5. DISCUSSION
5. Discussion

In this study, we have demonstrated the utility of an Expressed Sequence Tags (ESTs) based approach for gene discovery in rice. ESTs represent by far the largest constituent of DNA repositories in terms of sequence number and total nucleotide count in eukaryotes. These EST resources are exploited in genome annotation, gene expression analysis, gene discovery, and comparative genomic analysis. ESTs generated from a tissue specific library constructed during a defined developmental stage and, environmental conditions have been used to obtain gene expression profiles of that particular tissue (Rudd., 2003; Okubo et al., 1992). We have chosen Nagina 22, an indica rice cultivar, for EST generation and gene discovery, based on its phenology and the utility of this genotype in developing drought tolerant 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 its drought stress responsive transcriptome and in identification of stress responsive genes (SRGs) by in-silico and by cDNA microarray analyses. The data presented here gives a clear description of the utility of ESTs in gene discovery in chromosomal localization of candidate stress responsive genes which helped in generation of high density physical maps of drought transcriptome. Further, these ESTs constituted as primary substrate probes for cDNA microarray fabrication for expression profiling under field drought stress conditions. Our results highlight the utility of this approach in molecular dissection of genetic determinants for drought tolerance in rice.

The discussion is organized into two broad areas: technology improvement and utility, and identification and characterization of stress responsive genes in rice. We have considerably elaborated technology issues below since this is probably the first report on field drought stress array in indica rice and introduction of several new modifications of the primary array technology for a better resolution, precision and quality of data.
5.1 Efficiency of EST generation

The generation of high quality N22 ESTs was achieved by experimental optimization for high-throughput sequencing. Most EST sequencing projects have proven to be expensive due to a high clone redundancy (Reddy et al., 2002a). Normalization enriches cDNAs from relatively low-copy transcripts and should maximize the number of unique ESTs identified by random sequencing (Soares et al., 1994; Smith et al., 2001). Particularly, transcript profiling under drought stress were not carried out much in rice until we began our study to identify drought transcriptome through large scale EST generation from a normalized library. In the present study, the use of this normalized library which greatly reduced the levels of redundancy helped us in cost effective sequencing (Reddy et al., 2002). Subsequently we made available a large set of high quality and low redundant N22 ESTs through depositing in GenBank for access to rice research community. The major problem associated with most of the EST sequencing projects is low sequence quality obtained, since every individual cDNA clone is sequenced only once (single pass sequencing). Low sequence quality has several implications, importantly poor sequence quality can make detecting the overlap between EST sequences difficult, and thereby reducing both the readable sequence length of the assembled contigs, and the number of clones for which gene identity can be determined, since these two variables are correlated (www.phrap.org).

Furthermore, low sequence quality may lead to erroneous estimates of gene family sizes; especially if there are blocks of high and low quality stretches in ESTs. Finally, low sequence quality may also lead to an inability to determine homology with known sequences in the database. In addition to being a primary source of gene catalogues, high quality EST sequences are useful for SNP discoveries (Picoult-Newberg et al., 1999; Deutsch et al., 2001; Kota et al., 2003). However, a major difficulty with this approach is that the EST sequences contain sequencing errors, which have to be distinguished from true polymorphisms. Our optimized protocols for effective 3’ sequencing resulted in overcoming the above inherent limitations, with an average sequence read length of 568 bp (Fig. 2b, Table 1). The sequencing strategy employed in the present study successfully generated high quality sequences with an average read length of 568bp, around 400bp with Phred score of > 40, and about 500bp with a Phred
score of > 30 (Fig. 2a). These high quality sequences are found to be very useful in SNP discovery in a subsequent separate study (Reddy et al., data not shown).

Apart from the above reasons for incorrect nucleotides within EST sequences, there is a background of partially or completely incorrect sequence. Partially incorrect sequences contain stretches of vector or poly-linker sequence which we have cleaned manually. Completely incorrect sequences represent xeno-contaminants which are due to organellar or other types of DNA contaminations. In EST sequencing, cDNA library is constructed using mRNA as starting material from field harvested samples. However, we found relatively a few contaminants from cDNA library preparation which might have occurred during cloning stage, with organelle, viral or bacterial RNA, or DNA. This kind of contamination with bacterial, viral and other pathogens normally occur when the seedling are grown in greenhouse and field conditions. These types of contaminations may also cause problems in the assembly process, which in turn can lead to discrepancies in annotation and classification. To overcome this we have queried all our EST sequences against organellar and common vector and linker sequences to exclude these contaminating sequences. Upon analyzing the sequences we found few repetitive contaminants which we screened and used the sequences to clean up in all the batches of ESTs generated. In our case, 380 ESTs were removed prior to the submission of the 5815 sequences because they were viral sequences from Adenoviral type 2 encoding minor capsid protein VI (Table 1). Similarly, microbial origin sequence contaminants were also excluded from the analysis. Microbial contamination is an unavoidable outcome of EST analysis on greenhouse and 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.

In order to identify genes of other origin we have aligned and compared all the ESTs generated to the almost finished rice genome sequence. Upon homology searches for all the 5815 ESTs generated, 390 of which had 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), it
is likely that most or all of these ESTs are from microbial contaminants originated during RNA sampling.

5.2 Unigenes representation in N22 Library

In order to uncover the unique transcripts represented in our library, the ESTs generated in the present study were assembled using CAP3 assembly program (Huang and Madan, 1999. The assembly of these ESTs through CAP3 revealed a unigene set of 2,067 sequences presumably representing unique transcripts in N22. This included 1239 singlets and 828 contigs. Of the 2067, 390 unigenes did not show homology to rice genome, these 390 were omitted from further analysis and the resulting 1677 were identified as a unigene set derived from N22. We looked for similar unigenes in the EST and rice cDNA databases, and found that only 1343 N22 unigenes could be retrieved, and the remaining 334 did not show any similarity with the known rice EST or cDNA. These novel genes are described in a later section.

5.3 Novel genes (334) were identified from N22 unigene set

The Rice genome sequence (Feng et al., 2002; Goff et al., 2002; Sasaki et al., 2002; Yu et al., 2002; The Rice Chromosome 10 Sequencing Consortium, 2003; IRGSP The International Rice Genome Sequencing Initiative 2005), was expected to provide the gene space. However, identification of genes in the rice genome relied rather heavily on non-experimental methods such as ab initio gene prediction, sequence homology and motif analysis, which are limited by the insufficient ability of current gene-finding programs to effectively identify and annotate genes from complex genomes (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 reveal all the genomic coding information as they are biased mostly towards highly expressed genes. Not surprisingly, exhaustive efforts to uncover the rice transcriptome have only represented less than half of the predicted genes (Feng et al., 2002; Sasaki et al., 2002; Reddy et al., 2002; Markandeya et al., 2003, 2005; Yu et al., 2002; Zhao et al., 2004). The present study has identified 334 unigenes, which are novel rice genes uncovered in our library. Though these 334 unigenes have no homology to the
existing rice EST and cDNA databases, but were aligned onto rice genome sequence. These ESTs thus aided in accurate genome annotation for these genes by providing expressional evidence, which were earlier clubbed with hypothetical or predicted proteins. Annotation and comparative sequence analysis revealed 589 stress responsive genes in N22 unigene set.

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 showed 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 the known 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 (Table-2). 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.

The N22 unigenes were functionally classified based on different cellular processes (Bevan et al., 1998). 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, 2002a; Kreps et al., 2002; Ozturk et al., 2002) led to the identification of 589 putative stress responsive genes (SRGs) among the N22 unigenes. Interestingly the distribution of the 589 putative stress responsive ESTs among the functional categories (Fig. 4) showed that transcription factors were particularly well represented. Other well represented categories include proteins with known function in cellular defenses against abiotic and biotic stresses, and proteins involved in signaling and protein synthesis. This is in agreement with the earlier reported that transcription activators play an important role in stress response associated changes in gene expression (Chen et al., 2002).
5.4 Abundantly expressed genes and gene families under stress have been identified

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., 2002a), but still contains many more copies of some transcripts than others. 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 redundancies of the stress responsive genes were considered for in-silico northern analysis and expression profiles of these highly expressed genes are listed in Table 4. We studied the levels of redundancy among the contigs derived from the CAP3 assemblies. Of the 828 assembled sequences with more than one EST representation, the most highly represented transcripts were from metallothionein genes and gene families, followed by those genes involved in oxidative stress responses, novel genes and expressed proteins with no known function.

5.5 Analysis of N22 drought stress transcriptome

Drought stress is a complex trait, which is governed by many dispersed genes across the genome with a multitude of interconnected pathways and processes. We have catalogued and categorized genetically complex drought stress response-associated genes through EST analysis as our first step towards understanding molecular and cellular basis of stress response in rice. The stress response associated gene products are thought to protect, either directly or indirectly, against a variety of environmental stresses. Identification of specific genes with role in drought response function will help in developing transgenics with improved tolerance to drought stress. Genetic engineering of tolerance traits in crops through introduction of a small number of genes seems to be more attractive and rapid approach in improving stress tolerance (Cushman and Bohnert 2000). The success of these approaches has generally been limited by a lack of understanding of genes controlling metabolic flux, compartmentation, and response function (Holmberg, 1998; Shinozaki K, 1999).
The products of water stress induced genes can be classified into two groups (Shinozaki and Yamaguchi-Shinozaki, 1997). The first group includes the proteins that are directly involved in a function of stress tolerance. These include the genes which express to regulate osmotic adjustments inside a cell which is otherwise disturbed due to water loss. Basically these are genes involved in the synthesis and accumulation of osmolytes without disturbing the cellular functions (Ishitani et al., 1997). These encode enzymes required for the biosynthesis of various osmoprotectant molecules (sugars, Proline, Glycine-betaine etc.), and proteins that protect macromolecules and membranes (for e.g., LEA proteins, osmotin, antifreeze protein, chaperon, mRNA binding proteins etc.) The late embryonic abundant proteins such as dehydrins are abundantly expressed under stress are shown to protect cellular machinery (Lisse et al., 1996). Analysis of the N22 unigene set revealed many putative candidate genes, encoding enzymes required for biosynthesis of various osmoprotectants. Among these are the genes encoding proteins that are associated with osmotic stress response such as genes involved in osmoprotectant synthesis (BU673697, BU673025), the dehydration stress induced proteins (BU673123, BU672787) and the dehydration responsive proteins like RD22 (BU672774) as shown in table 5.

Data in Table 5 show a number of genes associated with sugar metabolism and antioxidant pathways, as well as osmolyte synthesis. Other important genes uncovered among Nagina 22 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 sequestration of ions that are concentrated during cellular dehydration. Numerous genes involved in membrane stability and thermo tolerance 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 thermo tolerance (Howarth and Ougham, 1993) and tolerance to other environmental assaults such as oxidative, chilling, high salt and heavy metal stresses. HSPs were also shown to regulate expression of other stress inducible genes (Liu and Thiele, 1996).

Aquaporins which are trans-membrane water channel proteins involved in regulation of water movement across membranes play an important role in
osmoregulation and avoidance of water deficit (Fray et al., 1994; Ruiter et al., 1997). 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 five fold up-regulation under stress (Seki et al., 2002a). Among the SRGs identified our N22 unigene set are many genes associated with water channels and transporters such as aquaporin (BU673363), an ABC transporter protein (BU673203) and an oligopeptide transporter protein (BU673275). It has been proved that as the leaves regain turgor, transcript level of this water channel protein increased in sea water treated M.crystallium (Yamada et al., 1995) indicating a role in recovery process.

The other genes among this category are the genes associated with detoxification, genes encoding detoxifying enzymes are also among this group (Ingram and Bartels., 1996). These include GSTs, catalases, super oxide dismutase, ascorbate peroxidase and a few others. Further, chaperonins and proteinases that may destroy inactive proteins (Williams et al., 1994) and enzymes involved in ATP production pathways (Riccardi et al 1998) also fall in this category. In plants, the inevitable production of reactive oxygen species (ROS) under any stress leads to singlet oxygen, superoxide, H₂O₂ and hydroxyl radicals. The mechanisms through which ROS detoxification occurs include both enzymatic and non-enzymatic., It was reported widely that stress increases ROS levels followed by up or down regulation of mRNA transcripts and protein levels and presumably leading to an accelerated turnover of components of detoxification systems which inturn exert a positive effect on plant performance (Noctor G, Foyer., 1998, McKersie BD., 1996, Van Camp., 1996, Roxasur., 1997). The genes in these groups usually either encode enzymes involved in removing toxic free radicals or proteins that directly mediate detoxification of toxic substances. The genes encoding detoxifying enzymes in our unigene set include GSTs, (two isoforms of glutathione-s-transferase BU673645), one showing 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
Homologues of these genes were identified in Nagina 22 EST collections and thus provide evidence for both these orthologues and paralogues might have evolved via duplications and acquired a new functional role in the due course of evolution. This mode of evolution, that is, duplication divergence of genes among higher plant is well demonstrated. Several ESTs were identified for genes that encode enzymes which break down H₂O₂ to water: catalase (BU673091, BU673392), ascorbate peroxidase (APX) (BU673288) showing homology to tomato APX (A3251882) and manganese superoxide dismutase (MnSOD) (BU673715) which is an orthologue of rice MnSOD (L34039) thought to provide tolerance to oxidative stress. The over expression of MnSOD in chloroplast conferred tobacco with 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 represents a group of genes that encode metallothioneins and metallothionein-like proteins, which are reported to be associated with metal detoxification. These are low molecular weight, cystein rich, soluble and metal-binding proteins found in both plant and animal tissues. These proteins sequester toxic metal ions, though precise mechanisms by which they accomplish is largely unknown. We found 7 groups or families of metallothioneins 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 H₂O₂, heat shock, abscisic acid and salicylic acid in shoots (Zhou et al., 2005) indicating their functional role during oxidative stress. Metallothionien promoter analysis revealed well known heat shock element motifs, besides many light responsive elements. Interestingly, N22 is a heat tolerant cultivar. This is one explanation for the accumulation of abundant transcripts of these genes under stress in rice tissue. Further characterization of these classes of genes is needed to elucidate their role in the drought stress response in rice. The other detoxifying proteins include thioredoxin (BU673762) showing sequence similarity to that of rice (AB053294) and the other to an orthologue of Arabidopsis (AY085055).

The second group consists of proteins involved in signal transduction and regulation of stress responsive gene expression. The transcripts of genes encoding several
of these proteins are shown to accumulate under drought conditions. The role of these gene products has been reviewed extensively (Shinozaki and Yamaguchi-Shinozaki, 2000). These include protein kinases, protein phosphatases, transcriptional factors, and enzymes in phospholipid metabolism, and other signaling molecules such as calmodulin-binding protein. Many kinases have been identified in the present study (Table 5); mitogen activated protein kinases (MAPKs) (BU672858, BI305201), calcium dependant protein kinase (BU673731), adenosine kinases and adenylate kinases (BU673745, BU672936). In addition, the signaling molecule calmodulin (BU673090, BU672925, BU673775), a common participant in the MAPK signal transduction cascade, was also found.

Our EST analysis has revealed many transcription factors which are presumably involved in regulation of drought stress-responsive gene expression. The identified transcription factors (Table-5) 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 (Singh, 2002; Park, 2001; Wu K, 2001; Thomashow, 1998; Seki et al., 2001; Shinozaki et al., 2003). Interestingly, transcription factor class represent 17% of our ESTs collection, the most abundant among them are the ethylene responsive factors, (ERF), the translated products of which are known to bind GCC box found in several PR gene promoters and CRT/DRE elements. These elements are reported to be involved in the expression of dehydration and low temperature responsive genes which confer ethylene responsiveness and are known to bind the CRT/DRE e( dehydration responsive element motifs) regulating the expression of dehydration responsive genes. These ERF proteins from one plant species have been shown to function in other plant species suggesting their potential utility in increasing the stress tolerance of across plants (Hep, 2001; Wu. K, 2001; Jaglo KR, 2001; Gu.YQ, 2002). The overexpression of tomato PT15, an ERF gene of arabidopsis, increased the transcript abundance of specific PR genes (Hep., et al 2001). Similarly Wu.K, (2001) reported that overexpression of an tomato ERF transcription factor PT14 showed upregulation of GCC and CRT/DRE motif-containing genes which further enhanced resistance to specific stresses. Also Park (2001) reported that transgenic tobacco plants
over expressing a single gene encoding an EREBP/AP2-type transcription factor showed enhanced resistance to osmotic stress and resistance to Pseudomonas syringae. This has proved that this class of transcription factors can enhance both abiotic and biotic stress responses. The ability of TSI proteins to bind both the GCC and the CRT/DRE motifs demonstrates that these two different stress pathways can be linked by a single ERF gene. Similarly a single ERF gene, the CBF1 (CRT/DRE-binding factor 1), conferred cold tolerance in transgenic arabidopsis showing for the first time that a single ERF gene can have a major impact on a complex plant stress response (Thomashow MF., 1998). Though over expression of a single ERFs has increased stress tolerance, the constitutive over expression of these genes seem to cause deleterious effects. This has been overcome by controlling the expression of ERF genes using stress inducible promoters and this approach has been successful for the DRE-binding factor DREB1A. Arabidopsis transgenic plants overexpressing DREB1A no longer showed deleterious effects. Infact, these plants showed enhanced protection against freezing, drought, and high salinity when controlled by a stress inducible promoter (Shinozaki. K., 1999). Recently, Fowler and Thomashow (2002) identified 306 cold-regulated genes and 41 DREB/CFB-regulated genes using Affymetrix Gene Chips. The Over-expression of the regulatory proteins such as DREB1A and DREB1B has resulted in an enhanced tolerance to drought, salt and freezing (Jaglo-Ottosen et al., 1998; Kasuga et al., 1999) in Arabidopsis. Our EST analysis has uncovered 4 families of DREB class of transcription factors which include DREB 1A, DREB 1B (AY166833), CBF (C Repeat binding factor, Acc No AY 5020522) and DREB 2A.

5.6 Physical maps of N22 rice unigenes

The unigene set derived from the N22 ESTs by CAP3 assembly has been localized onto rice genome (Fig. 5) and the functional annotations of the stress responsive genes have been associated with the unigenes (Fig. 6, 7, 8, and 9). Sequence and positional information at nucleotide level is expected to facilitate development of molecular probes and markers for drought tolerance in rice and most likely in other closely related cereal crops. A major constraint in map-based cloning approaches is the insufficient number of PCR-based molecular markers available to perform fine mapping (Drenkard, 2000). It is obvious that there is a need for generating a large set of EST
based PCR markers particularly that are specific in determining intravarietal differences among Indian cultivars for drought tolerance. They can be directly deployed in marker-assisted selection (MAS) for drought tolerance. The physical maps showing annotated unigenes of indica rice will serve as a reference source to rice researchers in particular and cereal researchers in general.

Particularly, the sequence and map data described here will help in designing gene specific markers in rice. Further, these resources can be used to convert candidate ESTs into PCR based markers, which we have been doing as part of our research program on functional genomics of drought tolerance in rice (data not shown). The EST resources can be used to study syntenic relationships among cereals as these are generally conserved among the grass genomes. We have used many ESTs as RFLP markers for syntenic studies in rice and sorghum (Prasad, PhD thesis).

The genetic association studies including the development of high-density maps constitute an important step in the positional cloning of genes underlying complex traits (Wolford et al. 2000). The sequence information in public databases will provide necessary tools for the creation of new molecular markers and identification of SNPs. Positional cloning in arabidopsis through SNP information is highly interesting (Lukowitz et al, 2000). Unlike other molecular markers, SNPs provide a way to generate highly saturated genetic maps and are amenable for automation (Wang et al, 1998). The SNP frequency pattern along rice chromosomes shows an uneven distribution of polymorphism-rich and poor regions (Feltus et al., 2004; Nasu et al., 2002). Current estimates of SNP frequencies in rice range from 1.70 SNP/Kbp to 11.7 SNP/Kbp (Feltus et al., 2004; W. Zhao, et al., 2004 Nasu et al., 2002; Zhang K. et al., 2004). Further, SNP frequency differs as much as 0.49% between indica genotypes (Nasu et al., 2002), which is considerably higher than japonica genotypes (0.03% to 0.05%). The data above suggests that there exists a high degree of polymorphism in indica cultivars, which makes it possible to develop markers even between very closely related cultivars, which has been difficult to find by conventional methods such as RFLP. The SNP profiles of N22 putative candidate genes will help breeders in understanding genetic determinants of drought tolerance and implementing strategies for efficiently introgressing these genes across rice lines. We have analyzed a few candidate genes for intra-specific sequence
variation using the ESTs (Reddy et al., unpublished), for which both the structural and functional information have been described here in this study.

The data presented here can be used as a resource in identification and analysis of QTLs for drought tolerance in rice. Localizing rice ESTs onto genomic sequence provides a direct route for drought tolerance gene discovery. The structural information of putative candidate genes (Table S-CS1) and genes linked to QTL (Table 5) will be an important resource for such studies. Candidate QTL genes can also be identified from expression profiling experiments, under the assumption that genes that show genotype-specific differences could be the causative agents for the variation in a trait. Physical map locations of rice ESTs observed by mapping EST sequences on to genetically anchored BAC/PAC clones of rice genomic sequences revealed known stress responsive genes in the QTLs associated with drought tolerance (Markandeya et al., 2005). Further experiments are underway to elucidate the precise role of these putative candidate genes in drought stress response in rice.

5.7 Gene expression profiling

Microarray technology which uses Expressed Sequence Tags was first demonstrated by analyzing 48 arabidopsis genes for differential expression in roots and shoots (Schena et al., 1995). The microarray data have already been analyzed concerning a number of plant processes, such as seed development (Girke et al., 2000), expression in response to mechanical wounding and insect feeding (Reymond et al. 2000), defence-signalling pathways using fungal pathogen and signalling molecules (Schenk et al., 2000), brassinosteroids (Goda et al., 2002), pathogen signaling (Schenk et al., 2000), nutrient-dependent changes in expression profiles (Wang et al., 2000; Thimm et al., 2001), and environmental stress responses (Kawasaki et al., 2001; Seki et al., 2001, 2002a, 2002b; Kreps et al., 2002; Ozturk et al., 2002). Model experimental systems in plants such as Arabidopsis and rice are highly amenable to gene expression profiling particularly dealing with abiotic stress. There are now several examples of plant abiotic-stress-related transcriptome profiling that have revealed many new components in stress response pathways (Kawasaki et al., 2001; Desikan et al., 2001; Ozturk et al., 2002; Chen et al., 2002; Kreps et al., 2002; Fowler and Thomashow 2002; Seki et al., 2001, 2002; Klok et al., 2002; Kim et al., 2003; Yu and Setter 2003).
In rice, gene expression profiling has been carried out through microarrays and SAGE (Serial Analysis of Gene Expression), Kawasaki et al., 2001; Rabanni et al., 2002). The results reported so far were derived differently by using various stress treatments, array formats, species, tissue types, and time courses making it difficult to make direct comparison among studies. Also, most of these experiments on expression profiles are conducted under laboratory conditions, and by inducing stress with exogenous compounds (ABA, GA, and SA etc) which are know to accumulate under abiotic stresses. These however, may not accurately mimic true field abiotic stress responses of the rice plant.

In the present study high quality microarrays using the above EST resources were constructed and were successfully hybridized with labeled cDNA from total RNA isolated from a series of field drought stress experiments.. These experiments were conducted under different field capacities using a rainout shelter minimizing environmental variation. The microarray analysis will throw light on gene expression in real field drought conditions. Till now, no lab has reported gene expression in rice under field drought conditions. The expression profiles were also obtained from seedling under different abiotic stresses viz dehydration, ABA, PEG, high salt at different time intervals to analyze cross talks between abiotic stresses.

5.8 MIAME Compliance microarray data

The strategies employed by us for generation of MIAME compliance microarray data has been a success owing to our careful experimental design, selection of array substrate probes, high quality array fabrication, target sample preparations and controlled hybridization protocols. All information related to MIAME guidelines has been strictly followed and appropriately incorporated as described by microarray gene expression data society (MGED) (Brazma et al., 2001). Our microarray gene expression data complies with that of MIAME standards. The critical issues concerning the MIAME compliance is discussed below

The success of our microarray experiment is a result of careful execution and inference under study, which in our case was gene expression under field drought stress at various levels of water stress. Since we conducted a successful artificial simulation of drought stress under rainout shelter, we could achieve proper target samples having a
right reference control sample. This was achieved by conducting a drought stress experiment in rainy season, though it took some time to execute this through conducting preliminary trails. The main reason for choosing rainy season for drought stress experiment is to minimize effect of all the environmental factors on reference control samples. Drought conditions are not just limited to water stress, but lot of other environmental factors during hot summer play a major role in contributing the effects of drought in control plants; these include effect of temperature which causes loss of osmotic turgor in leaves thereby triggering gene expression. Likewise, the other factors include heat waves, and long day lengths. All these will trigger common pathways that are also a part of drought stress response. Any outcome of the expression data will therefore will be taken as gene expression profile under water stress minus the common pathways that are triggered due to other environmental factors. Our microarray expression data was obtained using a null environmental effected control sample, which reflects the actual gene expression under water stress. Uniformity among biological replicates was achieved (Fig. 11).

5.9 High quality array constructed with defined negative and positive controls

The second crucial aspect is the probe substrates used to array onto microarray slides; these should reflect the nature of study for which it is meant. The design of a microarray mostly relies on the arraying of cDNA clones, which may be focused on transcripts associated with a particular library or tissue as in the case of our approach having cDNA clones generated from rice N22 drought specific library and cDNA clones of pearl millet generated from different stress libraries. The cross-hybridization data from rice vs. pearl millet is used for comparative genomics studies or the arrayed probes nature may also be more global, representing all or most of the transcriptome (Forster et al., 2003; Weeraratna et al., 2004). All the clones arrayed were printed in duplicates; these will reflect the consistency of hybridization, thus avoiding errors in the intensities read from hybridization artifacts. Analysis of few gene expression profile graphs revealed equal expression profiles. All the redundant clones were also used for arraying in order to avoid any kind of false interpretations obtained due to the data obtained by a single clone. Significantly, upon analyzing the redundant clones, all of them showed more or less similar kind of expression patterns.
We have achieved a better hybridization quality with high reproducibility by using several negative and positive controls in the chip. The negative controls included non-homologous genes to any existing plant or animal transcriptome, designed from yeast inter-genic regions and empty spots. This allowed us to measure non-specific fluorescence and possible cross-reactivity. For positive controls, we included spiked RNA at equal concentrations in both channels were used for normalization purposes. In all, we have used 23 yeast artificial genes which include calibration controls, to calculate normalization factor to normalize the data, since our microarray design was a reference design (Churchill, 2002; Yang, 2002; Dobbin et al., 2003; Kerr and Churchill, 2001). By using the smoothing factor, we have normalized the data using the above controls (Fig. 24a, b and c). By using LOWESS and the controls as a trained dataset, we normalized the intensities of both channels across all the data-points. The pre and post-normalization plots illustrated in (Fig- 25a, b, c and d) show that all the signal intensities were brought to zero centric.

5.10 Normalization

Normalization is done to minimize the inherent errors in the experiment including unequal quantities of RNA, differences in labeling, varied detection efficiencies between the samples and systemic biases in the measurement of intensities. We applied first level of transformation to adjust the individual hybridization intensities to balance them appropriately so that comparison can be made between the samples Keeping the systemic errors like systematic dependence of log2 values on the intensity, the Lowess (LOcally WEighted linear regreSSion) normalization technique was applied to all the data points and compared with the total intensity normalized plots. Lowess uses weighted function that reduces the emphasis of far away data points. Lowess normalization performs comparatively better and in our case it fits best with our protocols. A comparative analysis between the two different approaches was carried out to select the suitable normalization method.

LOWESS normalization has become our choice of normalization as it is a global normalization that uses all the genes on the array and considers house keeping genes and internal control in calculating normalization factor. In order to avoid a potential error to the data, known candidate genes expression pattern were used to reject the ineffective and
erroneously normalized data. Based on above considerations and initial trial and error runs, we have selected Lowess normalization using trained data from controls used on the slide for calculating smoothing factor and normalized the data. Data analysis was performed with background filtering and eliminating the flagged spots after log transforming the data. The MVA plots of pre and post-normalized data in this study are given in (Fig. 25 c and d).

The efficacy of normalization method and accuracy of normalization factor calculation were thoroughly examined with scatter plots, before proceeding to the clustering of co-regulated genes. It is evident from superimposing scatter plots of pre and post normalized data, Lowess performs well over the other methods (Fig. 25 a, b, c, and d) with our data.

5.11 Co-regulated genes under stress are identified

Finding the genes showing similar expression pattern is the basic idea of clustering in Microarray experiments. Since training data set is not available for supervised classification, we followed unsupervised classification and clustered the genes with hierarchical classification after normalizing expression data. In all the data points from 45 days to 81 days of the development stage, the A2, A3, A4 vectors (A2 denotes 60% FC, A3 denotes 40% FC, and A4 denote 15% FC) were taken for the stress regime analysis of the expression data. Only experiment normalizations were carried out in this preliminary analysis and no adjustment of the data was done since we achieved a good normalization of the data. Hierarchical clustering was carried out for either genes or experiments at an instance on whole slide data. Hierarchically grouped clustered sets of genes were examined and approximate number of clusters to be formed was estimated (Fig. 26a and b).

Analysis of data revealed that there are at least 2 clusters showing up regulation and 2 down regulation. The annotated genes falling in these clusters interestingly include a number of novel genes from N22 library along with some known candidate genes for drought tolerance. Some of the interesting clusters and the gene expression profiles are discussed below and also again when we discuss time series and across stress regimes.
5.12 Sets of co-regulated genes as a function of a constant stress at different growth stages have been identified

The group of genes revealed in gene expression profiling of A2 (60% FC) may play a key role in adaptation of plants to 40% water stress, since N22 genotype shows a host of adaptive mechanisms to drought as discussed earlier. The minimal stress induced throughout the plant’s life cycle at the fixed 60% FC may lead to identification of genes potentially involved in drought escape and adaptation mechanisms. The cluster analysis revealed that there are at least 361 genes from rice and pearl millet that are co-regulated during different developmental stages of the N22 at that stress level. These genes include both potential candidates for drought escape and developmental specific genes. Our results clearly showed a >3 fold upregulation of a set of well known genes till day 59 under 60% FC: these genes include DREB1A, DREB1B, RD22, Zinc finger protein, Zinc transporter, dehydrins, cytochrome c reductase, APF1, H+ ATPase, glycine rich proteins (GRPs), sodium symporter and calmodulin (Table S-A2L). Besides that, we have uncovered 54 genes whose functional role are not yet known, but are co-regulated with the above genes.

Transcription factors (TFs) of DREB family have been well characterized and their regulation mechanism through the recognition sequences in the promoter elements of several genes down-stream the signal transduction pathways in the model plant Arabidopsis are clearly defined (Yamaguchi-Shinozaki and Shinozaki 1994, Haake et al. 2002). The early response of both the TFs indicate that they act early in the stress signal transduction pathways switching on a number of genes, thereby assuming critical importance. Analysis of ABA-dependent pathway (Yamaguchi-Shinozaki and Shinozaki 1993) revealed the role of those genes which contain motifs belonging to MYC and MYB recognition sequences are essential for induction of expression of several genes, for instance RD22, by ABA and drought. Furthermore, ABA-inducible MYC and MYB transcription factors may function co-operatively in the ABA-dependent expression of RD22 and other related genes (Urao et al., 1993, Abe et al., 1997, 2003). These genes are over expressed until day 59 only and thereafter down-regulated > 4 fold. It seems reasonable to argue that these genes might have triggered the activation of other genes by that time (59th day in this case) i.e., entering into reproductive stage. However, we can
separate stress induced gene expression component from that of developmental changes from additional experimental data. The over-expression of sets of genes > 10 fold in cluster 9 between day 50 and 59 which includes RAB 28, Zinc Finger protein, cytochrome p 450, kinase, DREB2, catalase, BAG domain containing protein, EF hand Ca+2 binding protein, leucine zipper, Cyclin, AG motif binding protein substantiates our argument that these genes are involved in stress adaptation. Further, this cluster includes 70 genes of novel and unknown function. That several abiotic stress factors induce the expression of upstream transcription factors, such as DREB1 (cold) and DREB2 (osmotic change) is consistent with the recent mRNA profiling (Cheong et al. 2002, Seki et al. 2002) in model plants. The over expression of DREB2 by >10 fold (cluster 9), at day 52 in our experiments, is likely to be involved in osmotic adjustment as shown in many expression profiling experiments. In fact, these belong to the first group of genes which are expressed and demonstrated to be involved in osmotic adjustment (Ishitani et al., 1997; Cheong et al. 2002, Seki et al. 2002). There seems to be a greater cross talk among cold and drought stresses, since the over-expression of DREB1A early on, lead to triggering a cold and salt responsive protein RC12B, which showed sudden expression >3 fold on day 59 (cluster 10). The genes which belong to osmotic adjustment are glycine rich protein, RD22, dehydrins. Interestingly, genes involved in oxidative stress are triggered which show overexpression by > 10 fold.

The co-regulated genes under increased stresses in A3 (40% Field capacity) and A4 (15% Field capacity) show a large variation in their expressional patterns and also in the levels of expression which vary from 3 fold to 20 fold up-regulation till day 71. Further, as the growth stages is effected adversely from day 71 onwards, a subset of those genes are down-regulated by > 10 fold compared to that of plants in unstressed condition (Fig. 29b). There are at least 350 genes, whose expression seems to correlate with the extreme severity of the stress. The list of genes and their fold changes along with their Accession Numbers are given in (Table S-A3L, S-A4L). We see a large variation in expressional patterns as the severity of stress increases from A2 to A4 and likewise, the genes at A4 are grouped into many clusters which show varying degree of fold changes. Interestingly, many of these genes in A4 are from pearl millet (as a result of heterologous hybridization of rice transcripts under stress with pearl millet clones). In fact this
observation is valid and consistent with our scheme as the probes of pearl millet are from differential cDNA libraries from cold and drought stressed seedlings and pearl millet is known to be more drought tolerant than rice. This information will be an invaluable resource for identification of candidate orthologues of pearl millet in rice.

At stage A3, wherein the severity of stress is not extreme, a large number of genes involved in signaling, transport pathways, and hormone response in addition to genes of unknown and novel function have been identified. These include Ras, RAB28, serine threonine kinase (STK), EF hand calcium binding protein. These genes have been earlier reported to be induced by abiotic stress (Takahashi et al., 2000). Other genes include receptor like kinase, casein kinase, adenosine kinase, protein kinase, hexose transporter, ABC transporter, and NADH oxidoreductase. Interestingly, genes encoding anthocyanin reductase, 14-3-3, NADH oxidoreductase methionyl aminopeptidase, amine oxidase, auxin induced gene, jasmonic acid induced gene, leucine zipper protein and aldehyde dehydrogenase, all showed significant changes in expression consistent with their role in stress adaptation (Kawasaki et al., 2001; Seki et al., 2001, 2002). We observed down regulation of a gene encoding photosystem II 10 kDa gene. Earlier its reported that, psbO and psbR encoding the 33 kDa and 10 kDa proteins of photosystem II (PSII), respectively, when down-regulated deactivate the photosynthetic function and occurs during dehydration (Sherwin and Farrant 1998; Farrant et al. 1999). The expression pattern of these groups of genes is rather complicated since several genes of photosynthetic apparatus are differentially regulated at different stress regimes.

Gene expression profiles under severe stress A4 (15% FC), were distinct and interesting, in that a group of co-regulated genes in cluster 8 are up-regulated during the beginning of stress initiation and continue until day 63 followed by a drastic down-regulation by > 10 fold (Table S-A4L) These include genes encoding, ethylene forming enzyme, signal transduction associated histidine kinase, hydrolase, receptor like protein kinase, AG motif binding protein, Zinc finger protein, DREB2, DREB1b, elongation factor 1, RD22, chlorophyll a/b binding protein, lipase, aldolase, cytochrome P450, signal recognition particle, luminal binding protein, RAB, pectin esterase, jasmonate induced protein, extension like protein, bZIP, hydrophobic protein, RCI2B cold induced protein and Hsp90. Some of the observed changes in expression profiles of the above mentioned
might be developmental specific, since there is a transition from vegetative stage to seed filling stage when these genes down-regulate by >10 fold by day 71 (Fig. 29d) another set of 74 genes are observed to significant up-regulate (> 10 fold) in cluster 9 (Fig. 29f). Interestingly, the elevated expression of these genes starts up-regulating on day 67 and continues till the death of plant in A4 (day 71). In fact a few of those genes are up-regulated by as much as 200 fold. These include IMP dehydrogenase, adenylate cyclase, 4 unknown genes, and 3 novel genes (Table S-A4L) the same set of genes when compared across the stress regimes on the same day, are observed to have over expression only in case of A4. Presumably, these unknown genes may be associated with response to severe stress leading to plant death. This cluster also includes genes associated with DNAases, proteases and other degrading enzymes in the dying plant. The other large proportion of genes that are highly up-regulated (> 10 fold) during the extreme dry phase of the plants are NAM, alcohol dehydrogenase, alfa-amylase, Myb, p450, peroxidase, jasmonate induced protein, hexose transporter, xyloglucan endo-transglycoxyrase, phytocyanain, GST, glutamate dehydrogenase, arginine serine rich splicing factor, trehalose 6 phosphate synthase, hydratase, plastocyanin, and 26 unknown and 11 novel sequences.

5.13 Co-regulated gene clusters in panicle under stress have been identified

One of the interesting observations in this study is that the well known DREB families of genes are specifically up-regulated only under severe stress (A4) in panicles. None of these were shown to have elevated expression in A2 and A3 but were observed in minimum stress conditions (A2) in vegetative tissues, more importantly, the data revealed that these groups of genes begin to show up-regulation from the time of panicle development and reach to a maximum by day 67, followed by a decline. The following genes are co-regulated in A4 along with the DREB family of genes these include Zinc finger protein, cytochrome p450, catalase, lipase, RD22, RAB28, thioredoxin, pollen specific protein BAN102, BAG domain protein, thaumatin like protein, dormancy associated protein, gigantia, EF hand calcium binding protein, extension, bZip, RCI2b, glutaredoxin, Hsp90 besides 97 unknown genes. The other noticeable aspect in panicle gene expression profiling is that most of the co-
regulated genes are among the unknown category (Table S-A4P). This is mainly due to the fact that relatively few panicle development associated genes are characterized in rice and even in other grasses. This is obvious as identification and characterization of genes is mostly limited to vegetative phase of plants.

The expression profiles of panicles show that there are significantly more TFs showing stress associated changes in panicle tissues (Table S-A2P, S-A3P, S-A4P) than that of vegetative tissues. These include known stress responsive transcriptional factors MYB1, R2R3 MYB, MYBS3, AP2 domain containing protein, dnaJ, bZIP, WRKY, homeodomain leucine zipper protein, beside developmental specific TFs like MADS, NAM, NAC5 and AG motif binding protein. The expression of the last 3 TFs is related to flowering and transition of developmental stages in arabidopsis (Tzeng et al., 2002). Further the expression of combinations of transgenes belonging to AP1, PI, AP3, AG, SEP3 (MADS) families was demonstrated to be sufficient to transform leaf organs into flower organs. (Singh KB 1998; Veylder et al., 2002; Bovy et al., 2002; Pelaz et al., 2001; Honma et al., 2001). We have also identified various forms of NACs (NAC and NAM). These are known to encode a polypeptide containing a plant-specific, highly conserved N-terminal domain (Aida et al., 1997). In the arabidopsis genome, about 100 such putative members of NAC genes have been identified (Ooka et al., 2003; Riechmann et al., 2000). Among them, NAC and NAM have been reported to be targets for APETALA3, which are suggested to control cell expansion in specific flower organs, and reported to be associated with senescence, and also seem to be responding to abiotic stress (Sablowski and Meyerowitz, 1998; Takada et al., 2001; Vroemen et al., 2003; John et al., 1997). Apart from these, several new transcription factors have been identified which were not previously known for their role in abiotic stress.

Interestingly many members of MAPK family genes are represented among the differentially regulated clusters. MAPK6 starts showing up-regulation by > 3 fold in A2, on day 71 and continues to be up-regulated until day 75, while MAP2K and MAP3K are represented in A3 clusters, which reach a maximum expression level on day 71. This indicates all the above MAPKs are co-regulated at the same stress levels.
MAP kinases are known to be activated by multiple stress conditions including drought (Mizoguchi et al. 1996, Ichimura et al. 2000). The MEKK1, MEKK2 and two downstream MAP kinases, MPK4 and MPK6 have been proposed to be components of a MAP kinase module involved in salt and cold stress responses (Teige et al. 2004). In another study, the MPK4 has been shown to be weakly induced by high salt (Ichimura et al., 2000).

5.14 Co-regulated genes at vegetative stages across all stress regimes

We have investigated into a group of genes that are continuously up-regulated during the entire spectrum of stress intensities throughout vegetative growth. Accordingly, the expression profiles at A2 (60% FC), A3 (40% FC), and A4 (15% FC) stress regimes were analyzed and co-regulated genes on different days in vegetative stage (day 45, 52, and 55) have been grouped. The clusters revealed that at least 350 genes of rice and about 200 genes of pearl millet respond differentially during vegetative stage. These fall into a total of 8 up-regulated clusters which show greater than 3 fold changes in their expression, and consistently differ in their expressional levels across stress regimes. There are about 6 clusters which are down-regulated and reveal expressional patterns similar to that of the up-regulated ones. The genes identified during the initiation of drought stress include an early drought induced protein, proline rich protein, alfa tubulin, cyclophillin, salt tolerance protein, beta expansin, UV light related transcription factor, dehydrins, Hsp90, ethylene responsive transcription activator, transmembrane helix receptor, plasma-membrane major intrinsic protein, ABC transporter, H+ transporter, HOS 59, cytochrome P450, CEO protein, and genes involved in signaling, serine threonine kinase and adenylate kinase. These are followed by other signaling genes overexpressing >4 fold as the stress prolongs till day 59 which includes MAP3K, protein kinase, CK2 regulatory sub-unit, and among the TFs are DREB1A, AP2 domain containing protein, Myb, Zinc finger protein, AG MOTIF binding protein (Table S-45L, S-52L, S-55L).

Vegetative stage also represents a large number of genes involved in oxidative stress, including many genes which are yet to be characterized in rice. Interestingly, a group of well characterized genes for their stress responsive upregulation are seen to be
down-regulated in A3 at day 52: these include photolyase blue light receptor, aldolase, CCAAT binding transcription factor, hydrolase, MAP6K, DREB2, dehydration responsive protein, proline rich protein, Calcium dependant protein kinase (CDPK), peroxidase, cytochrome p450, WRKY, Zinc finger protein. There are a few genes which seem to be fluctuating across stresses, these include DREB1B, AP2 domain TF, NAC, NAM, chlorophyll a/b binding protein, c2 domain containing protein, low temperature and salt responsive protein, NSF vesicle attachment protein, PR protein, water channel protein, GRP, R2R3 MYB, EF and calcium binding protein, gigantia, nif U like protein, protein kinase c inhibitor, leucine zipper protein (LZP), wound induced protein, calmodulin, osmotin, proteinase, KH domain ZFP, disease responsive protein, DNA repair protein, RAD23, chitinase b, Jasmonate induced protein, MAPK, ASR1( ABA and stress inducible protein), far red impaired responsive protein and serine threonine kinase.

Gene expression profiling at the end of vegetative stage is represented on day 55, which has revealed much more complex patterns of co-regulation. Among all the genes, cluster 7 genes are highly expressed (>5 fold) and include DRE-binding protein 1B, extensin-like protein, GP28, lipid transfer protein precursor, lipid transfer protein LPT III, NADH dehydrogenase, R2R3MYB-domain protein, MYB1, P-Protein, putative anthocyanidin reductase, , Rice metallothionein, type I membrane protein, dof zinc finger protein, fiber protein Fb14, fruit-ripening protein, photosystem I antenna protein, phytochrome-associated protein, SNF2 related domain and 28 novel genes with unknown function.

5.15 Identification of stress responsive genes during seed setting and seed filling stages

We have identified many highly regulated genes at this stage of rice life cycle most of which have not been reported earlier. The genes which are highly expressed (>4fold) are observed in this stage. Interestingly, these are highly expressed only under severe stress. These include, receptor like protein kinase, beta tubulin, cellulose synthase, NADH dehydrogenase, putative CTP synthase, NBS-LRR-like protein, putative RAD23 protein, CDPK, acetyl-CoA carboxylase and pectin esterase, 3 hypothetical proteins, 4
expressed proteins, 4 novel and 6 unknown proteins. Most of these genes which are continuously up-regulated across the stress regimes represent those which may be developmental specific and also due to stress. These drought responsive developmental specific genes at this stage include 1-aminocyclopropane-1-carboxylate oxidase, 3-dehydroquinate dehydratase, allantoinase, cellulose synthase, cinnamyl alcohol dehydrogenase, homeodomain leucine zipper protein, putative naphthoate synthase, NHL repeat-containing protein, oligouridylate binding protein, pyrophosphate-fructose-6-phosphate-1-phosphotransferase, Rer1A protein, serine protease, pyruvate dehydrogenase complex, rac GTPase activating protein, Exportin 4, beta-glucosidase aggregating factor, zinc-finger homeodomain, sorbitol transporter, sugar ABC transporter, periplasmic sugar-binding protein, trehalose-6-phosphate synthase, zinc transporter these genes may be associated with seed filling, seed setting and appear to be developmental specific.

Along with these, there are also few drought stress responsive genes represented, in this cluster, among which few representative examples are C2 domain-containing protein-like Chaperonin, adenylate kinase, shikimate 5-dehydrogenase, DRE binding factor, copper chaperone late embryonic abundant-like protein, lipid transfer protein LPT III, myb3R1, peroxidase, GST, and early drought induced protein. Interestingly, this early drought induced gene was found to be in QTL region associated with chlorophyll ratio, 1000 seed wt and lodging incidence in rice (Harushima et al., 1998). This reveals that some of the genes identified earlier as stress inducible are not only stress responsive, but whose high levels of expression may be due to transition from one developmental stage to another. Since our data revealed that there are many stress inducible genes co-regulated along with developmental specific genes, the functional role of these needs to be characterized. Also it can be that a few of the developmental specific genes may actually be stress responsive. There are many variations observed during transitions from one developmental stage to another as the stress is prolonged.

One of the important observations is that the highly up-regulated gene clusters during later stages of growth, day 67, 71 and 75 represent the same genes which show up-regulation during earlier vegetative stages, but the levels are significantly higher. These groups of genes are up-regulated > 10 folds on day 71, followed a sudden drop on day 75. This may be interpreted as a result of severe physiological condition of the plants on day
75 in A3 and A4, where they are almost dried up and dead. Interestingly few genes in day 71 show an increase in their transcripts to an extraordinary level. The genes in cluster 10 (Day 71) exhibit >120 fold over expression and a subset of these genes over express > 200 fold in A4 (Table S-71L). However, most of these genes are of unknown function and a few are classified as novel. Functional characterization of these genes will provide leads in understanding, the key role played by these genes during terminal death phase of the plant. Some of the annotated genes in this group are, IMP dehydrogenase, GMP reductase and glutathione S-transferaseII, which are expressed >120 fold, and adenylate cyclase and expressed protein which are up-regulated > 200 fold. The list of all the genes along with their GenBank Accession Numbers, functional annotations for all the clusters mentioned in the results are given in accompanied CD-ROM. Most of the genes identified by in-silico analysis are seen often varying during transitions across developmental stages. Though it is difficult to discuss each and every detail about individual gene responses, the profiles clearly reflect the changes in the expression pattern of co-regulated genes under stress. The data presented in this study has focused mainly on transcript changes under field drought conditions. In our effort to understand the molecular genetic basis of stress response in rice, we succeeded in capturing drought induced transcriptional changes in near natural growth condition in field., The gene expression profile data in different combinations of stress treatments and developmental stages, throughout the life cycle of rice plant has provided valuable insights into the identification of many genes dispersed across the genome and also the nature of genetic x environmental interactions. As per our knowledge this is the first report on gene expression profiling studies under field drought stress covering all crucial stages of rice development. This substantial amount of data generated is expected to open up new insights in understanding stress responsive gene regulation and adaptation and eventually identification of candidate genes of drought tolerance in rice.