responsive genes in N22 seedlings*

| N22 EST accession | Full-length cDNA accession | Abundance | Putative function | Identical accession in GenBank |
|---|---|---|---|---|
| BI305796 | AK062796 | 101 | Rice metallothionein | AB002820 |
| BI306046 | AK061611 | 64 | Ribulose *bis*phosphate carboxylase, small subunit | L22155 |
| BI306560 | NF[a] | 60 | Heat shock protein 16.9C | L14444 |
| BI305614 | AK059196 | 55 | Thioredoxin *h* | D26547 |
| BI305481 | AK058313 | 49 | Metallothionein-like protein | AF001396 |
| BI305617 | AK058529 | 47 | Metallothionein-like protein type 2 | U57638 |
| BI305566 | AK106205 | 47 | Rd22 (dehydration-responsive protein) | D10703 |
| BI305843 | AK060920 | 41 | Triosephosphate isomerase | L04967 |
| BI306132 | NF | 31 | β-D-glucan exohydrolase | U46003 |
| BI306388 | NF | 29 | Jasmonate-induced protein | X98124 |
| BI306379 | AK104420 | 27 | Peroxidase | M73234 |
| BI305397 | AK058788 | 25 | Photosystem I PSI-K subunit | L12707 |
| CB964951 | AK065178 | 23 | Hypothetical protein | NM_127785 |
| BI306352 | AK070414 | 23 | Lipid transfer protein LPT IV | AF017361 |
| BI305945 | AK098931 | 20 | No hit | |
| CB966658 | AK109382 | 20 | Quinone oxidoreductase-like protein | NM_121703 |
| BI305557 | NF | 19 | Ubiquinol–cytochrome *c* reductase | X79275 |
| BI306687 | NF | 18 | Cys2/His2 zinc-finger protein | X60700 |
| BI306097 | AK104005 | 18 | Lipid transfer protein precursor | U29176 |
| BI305750 | AK065866 | 18 | Chitinase | AB027426 |
| CB967158 | AK058918 | 17 | Ribosomal protein l36e | AL132960 |
| BI306475 | AK070090 | 16 | Calmodulin 1 | AF042840 |
| BI305543 | AK073698 | 16 | Malate dehydrogenase, NAD-dependent | X78800 |
| BI305450 | AK105037 | 15 | Translation initiation factor SUI1 | AF094774 |
| BI305417 | AK100321 | 15 | Cytochrome P450 | AY072297 |
| BI305422 | AK104176 | 15 | Chlorophyll *a/b*-binding protein | U74295 |
| BI305391 | NF | 14 | 3-oxoacyl-(acyl-carrier-protein) reductase, putative | AJ243091 |
| BI306217 | AK058741 | 13 | Histone H4 | M12277 |
| BU673062 | NF | 13 | Expressed protein | NM_129142 |
| BI306338 | AK106979 | 13 | Hypothetical protein | AJ271079 |
| BI305440 | AK100908 | 13 | UDP-glucuronic acid decarboxylase | AB079064 |
| BI306390 | AK105055 | 12 | Photosystem II 10 kDa polypeptide psbr | U86018 |
| CB966380 | AK062463 | 12 | Lipid transfer protein | U88090 |
| BU673470 | NF | 12 | Dehydration stress-induced protein | AF314810 |
| BI305705 | AK061000 | 12 | Dof domain, zinc finger | AB028132 |
| BI305703 | U18404 | 11 | Metallothionein-like protein | U18404 |
| BI305752 | AK111242 | 11 | Glycosyl hydrolases family 16, putative | AF163820 |
| BI305253 | AK064960 | 11 | Glyceraldehyde-3-phosphate dehydrogenase, type I | AF251217 |
| CB966697 | AK062882 | 11 | AP2 domain transcription factor | NP_195167 |
| BI306264 | AK104987 | 11 | Glutamine synthetase | X14245 |
| BI305524 | AK070516 | 11 | Fructose-1,6-*bis*phosphatase | AB007193 |
| BI306389 | AK104912 | 11 | Xyloglucan endo-1,4-β-D-glucanase | X93175 |
| BI306059 | AK066834 | 10 | Osmyb1 transcription factor | D88617 |
| BI305933 | AK068686 | 10 | Cell division protein ftsh-like protein | NM_111112 |
| BI305835 | AK066933 | 10 | V-type $H^+$-translocating pyrophosphatase | D45384 |
| BU672803 | NF | 10 | Chitinase-B | AF402939 |
| BI305595 | AK064780 | 10 | Heat shock protein 82 | Z15018 |
| BI305683 | NF | 10 | Root-specific rcc3 | L27208 |
| BI306058 | AK065962 | 10 | Glutaredoxin | D86744 |
| BI305598 | AK068555 | 10 | Ribulose *bis*phosphate carboxylase, small subunit | AF017364 |
| BI306026 | AK104719 | 10 | Aldolase C-1 | D50307 |
| BU672976 | AK065027 | 10 | Disease resistance response protein | NM_123616 |
| BI305650 | AK065044 | 10 | Exoglucanase precursor | U46003 |
| BI305947 | AK072166 | 10 | Gigantea-like protein | AJ133787 |
| BU673030 | AK098982 | 10 | Ribosomal protein S31 | D38011 |
| CB965712 | AK067801 | 10 | Phenylalanine ammonia-lyase | Z15085 |
| BI306276 | AK066771 | 10 | Pathogenesis-related protein | U20347 |
| BI306443 | AK069446 | 10 | Catalase | D26484 |

[a] NF indicates no significant similarity found in full-length cDNAs of rice.

(Fig. 1). Interestingly, the distribution of the 589 putative stress-responsive ESTs among the functional categories showed that transcription factors were particularly well represented. The list of abiotic stress-responsive genes identified from our ESTs, along with the source for paralogues or orthologues from rice and other plants, respectively, is given in supplementary Table S1 at *JXB* online. All of the identified SRGs were mapped to the rice genomic sequence (IRGSP, 2005) (Table 5).

### *Digital northerns*

Apart from providing an efficient method for gene discovery, EST data sets can be used to provide low precision estimates of mRNA levels in a tissue through estimations of EST redundancy (Ohlrogge and Benning, 2000; Audic and Claverie, 1997). The EST library used in this study has relatively low redundancy because it was normalized (Reddy *et al*., 2002*a*), but still contains many more copies of some transcripts than others. The levels of redundancy among the contigs derived from the CAP3 assemblies have been studied. Of the 828 assembled sequences with more than one EST representation, the most highly represented transcripts were from metallothioneins, followed by transcripts involved in oxidative stress, novel genes, and expressed proteins with no known function. The *in silico* expression profiles are represented in Table 4.

### *Comparative analysis of expression profiles between leaf and panicle ESTs under drought stress*

Whether the identified stress-responsive genes also appear in other tissue under drought stress, they were compared with panicle ESTs of an IR64 (*indica*) library made from drought-stressed plants. Surprisingly, only 280 genes were found in common between the two libraries. Among these, 125 genes were identified as predicted stress-responsive genes (Table 6). Functional classification of common drought-responsive genes showed that a majority

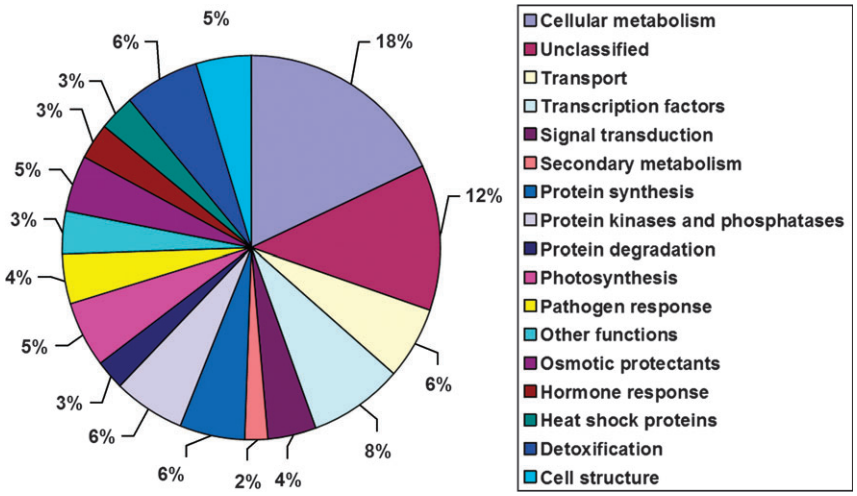


**Fig. 1.** Functional classification of 589 putative stress-responsive genes of rice.

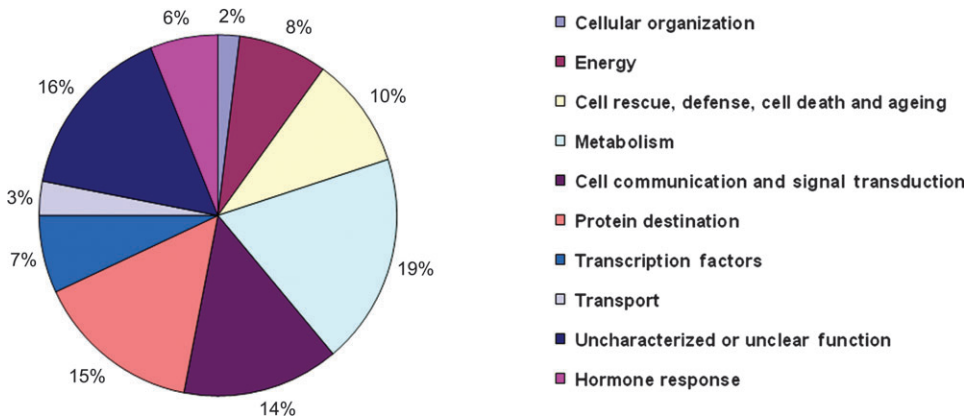


**Fig. 2.** Classification of 125 drought-stress responsive genes shared between leaf (N22) and panicle (IR64) tissues.

**Table 5.** *Distribution of putative stress-responsive genes (SRGs) in the rice genome*

| Chromosome | SRGs |
|---|---|
| 1 | 94 |
| 2 | 62 |
| 3 | 78 |
| 4 | 48 |
| 5 | 43 |
| 6 | 63 |
| 7 | 43 |
| 8 | 44 |
| 9 | 24 |
| 10 | 32 |
| 11 | 28 |
| 12 | 30 |
| Total | 589 |

of them (65%) are associated with metabolism, cellular communication and signal transduction, transcription factor, cellular defence, and protein destination categories (Fig. 2).

## Discussion

In this study the utility of an EST-based approach for gene discovery in rice has been demonstrated. Nagina 22, an *indica* rice cultivar, was chosen for EST generation and gene discovery, based on its phenology and the utility of this genotype in developing drought-tolerance lines. Nagina 22 is adapted for upland conditions and possesses a constellation of morphological and physiological characters such as early maturity, heat tolerance, two-point root system, accumulation and mobilization of carbohydrates, high regeneration and recovery processes, all associated with drought tolerance mechanisms in plants. The extensive EST resources from N22 were used in characterizing the N22 drought-stress transcriptome and in identification of drought-stress responsive genes. A classification of the unigene set revealed a significant number of novel genes with unknown functions. Since these are specific to the drought-induced *indica* library and are not represented in other stress libraries of rice, most of them presumably are stress-responsive genes. Molecular functional classification of 1677 unigenes shows a large number of genes that are predicted to be involved in signal transduction and transcriptional regulation (Table 2). Of the 1677 N22 unigenes, 81% showed homologous sequences to existing rice expressed genes and the remaining 19% have no expressional evidence for rice EST or cDNAs in databases. These 19% constitute novel rice genes which have been uncovered in this study. Analysis of the N22 unigene set revealed that 57% of them have a candidate functional role assigned and the remaining 43% belong to genes which have expressional evidence, but no functional role assigned. This suggests that there are many functionally unclassified genes that need to be characterized to discover new pathways and mechanisms adapted by plants to cope with drought stress.

Analysis of the N22 unigene set revealed many putative candidate genes for stress response that can be major targets for engineered stress tolerance. Among these are the genes encoding proteins that are associated with an osmotic stress response such as osmoprotectant synthesis (BU673697, BU673025), the dehydration stress-induced proteins (BU673123, BU672787), and the dehydration-responsive proteins like RD22 (BU672774). Data in Table 4 shows that the EST data revealed a number of genes associated with sugar metabolism and antioxidant pathways, as well as osmolyte synthesis. Of the two isoforms of glutathione-*S*-transferases (GST) (BU673645), one shows sequence similarity with *Zea mays* GST (AF244678) and the other, *OsGSTZ1*, to that of rice (AF309381). Evidence for a protective function of intracellular reactive oxygen species scavenging systems by glutathione *S*-transferase and glutathione peroxidase has been obtained from transgenic experiments in maize (Roxas *et al*., 1997). Homologues of these genes were identified through our Nagina 22 EST analysis, and thus provide both orthologues and paralogues that may have evolved during duplications and acquired a new functional role in the due course of evolution. Several Nagina 22 ESTs were identified from genes that encode enzymes which break down $H_2O_2$ to water; catalase (BU673091, BU673392), ascorbate peroxidase (APX) (BU673288) showing homology to tomato APX (A3251882), and manganese superoxide dismutase (MnSOD) (BU673715) which is a homologue of rice MnSOD (L34039) seem to provide tolerance to oxidative stress. The over-expression of MnSOD in chloroplasts conferred tobacco paraquat tolerance (Tsang *et al*., 1991). In a field study McKersie *et al*. (1996) reported that transgenic alfalfa expressing MnSOD suffered reduced injury from water-deficit stress.

The most abundant class of Nagina 22 drought-stressed transcripts represent a group of genes that encode metallothioneins and metallothionein-like proteins, which help in metal detoxification. These are low molecular weight, cysteine-rich, soluble, and metal-binding proteins found in both plant and animal tissues. These proteins sequester toxic metal ions. Seven groups of metallothioneins were found showing different levels of sequence similarity to rice metallothioneins (BU672908, BU672800, BU672917, BU673120, BU673768, BU672968, and BU672982). Rice metallothioneins expression is reported to be markedly increased under $H_2O_2$, heat shock, abscisic acid, and salicylic acid in shoots (Zhou *et al*., 2005), indicating their functional role during oxidative stress. Promoter analysis revealed heat-shock elements motifs, besides many light-responsive elements. Since the genotype under study is a heat-tolerant cultivar, these could be the reasons for high transcript abundance under drought stress. Further characterization of these classes of genes is needed to elucidate

**Table 6.** *Comparison of stress-responsive ESTs from drought-stressed N22 leaf and IR64 panicle libraries*

| N22 EST accession number (leaf) | Putative function | TIGR gene model | IR64 EST accession number (panicle) |
|---|---|---|---|
| | **Cellular metabolism** | | |
| BU673346 | Putative amine oxidase | 11670.t05414 | CA762096 |
| BI306458 | Ubiquinol–cytochrome *c* reductase | 11669.t05537 | CA766060 |
| BI305797 | Enolase | 11669.t01296 | CA759903 |
| BU672850 | Succinic semialdehyde dehydrogenase | 11668.t00676 | CA763533 |
| BI306457 | Cytosolic glyceraldehyde-3-phosphate dehydrogenase GAPDH | 11670.t03801 | CB096682 |
| CB964504 | Trehalose-6-phosphate synthase | 11682.t03982 | CA765766 |
| BI306315 | Similar to ATP-citrate-lyase | 11687.t04301 | CA767579 |
| CB967361 | Aldolase | 11667.t00188 | CA765072 |
| BI305193 | Respiratory burst oxidase homologue | 11687.t02970 | CA760620 |
| CB964609 | Biotin synthase | 11674.t04125 | CA762191 |
| BI306290 | $Ca^{2+}$ sensitive 3′(2′),5-diphosphonucleoside 3′(2′) phosphohydrolase | 11686.t00729 | CA759552 |
| CB965408 | Phosphoethanolamine methyltransferase | 11667.t04614 | CA762802 |
| CB964525 | Methionyl aminopeptidase-like protein | 11673.t02294 | CA759430 |
| BI305434 | Acyl-CoA:1-acylglycerol-3-phosphate acyltransferase | 11667.t05351 | CA762908 |
| BI305360 | Cytochrome P450 | 11667.t04030 | CA763743 |
| BI306288 | Putative copper amine oxidase | 11670.t03710 | CA759334 |
| BI305831 | Sucrose synthase | 11680.t00847 | CA761643 |
| BU673036 | Putative phospholipid cytidylyltransferase | 11687.t00205 | CA766827 |
| BI306060 | GF14-c protein | 11674.t03189 | CA760002 |
| BI305999 | Expressed protein | 11668.t04594 | CA760466 |
| | **Structural proteins** | | |
| CB967019 | Histone H3.2 protein | 11669.t02534 | CA761515 |
| BI306497 | Ubiquitin (mub1) gene | 11667.t02054 | CA763276 |
| BU673900 | Actin | 11669.t04750 | CA766273 |
| | **Defence** | | |
| BI306248 | Thioredoxin *h* | 11673.t00784 | CA763750 |
| BU673649 | Glutathione *S*-transferase OsGSTZ1 | 11686.t00973 | CA759451 |
| BU673288 | Ascorbate peroxidase (TL29) | 11670.t04787 | CA762991 |
| CB966179 | Phospholipid hydroperoxide glutathione peroxidase | 11669.t02241 | CA763032 |
| CB967248 | Glycolate oxidase | 11673.t03998 | CA764333 |
| BI306655 | Glyoxalase I | 11674.t00825 | CB096525 |
| BI306573 | Glutathione-dependent dehydroascorbate reductase precursor | 11680.t01165 | CA767174 |
| BU673645 | Glutathione *S*-transferase GST 13 | 11669.t00325 | CA766885 |
| | **Pathogen response** | | |
| CB965601 | Thaumatin-like protein | 11680.t04523 | CB097147 |
| BU673639 | Wound-inducive gene | 11674.t02539 | CB096245 |
| BU672887 | Sgt1 | 11667.t04013 | CA759388 |
| BI305746 | Cyclophilin CYP5 | 11680.t04712 | CA767313 |
| | **Hormone response** | | |
| BU673400 | Indole-3-glycerol phosphate synthase | 11670.t03633 | CA763617 |
| BI305739 | Abscisic acid- and stress-inducible protein (Asr1) | 11687.t00573 | CA759579 |
| BI306538 | 1-aminocyclopropane-1-carboxylate oxidase | 11673.t02534 | CA765577 |
| BU673190 | Putative IAA1 protein | 11669.t04977 | CA765289 |
| BI306117 | Elongation factor EF-2 | 11670.t00182 | CA759898 |
| BI305642 | Phytochrome-associated protein | 11667.t00846 | CA759750 |
| CB965518 | Auxin-induced protein | 11670.t02490 | CA760053 |
| | **Heat shock proteins and osmotic protectants** | | |
| BI306480 | High mobility group I/Y-2 | 11674.t03171 | CA761966 |
| BU673322 | Luminal binding protein 2 precursor (BiP2) | 11680.t01001 | CB097040 |
| BI306214 | Chaperonin 21 precursor | 11681.t02378 | CA764694 |
| BI305618 | GrpE protein | 11668.t03643 | CA764945 |
| BI306513 | Mitochondrial chaperonin-60 | 11676.t02764 | CA759723 |
| BI306548 | 16.9 kDa heat shock protein | 11667.t00335 | CA761072 |
| BI305213 | Heat stress transcription factor Spl7 | 11668.t02914 | CA760455 |
| BI306657 | Heat shock protein 82 | 11681.t02749 | CA760049 |
| BI306343 | Glycine-rich protein | 11680.t03864 | CB096773 |
| BI305248 | Dehydrin | 11687.t02337 | CA766722 |
| | **Protein degradation** | | |
| BI306554 | Serine carboxypeptidase | 11687.t02109 | CA764189 |
| BI305677 | Ubiquitin protein fused to a ribosomal protein | 11669.t01168 | CB097190 |
| | **Protein kinases and phosphatases** | | |
| BI306067 | OsCDPK7 | 11670.t04657 | CA765008 |
| BI306714 | Phosphoribulokinase | 11668.t04309 | CA764349 |
| BI305348 | Calcium-dependent protein kinase | 11673.t03565 | CA759704 |
| BI306130 | Protein kinase, putative | 11669.t01452 | CA762064 |
| BI305344 | Serine/threonine kinase | 11669.t01890 | CA762856 |

**Table 6.** (*Continued*)

| N22 EST accession number (leaf) | Putative function | TIGR gene model | IR64 EST accession number (panicle) |
|---|---|---|---|
| CB967004 | Protein kinase | 11682.t03978 | CA761897 |
| BU672858 | Mitogen-activated protein kinase homologue MMK2 | 11676.t03405 | CA763089 |
| BI305269 | MAP3K-β-1 protein kinase | 11669.t01408 | CA760582 |
| CB966430 | Mitogen-activated protein kinase | 11680.t00510 | CA760368 |
| BI305458 | Nucleoside diphosphate kinase | 11676.t03651 | CB096334 |
| BI305224 | Contains similarity to protein phosphatase-2c~gene | 11670.t03111 | CB096996 |
| CB964933 | Protein phosphatase 2C-like protein | 11669.t00343 | CA766893 |
| BI306327 | Protein phosphatase 2C-like protein | 11670.t05303 | CA765327 |
| | **Photosynthesis** | | |
| BI306021 | Chlorophyll *a*/*b*-binding protein | 11673.t03477 | CB097064 |
| BU673906 | Ribulose bisphosphate carboxylase/oxygenase | 11686.t01663 | CA767270 |
| BU673889 | Chloroplast apocytochrome b6 (petB) | 11667.t05409 | CA764607 |
| BI305598 | Small subunit of ribulose-1,5-*bis*phosphate carboxylase | 11686.t01843 | CB096380 |
| BU672866 | Putative chlorophyll synthase | 11676.t03688 | CA765915 |
| BI305564 | CP26, partial sequence | 11687.t01243 | CA760967 |
| BI305247 | Putative chloroplast RNA helicase VDL isoform 1 | 11667.t06958 | CB097044 |
| BI306736 | Photosystem II D1 protein | 11669.t01915 | CB096561 |
| BI305816 | Photosystem I chain IV precursor | 11673.t02296 | CA765338 |
| BI305763 | Triosephosphate isomerase (Rictipi2) gene | 11667.t00449 | CA763752 |
| | **Protein synthesis** | | |
| CB967287 | 40S subunit ribosomal protein | 11680.t00329 | CA760617 |
| BI306102 | EF-1α | 11669.t00701 | CA759893 |
| BI306120 | EREBP-like protein | 11669.t00750 | CA766852 |
| CB965835 | *S*-ribonuclease binding protein SBP1 | 11668.t00276 | CA759415 |
| BI306102 | EF-1α | 11669.t00701 | CA759893 |
| BU673172 | Elongation factor 1α | 11669.t00701 | CA759893 |
| CB964857 | No hit | 11687.t00576 | CB096917 |
| BI306632 | Ribosomal protein | 11673.t03378 | CB096625 |
| BU673302 | Translation initiation factor 4A | 11669.t03349 | CA763980 |
| CB967086 | RSZp22 splicing factor | 11680.t00786 | CA762148 |
| BU673172 | Elongation factor 1α | 11669.t00701 | CA759893 |
| | **Secondary metabolism** | | |
| BI306467 | Putative strictosidine synthase-like | 11669.t05057 | CA763632 |
| BI305578 | γ-tocopherol methyltransferase | 11668.t04338 | CA759409 |
| | **Signal transduction** | | |
| BI305552 | Small GTP-binding protein (Ran1) | 11667.t03912 | CA763744 |
| BU673756 | Signal recognition particle receptor α | 11667.t06848 | CA763094 |
| BI305605 | Vesicle soluble NSF attachment protein receptor | 11669.t02342 | CA759161 |
| BI305572 | Small GTP binding protein RACDP (RACD) | 11668.t05480 | CA763414 |
| BU673747 | Putative GTP-binding protein | 11667.t00746 | CA762975 |
| | **Transcription factors** | | |
| BU673061 | Zinc finger protein | 11670.t03862 | CA765791 |
| BI306209 | RING finger protein | 11682.t00608 | CA761322 |
| BU673870 | HOS59 | 11680.t04149 | CA760336 |
| BU672942 | Small zinc finger-like protein (TIM9) | 11686.t03641 | CA764474 |
| BI305867 | RING3-like bromodomain protein | 11680.t00364 | CA767235 |
| BI305994 | Similar to lipase | 11667.t02472 | CA764973 |
| BI306221 | Putative RING zinc finger protein | 11676.t01170 | CB096539 |
| BI306016 | Small nuclear ribonucleoprotein | 11668.t00244 | CA767531 |
| | **Transport** | | |
| BU673203 | ABC transporter family protein | 11668.t05262 | CB096918 |
| BI306386 | Vacuolar $H^+$-ATPase (vatp-P1) | 11687.t00590 | CA763376 |
| BU672768 | Major intrinsic protein | 11668.t01189 | CA760846 |
| BU673507 | Vacuolar membrane ATPase subunit G | 11670.t04784 | CA767335 |
| BI305935 | DNA binding protein, putative | 11670.t03474 | CA760475 |
| | **Unclassified** | | |
| BI305711 | Expressed protein | 11681.t02973 | CA765335 |
| CB967067 | Amino acid transporter family | 11680.t04035 | CA767198 |
| BU673532 | Unknown protein | 11674.t02230 | CA762677 |
| BI306384 | Expressed protein | 11670.t03940 | CA764882 |
| BI306017 | Unknown protein | 11670.t05378 | CA761712 |
| BI306104 | No hit | 11681.t02166 | CA761314 |
| CB966166 | Expressed protein | 11667.t03957 | CA765382 |
| BI306086 | Hypothetical protein | 11682.t03722 | CA767428 |
| BI306484 | Expressed protein | 11669.t02872 | CA766081 |
| BI306324 | Unknown protein | 11667.t02941 | CA767540 |
| CB967400 | Unknown protein | 11669.t06006 | CB096667 |

**Table 6.** (*Continued*)

| N22 EST accession number (leaf) | Putative function | TIGR gene model | IR64 EST accession number (panicle) |
|---|---|---|---|
| BU673135 | Unknown cold-induced protein | 11682.t04051 | CB096988 |
| BI306518 | Timing of CAB expression 1-like protein | 11681.t03083 | CA764459 |
| BI306163 | Early nodulin | 11668.t01140 | CA764447 |
| CB967012 | Pollen allergen-like protein | 11670.t03620 | CB096861 |
| BU673348 | Non-phototrophic hypocotyl 1b | 11670.t02192 | CA763992 |
| BI306519 | Putative pumilio/Mpt5 family RNA-binding protein | 11668.t05346 | CB096802 |
| BI306073 | Wound induced protein homologue | 11667.t00305 | CB096984 |

their role in the drought-stress response in rice. The other detoxifying proteins include thioredoxin (BU673762) showing homology to that of rice (AB053294), and the other showing homology to a gene in *Arabidopsis* (AY085055).

The stress-responsive gene sets also include those associated with water channels and transporters such as aquaporin (BU673363), an ABC transporter protein (BU673203), and an oligopeptide transporter protein (BU673275). The recently discovered aquaporins act as water channels and their transcript levels are shown to be influenced significantly by a wide variety of environmental stimuli (Weig *et al*., 1997). These are reported to be involved in water uptake and may function in metabolite or ion transport. These transport proteins are reported to show a 5-fold up-regulation under stress (Seki *et al*., 2002*a*).

Other important genes uncovered in Nagina 22 drought-stress ESTs include the membrane-stabilizing proteins and late embryogenic abundant proteins which enhance water-binding capacity, creating a protective environment for other proteins or structures, referred as dehydrins (BI305248). They play a major role in the sequestration of ions that are concentrated during cellular dehydration. Numerous genes involved in membrane stability and thermotolerance have been identified from the present EST collections. These include heat shock proteins (HSPs), which have been widely hypothesized to be a major factor in cell thermotolerance (Howarth and Ougham, 1993) and tolerance to other environmental assaults such as oxidative, chilling, salt, and heavy metal stresses. HSPs were also shown to regulate expression of other stress-inducible genes (Liu and Thiele, 1996).

Another group of genes uncovered include those encoding proteins involved in signal transduction and the regulation of gene expression. It is probable that these play a regulatory role in the plant stress response. These include protein kinases, protein phosphatases, transcription factors, and enzymes in phospholipid metabolism and other signaling molecules such as calmodulin-binding protein. Many kinases were observed in the collection (see supplementary Table S1 at *JXB* online), including mitogen activated protein kinases (MAPKs) (BU672858, BI305201), calcium-dependent protein kinase (BU673731), adenosine kinases, and adenylate kinases (BU673745, BU672936). In addition, the signalling molecule calmodulin (BU673090, BU672925, BU673775), a common participant in the MAPK signal transduction cascade, was found in the Nagina 22 EST libraries studied. The present EST analysis also revealed many more candidate signalling genes, such as MAP kinases and various transcription factors.

The identified transcription factors include proteins having typical DNA binding motifs such as bZIP, MYB, MYC, EREBP/AP2, and ZINC fingers. The role of various transcription factors in stress-responsive gene regulation has been investigated in plants, and several target genes and pathways have been identified (Thomashow, 1998; Park *et al*., 2001; Seki *et al*., 2001; Singh *et al*., 2002; Shinozaki *et al*., 2003). Overall, the normalized library proved to be a rich source of stress-responsive rice genes.

The EST data and analysis presented here are a first global overview of the transcripts that are expressed in *indica* rice under water stress. The annotation and comparative analysis of these ESTs have identified many genes associated with or having a potential role in drought-stress tolerance. These genes provide a starting point for understanding the nature of molecular mechanisms of a plant's response and tolerance to drought. EST analysis has uncovered numerous novel genes and transcriptional activators, the master switches that influence the expression of cascades of genes associated with a stress response.

Comparative analysis of SRGs of N22 (the present study) with IR64 panicle ESTs generated under drought-stress revealed that 125 (40%) of them are in common, demonstrating similar patterns of regulated pattern of gene expression between leaf and panicle tissues (Table 6). This pattern is largely similar to the one reported earlier (Tang *et al*., 2005). These genes are presumably associated with drought-stress response and tolerance during different growth stages of the rice plant. The remaining 60% of SRGs uncovered in this library could be genotype-specific or tissue-specific. However, whether these genes are actually genes involved in rice drought-tolerance cannot be definitely determined without further expression profiling, allele mining, QTL mapping, and reverse genetic experiments.

Identification of the genes in the rice genome has relied heavily on non-experimental methods such as *ab initio* gene prediction, sequence homology, and motif analysis. These efforts were limited by the insufficient ability of current gene-finding programs to identify and annotate genes from complex genomes effectively (Guigo *et al*., 2000; Mathe *et al*., 2002; Zhang *et al*., 2002; Bennetzen *et al*., 2004). So far, the identification of coding regions on a genome scale in rice has focused on EST and full-length cDNA analyses (Kikuchi *et al*., 2003). However, the available EST and cDNA resources do not comprehensively reveal all the genomic coding information as they are biased mostly toward highly expressed genes. Not surprisingly, exhaustive efforts to uncover the rice transcriptome have represented less than half of the predicted genes (Feng *et al*., 2002; Reddy *et al*., 2002; Sasaki *et al*., 2002; Yu *et al*., 2002; Markandeya *et al*., 2003, 2005; Zhao *et al*., 2004). Our EST data has aided in providing expression evidence for an additional 334 unigenes. Most EST sequencing projects have proven to be expensive due to high clone redundancy (Reddy *et al*., 2002*a*). In particular, transcript profiling under drought stress had not been carried out much in rice until our study to identify the drought transcriptome through large-scale EST generation. The EST resources of N22 have been found to be useful in the generation of high-density physical maps of stress-responsive genes in rice (Markandeya *et al*., 2005), to develop candidate gene molecular markers across selected cereals (Sivarama Prasad, 2005), including EST-PCRs (Chandrasekhar, 2005), and to identify SNPs (Lachagari *et al*., unpublished data). Further these ESTs are now being used as target probes in the fabrication of cDNA microarrays for expression profiling studies under field drought stress.

## Supplementary data

Supplementary data are available at *JXB* online.

## Acknowledgements

This work was financially supported by the Rockefeller Foundation, USA, through a grant to ARR. We thank the Department of Biotechnology, India and the Council of Scientific and Industrial Research, India for fellowships to MG, VBRL, and AMMR.

## References

**Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSIBLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25,** 3389–3402.

**Audic S, Claverie JM.** 1997. The significance of digital gene expression profiles. *Genome Research* **7,** 986–995.

**Babu PR, Sekhar AC, Ithal N, Markandeya G, Reddy AR.** 2002. Annotation and BAC/PAC localization of non-redundant ESTs from drought-stressed seedlings of an *indica* rice. *Journal of Genetics* **81,** 25–44.

**Bennett J, Arumugam K, Lafitte R, Wen J, Rudd S, Bruskiewich RM.** 2002. IRRI Drought Stress Panicle cDNA Library. www.ncbi.nlm.nih.gov/dbest.

**Bennetzen JL, Coleman C, Liu R, Ma J, Ramakrishna W.** 2004. Consistent over-estimation of gene number in complex plant genomes. *Current Opinion in Plant Biology* **7,** 732–736.

**Boyer JS.** 1982. Plant productivity and environment. *Science* **218,** 443–448.

**Bray E, Bailey-Serres J, Weretilnyk E.** 2000. Responses to abiotic stresses. In: Buchanan B, Gruissem W, Rockville JR, eds. *Biochemistry and molecular biology of plants*. American Society of Plant Biologists, 1158–1203.

**Chandrasekhar A.** 2005. Molecular mapping using SSRs and EST-PCR based markers, and identification of quantitative trait loci (QTLs) for drought tolerance in a doubled haploid population of rice (*Oryza sativa* L.). PhD thesis, University of Hyderabad.

**Dey MM, Upadhaya HK.** 1996. Yield loss due to drought, cold, and submergence in Asia. In: Evenson RE, Herdt RW, Hossain M, eds. *Rice research in Asia, progress and priorities*. Cary, NC: Oxford University Press, 231–242.

**Echenique V, Stamova B, Wolters P, Lazo G, Carollo V, Dubcovsky J.** 2002. Frequencies of Ty1-copia and Ty3-gypsy retroelements within the Triticeae EST databases. *Theoretical and Applied Genetics* **104,** 840–844.

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

**Ewing RM, Ben Kahla A, Poirot O, Lopez F, Audic S, Claverie JM.** 1999. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Research* **9,** 950–959.

**Feng Q, Zhang Y, Hao P, *et al.*** 2002. Sequence and analysis of rice chromosome 4. *Nature* **420,** 316–320.

**Fernandes J, Brendel V, Gai X, *et al.*** 2002. Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and microarray hybridization. *Plant Physiology* **128,** 896–910.

**Gene Ontology Consortium.** 2001. Creating the gene ontology resource: design and implementation. *Genome Research* **11,** 1425–1433.

**Gibbings JG, Cook BP, Dufault MR, Madden SL, Khuri S, Turnbull CJ, Dunwell JM.** 2003. Global transcript analysis of rice leaf and seed using SAGE technology. *Plant Biotechnology Journal* **1,** 271–285.

**Goff SA, Ricke D, Lan TH, *et al.*** 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296,** 92–100.

**Gotoh O.** 1982. An improved algorithm for matching biological sequences. *Journal of Molecular Biology* **162,** 705–708.

**Gowda M, Jantasuriyarat C, Dean RA, Wang GL.** 2004. Robust-Long SAGE (RL-SAGE): a substantially improved Long SAGE method for gene discovery and transcriptome analysis. *Plant Physiology* **134,** 890–897.

**Guigo R, Agarwal P, Abril JF, Burset M, Fickett JW.** 2000. An assesment of gene prediction accuracy in large DNA sequences. *Genome Research* **10,** 1631–1642.

**Howarth CJ, Ougham HJ.** 1993. Gene expression under temperature stresses. *New Phytologist* **125,** 1–26.

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

**International Rice Genome Sequencing Project.** 2005. The map-based sequence of the rice. *Nature* **436,** 793–800.

**Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert HJ.** 2001. Gene expression profiles during the initial phase of salt stress in rice. *The Plant Cell* **13,** 889–905.

**Khush GS.** 2005. What it will take to Feed 5.0 Billion Rice consumers in 2030. *Plant Molecular Biology* **59,** 1–6.

**Kikuchi S, Satoh K, Nagata T, *et al.*** 2003. Collection, mapping and annotation of over 28 000 cDNA clones from japonica rice. *Science* **301,** 376–379.

**Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF.** 2002. Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiology* **230,** 2129–2141.

**Liu XD, Thiele DJ.** 1996. Oxidative stress induced heat shock factor phosphorylation and HSF-dependent activation of yeast metallothionein gene transcription. *Genes and Development* **10,** 592–603.

**Ma J, Bennetzen JL.** 2004. Rapid recent growth and development of rice nuclear genomes. *Proceedings of the National Academy of Sciences, USA* **101,** 12404–12410.

**Markandeya G, Babu PR, Lachagari VBR, Feltus FA, Paterson AH, Reddy AR.** 2005. Functional genomics of drought-stress response in rice: transcript mapping of annotated unigenes of an *indica* rice (*Oryza sativa* L. cv. Nagina 22). *Current Science* **89,** 496–514.

**Markandeya G, *et al.*** 2003. ESTs from a normalized cDNA library of drought-stressed rice seedlings (Oryza sativa L. cv. Nagina 22). www.ncbi.nlm.nih.gov/dbest

**Mathe C, Sagot MF, Schiex T, Rouze P.** 2002. Current methods of gene prediction, their strengths and weaknesses. *Nucleic Acids Research* **30,** 4103–4117.

**Matsumura H, Nirasawa S, Kiba A, Urasaki N, Saitoh H, Ito M, Kawai-Yamada M, Uchimiya H, Terauchi R.** 2003. Overexpression of Bax inhibitor suppresses the fungal elicitor-induced cell death in rice (*Oryza sativa* L.) cells. *The Plant Journal* **33,** 425–434.

**Matsumura H, Nirasawa S, Terauchi R.** 1999. Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *The Plant Journal* **20,** 719–726.

**McKersie BD, Bowley SR, Harjanto E, Leprince O.** 1996. Water-deficit tolerance and field performance of transgenic alfalfa overexpressing superoxide dismutase. *Plant Physiology* **111,** 1177–1181.

**Michalek W, Weschke W, Pleissner KP, Graner A.** 2002. EST analysis in barley defines a unigene set comprising 4000 genes. *Theoretical and Applied Genetics* **104,** 97–103.

**Ohlrogge J, Benning C.** 2000. Unraveling plant metabolism by EST analysis. *Current Opinion in Plant Biology* **3,** 224–228.

**Ozturk ZN, Talame V, Deyholos M, Michalowski CB, Galbraith DW, Gozukirmizi N, Tuberosa R, Bohnert HJ.** 2002. Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Molecular Biology* **48,** 551–573.

**Park JM, Park CJ, Lee SB, Ham BK, Shin R, Paek KH.** 2001. Overexpression of the tobacco *Tsi1* gene encoding an EREBP/AP2-type transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. *The Plant Cell* **13,** 1035–46.

**Rabbani MA, Maruyama K, Abe H, *et al.*** 2003. Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiology* **133,** 1755–1767.

**Reddy AR, *et al.*** 2002*b*. Novel EST enrichment with normalized cDNA libraries from drought-stressed rice (*Oryza sativa* L. cv. Nagina 22). www.ncbi.nlm.nih.gov/dbest.

**Reddy AR, Ramakrishna W, Sekhar AC, Ithal N, Babu PR, Bonaldo MF, Soares MB, Bennetzen JL.** 2002*a*. Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. *indica* cv. Nagina 22). *Genome* **45,** 204–211.

**Roxas VP, Smith Jr RK, Allen ER, Allen RD.** 1997. Over-expression of glutathione *S*-transferase/glutathione peroxidase enhances the growth of transgenic tobacco seedlings during stress. *Nature Biotechnology* **15,** 988–991.

**Sasaki T, Matsumoto T, Yamamoto K, *et al.*** 2002. The genome sequence and structure of rice chromosome 1. *Nature* **420,** 312–316.

**Sasaki T, Song J, Koga-Ban Y, *et al.*** 1994. Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. *The Plant Journal* **6,** 615–624.

**Seki M, Ishida J, Narusaka M, *et al.*** 2002*b*. Monitoring the expression pattern of around 7000 *Arabidopsis* genes under ABA treatments using a full-length cDNA microarray. *Functional Integrative Genomics* **2,** 282–291.

**Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y, Shinozaki K.** 2001. Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. *The Plant Cell* **13,** 61–72.

**Seki M, Narusaka M, Ishida J, *et al.*** 2002*a*. Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *The Plant Journal* **31,** 279–292.

**Shinozaki K, Yamaguchi-Shinozaki K, Seki M.** 2003. Regulatory network of gene expression in the drought and cold stress responses. *Current Opinion in Plant Biology* **6,** 410–417.

**Singh KB, Foley RC, Sánchez LO.** 2002. Transcription factors in plant defence and stress responses. *Current Opinion in Plant Biology* **5,** 430–436.

**Sivarama Prasad L.** 2005. Comparative genetic mapping in rice (*Oryza sativa* L.) and sorghum (*Sorghum bicolor* L. Moench) using rice ESTs as RFLPs. PhD thesis, University of Hyderabad.

**Smith TF, Waterman MS.** 1981. Identification of common molecular subsequences. *Journal of Molecular Biology* **147,** 195–197.

**Tang J, Xia H, Li D, *et al.*** 2005. Gene expression profiling in rice young panicle and vegetative organs and identification of panicle-specific genes through known gene functions. *Molecular Genetics and Genomics* **274,** DOI 10.1007/s00438-005-0043-2.

**The Rice Full-Length cDNA Consortium.** 2003. Collection, mapping, and annotation of over 28 000 cDNA clones from japonica rice. *Science* **301,** 376–379.

**Thomashow MF.** 1998. Role of cold-responsive genes in plant freezing tolerance. *Plant Physiology* **118,** 1–8.

**Tsang EW, Bowler C, Herouart D, Van Camp W, Villarroel R, Genetello C, Montagu VM, Inzé D.** 1991. Differential regulation of superoxide dismutases in plants exposed to environmental stress. *The Plant Cell* **3,** 783–792.

**Uchimiya H, Kidou S, Shimazaki T, *et al.*** 1992. Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.). *The Plant Journal* **2,** 1005–1009.

**Umeda M, Hara C, Matsubayashi Y, Li H, Liu Q, Tadokoro F, Aotsuka S, 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 Molecular Biology* **25,** 469–478.

**Weig A, Deswarte C, Chrispeels MJ.** 1997. The major intrinsic protein family of *Arabidopsis* has 23 members that form three distinct groups with functional aquaporins in each group. *Plant Physiology* **114,** 1347–57.

**Yamamoto K, Sasaki T.** 1997. Large-scale EST sequencing in rice. *Plant Molecular Biology* **35,** 135–144.

**Yu J, Hu S, Wang J, *et al*.** 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296,** 79–92.

**Yu J, Wang J, LinW, *et al*.** 2005. The genomes of *Oryza sativa*: a history of duplications. *PLoS Biology* **3,** e38.

**Zhang J, Feng Q, Jin C, Qiu D, Zhang L, Xie K, Yuan D, Han B, Zhang Q, Wang S.** 2005. Features of the expressed sequences revealed by a large-scale analysis of ESTs from a normalized cDNA library of the elite *indica* rice cultivar Minghui 63. *The Plant Journal* **42,** 772–780.

**Zhang K, Deng M, Chen T, Waterman MS, Sun F.** 2002. A dynamic programming algorithm for haplotype block partitioning. *Proceedings of the National Academy of Sciences, USA* **99,** 7335–7339.

**Zhao W, Wang J, He X, *et al*.** 2004. BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Research* **32,** 377–382.

**Zhou GK, Xu YF, Liu JY.** 2005. Characterization of a rice class II metallothionein gene: tissue expression patterns and induction in response to abiotic factors. *Journal of Plant Physiology* **162,** 686–696.