4. RESULTS
4. Results

4.1 Generation of EST Resource

4.1.1 High Quality Expressed Sequence Tags Generation from a normalized cDNA library

Recombinant cDNA colonies, 6144 in all, were picked from the normalized cDNA library to make high quality plasmid minipreps. The quality and concentration of the recombinant clones was estimated by running on 1% agarose gel electrophoresis (Fig. 1). All the 6144 plasmid clones were 3’ end sequenced. Our optimized sequencing efforts, through preparation of high quality, uniform concentrations of sequencing templates, use of reduced dye chemistries, and employing cost-effective purification methods, drastically reduced the time and costs of single-pass sequencing. The sequencing strategy proved to be very efficient with a success rate of ~85%. The raw data from the sequence analyzer of MegaBACE were exported and were base called using PHRED with Codoncode Interphase, from which we got 5044 readable sequences with high quality index and PHRED score of greater than 20. The PHRED quality values and chromatogram of a representative sequence are displayed in Fig. 2a.

The fragment read length in a batch of 4000 sequences varied from 200bp to more than 1000bp, with 3174 sequences showing more than 500bp. Of these, 1028 sequences were of 750bp and 706 sequences were 850bp or more in length. The average read length of the generated ESTs is 568 bp (Fig. 2b). These sequences were considered for further sequence analysis. All sequences were screened for quality using base quality scores. Sequences greater than 800bp on an average shows 500-600bp having PHRED score of more than 20, about 400 – 500 bases having PHRED score of more than 30 and around 300 – 400 bases have PHRED score of more than 40. The summary of EST generation is given in (Table 3).

4.1.2 Data preprocessing and identification of biological contaminants

Before annotation, the raw sequences were processed for identification of any biological contaminant, especially vector sequences, and contaminants that were incorporated during library construction. The vector sequence PT₇T₃Pac was screened
using Cross_match program using CodonCode interphase and trailing vector sequences at 5’ and 3’ end sequences from the cloning site were traced and removed. The sequences which showed high homology to the cloning vectors using a gapped BLASTN program with a restrictive expect value of 1e-20 identified 354 vector sequences, which were also eliminated from further processing. Since we used a Normalized library we searched for any biological contaminant using a batch of 100 sequences and analyzed the most redundant clones. Among these we found highly redundant sequence showing homology to viral origin (Adenovirus type 2). All these were removed from further sequence analysis. Homopolymers of terminal poly A or poly T were identified, low quality sequence regions were trimmed and sequences less than 100 bp in length were also excluded. Of these, 390 were found to have no homologues in the nearly-completed Nipponbare rice genome sequence (IRGSP, 2005). Since the gene content and order between indica and japonica rice genomes are virtually identical (Bennetzen et al., 2004; Ma and Bennetzen, 2004), it is likely that most or all of these ESTs are from microbial contaminants from greenhouse grown rice seedlings from which the ESTs have been generated. For instance, 380 ESTs were removed prior to the submission of the 4240 sequences because it was clear they were viral sequences from Adenoviral type 2 encoding minor capsid protein VI (Table 3). Microbial contamination is an unavoidable outcome of EST analysis on greenhouse or 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 3.

### 4.1.3 GenBank Submissions

In the present study only cleaned sequences after pre-processing and removing sequences less than 100bp length were submitted to GenBank in two batches of 1152 EST sequences and 3088 EST sequences. The first batch of 1152 sequences were released in GenBank (GenBank release 07-OCT-2002A; Accession numbers: BU672765 to BU673915) under the title “Novel EST enrichment with normalized cDNA libraries from drought” (Reddy, A.R., Markandeya, G., et al., 2002). The second batch of 3088 sequences were released in GenBank (GenBank release 29-APR-2003; Accession numbers: CB964418 to CB967504), entitled “ESTs from a normalized cDNA library of
Results

drought stressed rice seedlings (*Oryza sativa* L.cv Nagina 22)” (Markandeya, G et al., 2003)

4.1.4 Construction of N22 unigene set

In order to uncover the unique transcripts represented in the N22 library, we have clustered 5815 ESTs using CAP3 program (Huang and Madan, 1999), which include 4240 ESTs generated in the present study and 1575 ESTs generated earlier. Since these ESTs are generated from 5’as well as 3’ends, CAP3 program was of our choice. The assembly of these ESTs through CAP3 revealed a unigene set of 2,067 sequences presumably representing unique transcripts from our library. Of these, 390 sequences were removed as microbial contaminants leading to the identification of 1677 N22 unigene set. The clustered unigene set assembled into 828 contigs and the remaining 1239 were unique transcripts. The average read length of the assembled sequences ranged between 400-700bp.

4.1.5 Digital Northerns

Apart from providing an efficient method of 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). However, it still contained many more copies of some transcripts than others. 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 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 (Fig. 3).

4.1.6 Annotation

Annotation of the assembled unigene set was done through homology search in the NCBI nr nucleotide and protein databases, using BLASTN and BLASTX programs respectively. The annotation revealed that 57% of the unigene set has hits with known putative function, the remaining 43% of the unigene set comprised hits with no functional characterization which include, expressed proteins, unknown proteins, hypothetical
proteins, putative proteins, predicted proteins and ribosomal proteins. These also include 25% of unigenes which had no known hits in the databases. To differentiate spliced variants and paralogues, all the 5815 EST were searched for homology in nr EST databases. Among these ESTs, 334 did not show any homology to rice dbEST or rice cDNAs but were localized onto the rice genome sequence (IRGSP, 2005). This revealed that these were unique to our library and constitute novel sequences identified from our library and constitute 19% of the total N22 unigene set. These novel ESTs may provide evidence for the insilico predicted genes and will assist in annotation and expressional evidence. 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 has been annotated and functionally classified based on GO database (Gene Ontology Consortium, 2001). The transcription factor class constitutes the third highest category of annotated genes followed by that of cellular metabolism and protein synthesis (Table 4). The analysis of the annotated transcriptome revealed many potential stress responsive genes. These were earlier identified through comparative in silico analysis of paralogues from multiple sources of rice and orthologues from other plants (Babu PR., PhD thesis; Markandeya et al 2005 communicated). This deep coverage and analysis of transcriptome of the drought stressed leaf library resulted in the identification of 589 potential stress related genes. As these are from a normalized library constructed from drought stressed seedling tissue, the profiles may provide clues in identification of drought stress responsive genes. The classification of the stress responsive genes into 17 functional groups (Fig. 4) revealed that transcription factors were efficiently captured.

Four of the of the well known DREB family transcription factors OsDREB1B, OsDREB1A, OsDREB2 and OsCRT/DRE binding factor were identified among the ESTs and were targeted for further analysis. Two of these full length cDNAs (DREB1B and CRT/DRE binding factor) have been isolated and have been deposited in GenBank Accession Numbers (AY166833, AY502052). These were initially cloned in pT7T3Pac and TA cloning vector. The cDNAs were then excised and cloned in pBluescript KS+.
The fragments were then excised from pBluescript and cloned into plant expression vector pE1805 driven by a super promoter. In parallel these were also cloned in frame in pET28 series for bacterial expression. In a separate study analysis of transgenics of these genes were developed and analyzed.

4.2 Transcript mapping and chromosomal localization of N22 unigenes

Transcript mapping of N22 unigenes were done using the 2094 N22 unigene set and were localized on to rice genome by a search against the TIGR japonica rice assembly through BLASTN algorithm. Functional annotations were then associated to the mapped unigenes and displayed using cMap software (Fang et al., 2003). Not all unigenes could be displayed in regions of high density due to pixel limitations of the image file. The distribution of the identified chromosomal locations of the unigene set among the 12 rice chromosomes (Fig. 5) was displayed in which 810 were mapped onto chromosome 1, 667 on chromosome 2, 738 on chromosome 3, 484 on chromosome 4, 537 on chromosome 5, about 400 to 500 on chromosome 6, 7, 8 and about 300 to 200 on chromosome 9, 10, 11. The chromosomal positions of unigenes and the location of identified stress responsive genes are represented in Fig. 6, 7, 8 and 9. The number of exons and their sizes of these unigenes are given in table S-CS1 in accompanied CD-ROM.

Gene organization analysis by aligning the Unigene set sequences onto the genomic sequences revealed genes with single exon to twelve exons. Single exon genes were mainly those having small coding sequences. The structural analysis of these 589 putative stress responsive genes revealed extensive variation both in number and length of exons (20bp - 1600bp) and introns (37bp – 2000bp). Few target QTL regions were extracted from reference genetic maps (Harushima et al., 1998) and QTL studies associated with drought and yield traits (Price et al. 2002; Zhang et al. 2001; Babu et al., 2003). Putative candidate genes spanning such quantitative trait loci were identified for a few of the identified stress responsive genes (Table 5).
4.3 Artificial simulation of drought stress under field condition for expression profiling studies

4.3.1 Field drought stress response

Field drought experiment by means of artificial simulation of drought in various levels of drought, (60% FC, 40% FC, 15% FC), was achieved using high precision electronic irrigation controllers. The samples from these experiments showed RWC ranging from 60% to 90% indicating severe drought compared to normal unstressed plants. The experiment was conducted in such a way that reference control had negligible effects of drought by chosen rainy season to induce drought. The design of experimental plots, rain out shelter and electronic irrigation controller are shown in Fig-10a & b. Validation of the whole experiment was put on trail before going for the final experiment.

4.3.2 Avoidance of variation in biological replicates

We have used 30 biological replicates in each plot which were sufficient enough to yield samples across the experiment and we did not observe any variations during the course of the experiment. The variations in biological replicate was overcome with filling the plots with uniform soil strata and arranging high precision water inlets for each plant, such that a uniform irrigation pattern is maintained for all the plants. The water inlets used were pipes having internal drippers which delivered water at 99.999% accuracy. Hence an error of ±0.0001% was found to be negligible. The design of biological replicate is shown in (Fig. 11). The methodology of calculation of field capacities and irrigation regulation is described in materials and method section. The N22 plant have a life cycle of 95-105 days, in which the vegetative stage spans from day 5 till day 55, pre-blooming stage from day 55-60, milking stage from day 60-70 and seed setting stage spans from day 71 till day 105. The observations at different stages of rice life cycle (vegetative, pre-booting, milking, and seed setting stage) are described in the following sections. The complete experiment setup is shown in (Fig. 12a), and the result of the experiment where a stress gradient is clearly observed in shown in (Fig. 12b).
4.3.3 Vegetative stage and Pre-bloom or Pre-emergence stage

The induction of various stress levels (60%, 40%, 15% FC’s) began on day 30\textsuperscript{th}, following which the appropriate stress levels were obtained on day 38 (Fig. 13). The phenotypic response was observed on day 42 with leaf rolling symptoms at 15% FC. Vegetative sampling started on day 42 onwards till day 55(42, 45, 49, 52, and day 55). For microarray expression analysis, only day 45 and day 55 samples were considered. Pre bloom stage in our experiment was observed from day 52 till day 60 across different field capacities. The N22 genotype generally boots from day 55 till day 60. Contrary to this booting occurred at day 52 in A4 (15%FC) (Fig. 14). This observation of early booting is presumably the drought escape phenotypic response of N22. Accordingly samples were collected on day 59 for all A2, A3 and on day 52 for A4 sample. Expression analysis was done for this lone data sample.

4.3.4 Milking stage or seed filling stage

Milking stage varies from day 57 to 60 and continued till day 70. Significance phenotypic variation is observed in the plant growth pattern during this stage. These include; number of tillers varied drastically between A2, A3, and A4 ranging from 4 tillers in A4 to 28 tillers in A1 as shown in (Fig. 15). Plant height was observed to be decreased drastically as the stress increases, and was found to be 51, 48, 41 and 32 inches in A1, A2, A3 and A4 respectively. The comparison of plants under different field capacities is shown in (Fig. 15); yellowing of leaves in A4 was observed on 64\textsuperscript{th} day indicating an initiation stage of wilting; the milking during its initial period in the seeds was almost reduced to 30% compared to A1 control, and further those 30% also stopped due to increase in the severity of the stress. Similarly, it was 45% milking in A3, of which 30% continued milking until the initiation of seed setting stage and in A2 80% of which continued milking and 70% of the seeds entered the maturity stage. Samples were collected on day 63 and day 67, which were used for microarray studies.

4.3.5 Seed setting stage

This stage bought much data as it was inferred through Relative Water Content, which dropped considerably in A4. The RWC dropped to 61%, which is know to be the extreme drought conditions. The rice plants were totally lodged in the beginning of this
stage at day 71 as observed in A4 and in A3 they were just at the beginning of lodging (Fig. 16). The fresh weight was reduced in A2 to a minor extent. Further stress until day 75 dried up the rice plants almost completely in A4 and A3 (Fig. 17). The extent of seed setting in A4, A3, and A2 were 0%, 35%, and 90% (but with reduced grain weight compared to A1 control). The relative water content (RWC) has been measured throughout the stressed period and observed to follow the expected stress levels as represented in (Fig. 18).

4.3.6 Drought Stress Recovery

Test plots were recovered from drought stress in A4R on day 67 in milking stage, the recovery by providing irrigation equivalent to that of control and the first recovery samples were collected on day 71 in A4R. Subsequently the samples that were collected during recovery are, A4R on day 71, 75, 78 and 81, A3R samples were collected on day 78 and 81, and A2R samples were collected on day 81. We successfully collected samples across all the developmental stages as described in materials and methods which include stressed samples from 52 field drought stress regimes and 28 seedling stressed samples, and 80 samples from unstressed field and unstressed seedlings for control samples (Fig. 19).

4.4 Fabrication of 15.5 K cDNA chip

4.4.1 Gene resources for microarray expression profiling

In all 7332 clones went into the fabrication of cDNA microarray these include 6144 of rice ESTs generated in the present study and 1152 ESTs of Pearl Millet (Rabi Ph.D thesis). These were PCR amplified, scored and quality was monitored by running on 1% agarose gel. Various controls (Table 6) were spiked in during cDNA target labeling which are used in normalization of intensities and in calculating fold changes.

4.4.2 Quality of microarray fabrication

Samples from different microarray manufactures were procured and scanned in a fluorescent scanner for background noise, among all Superamine substrates from Telechem showed lower background and our analysis suggests, however, that aminosilane offers a more consistent surface with lower background fluorescence. aminosilane coated glass microscope slides have been the most consistent. In addition,
the spot morphology on Superamine slide is much more uniform, with fewer doughnuts than on any of the alternatives we have investigated (Fig. 20). The cDNA microarrays were fabricated using a reference design, with 24 yeast artificial genes, λ DNA, besides 6144 ESTs from rice and 1152 pearl millet printed in duplicate.

4.4.3 cDNA Array

The cDNA microarray chip consisting of 15552 features with 6144 ESTs from drought stress induced cDNA library of Nagina22, 1152 ESTs from drought stressed *Pennisetum glaucum*, were arrayed onto superamine coated glass substrate using Omnigrid (GeneMachines). We are able to print 15552 featured chips with equal spacing between spots, between subarrays and arrays, consisting of 162 cDNA clones in each subarray stamped in duplicates (Fig. 21). Target labeling was done for 80 data points of stressed samples and 80 control samples with fluorescently labeled Alexa flour dyes coupled to aminoallyl modified dUTP, and aminohexyl modified dATP using indirect labeling method at all the data points successfully. The efficiency of fluorescent dye incorporation was calculated as described in materials and methods, the ratios of base dye ratio and concentrations of fluorescently labeled cDNA were considered for hybridization.

4.5 Analysis of microarray data

4.5.1 Quantification of Expression levels

Overlay of the two channel intensities revealed the hybridization is excellent (Fig. 21) and more or less similar intensities were observed in adjacent in-slide replicas. We are successful in achieving such accuracy in all the data points. All the spots were girded and a sum of 15552 cells were generated (Fig. 22a, b) with 5 - 23 pixel spot size adjusted as to cover all the features of the superimposed slide. Some custom perl scripts were used to facilitate the annotation to the features on the array. Efficiency and accuracy of quantification was examined manually throughout the slide and artifacts were flagged off. The slide intensities were examined with scatter plot of two channel intensities (Fig. 23)
and QC plots for spot saturation. If the intensity values approaches to maximum ($2^{16}$) per pixel for a 16 bit scanner comparison are no longer meaningful as the array elements become saturated. In all the data points studied we found negligible number of saturated spots and quantified spots from each analysis were exported for further analysis.

4.5.2 Normalization channel intensities
The systematic bias in labeling efficiencies was overcome by using a reference design using artificial genes as controls. A total of 25 calibration and ratio controls printed in triplicate and the intensities of these controls were used to calculate a scaling factor for normalization using LOWESS. We have tested various methods for choosing appropriate normalization method to fit our data and lowess normalization with spike-in controls has become our choice of normalization. The plots of the controls before and after normalization from representative slides are shown in Fig. 24a, 24b and 24c. Similarly the intensities of all the spots were bought to zero centered for effective normalization. It has been observed based on pre and post normalized RI plot (log ratio intensity vs. log intensity product) that the normalization was effectively done (Fig. 25a, b, c, d)

4.5.3 Clustering of Co-regulated genes
4.5.3.1 Cluster analysis of co regulated genes across developmental stages from leaf tissue with in stress regime (Time series)
Hierarchically grouped clustered sets of genes were examined and approximate number of clusters to be formed was estimated and k-Means clustering was carried out on the genes. Clustering based on experiments was also considered at certain points to evaluate data integrity and relatedness. Since there were two technical replicates (not dye swap) at certain data points we considered ANOVA results and also each replicate analyzed separately as they are neither dye swap replicates nor biological replicates. Significant genes from ANOVA were hierarchically clustered and further k-means clustered to ~10-20 groups based on hierarchical clustering results (Fig. 26a, b). This analysis revealed such replicates are more or less similar in expression profiles, which is
an indicative negligible artifact during hybridization process. The number of differentially expressed genes of leaf samples at various drought stress levels at 60% FC (A2), 40% FC (A3) and 15% FC (A4) and details of the pattern of gene expression at each stress regime from day 45 till day 75 is described in the following sections. The classes of genes and their functional roles are described in discussion part of this thesis.

4.5.3.1.1 Expression pattern from time series analysis of A2 (60% FC) leaf sample
Data from the same stress regime at different days from 45-75 for time series analysis in A2 was k-means clustered (Fig. 27a). The clusters reveal that there are 361 genes showing altered gene regulation in the stress across the developmental stages. These are clustered into 3 groups representing 162, 112, and 87 genes based on their expression pattern and were shown in clusters 8, 9 and 10 respectively. Cluster 8 (Fig. 27b, c) reveals expression pattern of 162 genes showing at least 3 fold up-regulation from day 45 to day 59 and their after these genes show at least 4 fold down-regulation as the development stage progresses to seed setting stage (day 71) and the same set of genes show more or less constitutive expression by day 75. The data in cluster 9 (Fig. 27d, e) reveals a total of 112 genes show 4 fold down regulation by day 52 and thereafter these subset of genes show at least 10 fold up-regulation by day 59. Further these genes start to down-regulate >40 fold by day 71 and thereafter show constitutive expression by day 75. A subset of 87 genes represented in cluster 10 (Fig. 27f, g) shows no change in their expressional pattern all the way from day 45 to day 75, but shows at least 3 fold up-regulation at day 59. The lists of interesting genes from this part of the analysis are given in S-A2L section of accompanied CD-ROM as supplementary tables.

4.5.3.1.2 Expression pattern from time series analysis of A3 (40% FC) leaf sample
The expression profiles (Fig. 28a) show that there are at least 251 genes falling into 2 clusters (cluster 9 and 10) which show altered gene expression under stress (40% FC) across the developmental stage from day 45 to day 75. Cluster 9 (Fig. 28b) represents 153 genes showing constitutive expression up to day 55 but shows at least 3
fold up-regulation by day 59 and thereon shows at least 4 fold down regulation as the development stage reaches to day 71, and thereafter shows no change in gene expression as by day 75 the physiological condition at this stage showed total drying up of the plant (Fig. 17). Cluster 10 represents 98 genes showing more or less constitutive expression till day 67, but shows at least 2 fold up-regulation on day 71 (Fig. 28c) and thereafter shows no change in the gene expression. The lists of interesting genes from this part of the analysis are given in S-A3L section of accompanied CD-ROM as supplementary tables.

**4.5.3.1.3 Expression pattern from time series analysis of A4 (15% FC) leaf sample**

The k-means clustering (Fig. 29a) analysis of data from A4 (15% FC) from day 45 to day 75 revealed that there are at least 498 genes showing altered gene expression. These genes fall under 4 clusters represented in cluster 7, 8, 9 and 10. Cluster 7 (Fig. 29b, c) represents 156 genes showing three subset of genes (7_1, 7_2, and 7_3). The subset 7_1 represents 69 genes showing no expressional changes across day 59, but shows at least 3 fold down regulation by day 63, interestingly these set of genes continue to up-regulate through day 67 and reaches a peak on day 71 representing at least 20 fold over expression compared to day 63 and this set of genes shows no change in expression by day 75 where the physiological condition of the plants was totally dry. Cluster 7_2 represents at least 47 genes showing more or less constitutive expression till day 59 and on day 63 and 67 shows a 2 fold down regulation compared to day 59 and by day 71 these genes show greater than 10 fold up-regulation by day 71. These genes drop suddenly in the same proportion (greater than 10 fold down regulation). In cluster 7_3, which represent 40 genes, show at least 3 fold changes on day 52, 59 and 71 and these genes show at least 4 fold down regulation on day 63 and 67 compared to day 59 or 7, as expected this set of genes show a sudden drop by greater than 4 fold by day 75.

Cluster 8 (Fig. 29d, e) represents 187 genes showing no expressional difference till day 55 and shows at least 3 fold up-regulation by day 59 and continue to be up to day 63, the same set of genes starts down-regulating greater than 10 fold as it reaches to day 71 and they again show no expression change on day 75 as expected. Cluster 9 (Fig. 29f, g) shows 74 genes which shows no considerable expression change till day 67 and shows greater than 10 fold up-regulation by day 71 and drops by at least 10 fold by day 75. Cluster 10 (Fig. 29h, i)
represents 81 genes showing no considerable change till day 55 and shows at least 7 fold up-regulation on day 59 and continue to be same till day 63 and thereafter starts down-regulating by at least 4 fold by day 67, and by >70 by day 71 and the expression of these genes on day 75 seem to be more or less constitutive. The lists of interesting genes from this part of the analysis are given in S-A4L section of accompanied CD-ROM as supplementary tables.

4.5.3.2 Cluster analysis of co-regulated genes across developmental stages from panicle tissue with in stress regime (Time series).

Sampling from panicle has been taken up from pre emergence of the panicle (day 59) till seed setting stage which includes samples from day 59, 63, 67, 71 and 75. The results of time series analysis of these samples with in each stress regime are described below.

4.5.3.2.1 Expression pattern from time series analysis of A2 (60% FC) panicle sample

The clustering of co-regulated genes (Fig. 30a) showed there are at least 388 genes which are shown to be differentially regulated; these genes are represented in clusters 7 and 10. The genes (309) in cluster 7 (Fig. 30b, c) shows a 3 fold increase by day 67 and continues to be same till day 71, and thereafter shows another 3 fold increase by day 75. In cluster 10 (Fig. 30d, e), 79 genes showing no considerable change till day 71, but as the 75 day sets, this set of genes shows a 3 fold up-regulation. The lists of interesting genes from this part of the analysis are given in S-A2P section of accompanied CD-ROM as supplementary tables.

4.5.3.2.2 Expression pattern from time series analysis of A3 (40% FC) panicle sample

The expression profiles (Fig. 31a) show that there are at least 1160 genes falling into 4 clusters (cluster 1, 3, 6 and 10) which show altered gene expression under stress (40% FC) across the developmental stage from day 45 to day 75. Cluster 1 (Fig. 31b, c) represents a set of 250 genes showing at least 2 fold change by day 71 compared to day 63, and by day 75 comes to a normal. Cluster 3 (Fig. 31d, e) represents 434 genes showing at least 3 fold change by day 71, and comes to normal by day 75. Cluster 6 (Fig.
31f, g) represents 260 genes showing a two fold change by day 71 and continue to be till day 75. Cluster 10 (Fig. 31h, i) represents a set of 217 genes showing at least 4 fold up-regulation from day 67 to 75. The lists of interesting genes from this part of the analysis are given in S-A3P section of accompanied CD-ROM as supplementary tables.

4.5.3.2.3 Expression pattern from time series analysis of A4 (15% FC) panicle sample

The k-means clustering (Fig. 32a) analysis of data from A4 (15% FC) from day 59 to day 75 revealed that there are at least 580 genes that are differentially expressed and are represented in clusters 1, 5, and 7. Cluster 1 (Fig. 32b, c) shows 221 genes showing a 2 fold drop in gene expression from day 59 to 63, and thereafter shows at least 4 fold up-regulation as the day 71 is reached of 221, 206 show day 75 this set of genes starts down regulating. But 15 of these genes continue to be over expressed till day 75. Cluster 5 (Fig. 32d, e) reveals 200 genes showing slight down-regulation pattern initially but shows a 3 fold increase by day 67 and thereafter continue to be down-regulated gradually. Cluster 7 (Fig. 32f, g) represents genes 161 genes shows a 2 fold up-regulation on day 67 compared to day 59, thereafter gene expression continue to be declined as it reached day 75. The lists of interesting genes from this part of the analysis are given in S-A4P section of accompanied CD-ROM as supplementary tables.

4.5.3.3 Cluster analysis of co regulated genes in leaf tissue across field drought stress regimes

The gene expression profiles captured from different stress regimes from day 45 to day 81 were analyzed for gene expression change across 3 different stress regimes viz. 60% FC, 40% FC, 15% FC referred as A2, A3, A4. The number of differentially expressed genes of field leaf samples at various drought stress levels at different days of development, and details of the pattern of gene expression at each data point from day 45 till day 75 is described below. The classes of genes and their functional roles are elaborated in discussion part of this thesis.

4.5.3.3.1 Expression Pattern of field leaf samples across stress regimes on day 45

This is the stage where the onset of drought has just began (7 days after stress initiation), the k-means clustering of these group of genes in A2 (60% FC), A3 (40% FC), and A4 (15% FC) (Fig. 33a), revealed at least 192 genes which are expressed
differentially in stress regimes. These include a set of 154 genes showing up-regulation by at least 3 fold in A3 and A4 compared to A2 represented in cluster 8 (Fig. 33 b, c). The other cluster which showed down-regulation by >3 fold, these differ in the expression pattern across A3 (40% FC) and A4 (15% FC). A set of 38 genes seem to be down-regulated by at least 3 fold in A4 compared to A2 and A3 and are grouped in cluster 10 (Fig. 33 d, e). The lists of interesting genes from this part of the analysis are given in S-45L section of accompanied CD-ROM as supplementary tables.

4.5.3.3.2 Expression Pattern of field leaf samples across stress regimes on day 52

The differentially expressed genes observed through gene expression profile plots (Fig. 34a) shows at least 393 genes are up-regulated falling into four different clusters (cluster 4, 5, 7 and 10) and three clusters (6, 8 and 9) which are down regulated. In cluster 4 (Fig. 34b, c) representing 43 genes shows a 3 fold change from A2 to A3 and the same set of genes seem to have no considerable expression change in A4. A set of 37 genes shows no considerable expression change from A2 to A3, but shows a 3 fold increase from A2 to A4 in cluster 5 (Fig. 34d, e). A total of 212 genes have two kinds of expression patterns in cluster 7. One subset of 182 genes in groups 1 and 2 show more or less similar kind of profile plots and hence were clubbed in one cluster (7_1 & 2), the expression graph shows that these genes are consistently up-regulated across stress regimes (Fig. 34f, g), whereas the other group of genes (30) in cluster (7_3) shows a 3 fold increase in A3 to A4. The other cluster, 10 (Fig. 34h, i) with 101 genes, shows an at least 4 fold increase from A2 to A3 and continue to have similar expression in A4. Among the down regulated genes 527 of them seem to be down-regulated and are represented in cluster 6, 8 and 9. Cluster 6 (Fig. 34j, k) represents 182 genes which are down-regulated only in A3 by 2 fold. Cluster 8 (Fig. 34l, m) represents 204 genes showing 2 fold down-regulations in A3 and A4. Cluster 9 (Fig. 34n, o) represents 141 genes showing a consistent down-regulation from A2 to A4 by at least 2 fold. The lists of interesting genes from this part of the analysis are given in S-52L section of accompanied CD-ROM as supplementary tables.
4.5.3.3 Expression Pattern of field leaf samples across stress regimes on day 55

This is the end of vegetative stage, the expression profile plots (Fig. 35a) reveals that there are several clusters obtained at this stage that show complex regulation, altogether there are 8 clusters (fall into 3 broad clusters cluster 7, cluster 9 and cluster 10) that are up-regulated differently across stress regimes, and 6 clusters (also fall into 3 broad clusters cluster 5, cluster 6 and cluster 8) which are down regulated in different patterns. This kind of expression pattern is observed only at this stage, and the reasons for these may be due to change in stage of the rice life-cycle. Clusters 7, 9 and 10 shows different sub set of genes within these clusters, which co-regulate differentially and all of these genes which are differentially expressed are up regulated > than 3 fold as seen in profile plots and heat maps (Fig. 35b, c, d, e, f & g). Genes (38) in cluster 7_3 (Fig. 35h, i) are 2 fold up-regulated in A3, A4 and as stress increases. In cluster 9_1, 57 of genes are up regulated >2 fold in A3 and as the severity of stress is increased the transcripts levels remain unaltered, as seen in profile plot and heat map of cluster 9_1 (Fig. 35j, k). There are another group of 8 genes which are up regulated only under severe stress (A4), these set of genes show no changes in their transcripts levels in A3 as observed in the profile and heat maps of cluster 10_2 (Fig. 35l, 35m), the genes under these cluster are up regulated > 3 fold in A4. Cluster 10_3 in which 27 genes initially down-regulate by one fold in A3 and in A4 the same set of genes are up-regulated by more than 3 folds. Clusters 9_2 and 9_3 show that 46 genes that are up-regulated in A3 and are not as much as observed in A4, this kind of profiles show that these genes might play a role in adaptation rather than tolerance. A set of 58 genes which show no altered regulation in A4, but show altered regulation in A3 are observed in cluster 10_1. These genes are 2 fold down regulated in A3. The cluster plots shows there are at least 5 sub clusters that are down-regulated between 2 to 3 fold which fall into 3 major clusters i.e. cluster 5, cluster 6 and cluster 8 (Fig. 35n, o, p, q, r & s). The set of 95 genes in cluster 5_3, 5_4 and 5_5 show initial up-regulation in A3 and the same down-regulate in A4. In cluster 6_2, 52 genes are down regulated only in A4. The only cluster showing down-regulation of 77 genes in all the stress regimes are in cluster 8_2 and 8_3. The lists of interesting genes from this part of the analysis are given in S-55L section of accompanied CD-ROM as supplementary tables.
4.5.3.3.4 Expression Pattern of field leaf samples across stress regimes on day 59

The rice life cycle has finished its booting stage and the clustering for co-regulated genes at post-booting stage revealed that most of the differentially expressed genes show up-regulation, the variation in their expressional pattern is shown as cluster plots in Fig. 36a. There are altogether 4 up-regulated clusters showing various pattern of gene expression which fall into two major clusters i.e. cluster 8, cluster 10. Cluster 8 has 178 genes which show atleast 3 fold up regulation in A3 compared to A2 and the same expression level also seen in A4 (Fig. 36b, c). Cluster 10, has been further k-means clustered into 3 sub-clusters (Fig. 36d, e, f, g, h & i), genes in these clusters show 3 to 7 fold up-regulation. 34 Genes in cluster 10_1 are 3 fold up-regulated (Fig. 36d, e) only in A4 and show no expressional difference in A3, the other subsets of this cluster show 7 fold up-regulation, the genes (22) in cluster 10_2 slightly down regulate in A3 and under severe stress these are seen to be up-regulated more than 7 fold (Fig. 36f, g), and in cluster 10_3, 20 genes in this cluster are shown to up-regulate >5 fold only in A4 (Fig. 36h, i). There exist atleast 19 genes which show down regulation all along the stress gradients are observed in cluster 6, which shows a several fold down regulation from A2 to A3 and also in case of A4 (Fig. 36j, 36k). Another set of ~200 genes from the same cluster shows similar pattern but fold change is only 2 fold (Fig. 36l, m). The lists of interesting genes from this part of the analysis are given in S-59L section of accompanied CD-ROM as supplementary tables.

4.5.3.3.5 Expression Pattern of field leaf samples across stress regimes on day 63

The k-means clustering of co-regulated genes into 10 clusters shows that there are not much complex expressional patterns at this data-point under study. The expression profile plots of the clusters (Fig. 37a) reveal 4 clusters which are up-regulated between 3 to 5 fold, and 2 clusters which are down regulated between 3 to 4 fold. The clusters which are up-regulated >3 fold, are cluster 6 (Fig. 37b, c) and 9 (Fig. 37d, e). The genes in cluster 6 show different expressional patterns and fold changes of > 3 and 5, and have been further sub-clustered into 3 subsets (cluster 6_1, 6_2 and 6_3). The genes (16) in cluster 6_3 continue to up-regulate across stress regimes, whereas the same subset of genes in cluster 6_1 (86 genes) and 6_2 (93 genes) up-regulate only in A4.
other cluster which is up-regulated > 5 fold is cluster 9 (140 genes) (Fig. 37d, e). These
genes are up-regulated only in A4; in A3 these are slightly down-regulated. Cluster 10, 8
and 7 show down regulation of 11, 176, and 152 genes respectively, to the level of > 3
to fold. The expression plots and heat maps of these down regulated genes are shown in Fig.
37f, g, h, i, j & k). The lists of interesting genes from this part of the analysis are given in
S-63L section of accompanied CD-ROM as supplementary tables.

4.5.3.3.6 Expression Pattern of field leaf samples across stress regimes on day 67

The cluster analysis revealed that there are 2 clusters which show up-regulation, 3
clusters which are down-regulated and 2 clusters which are up and down regulated (Fig.
38a), the cluster 10 (176) (Fig. 38d, e) and cluster 7 (154 genes) (Fig. 38b, c) each having
a set of genes ~ 330, which are shown to be up-regulated >5 fold. The difference in these
2 clusters is the genes in cluster 10 continuously up-regulate across all stress regimes
whereas in cluster 7, the genes in A3 do not show any expressional changes and the
same in A4 are up-regulated > 5 fold. The clusters which show down regulation >2.5 fold
are 211 genes in cluster 5 (>2 fold down regulation) these are down regulated across all
the stress regimes. In cluster 8 and 9, 149 and 167 genes show down-regulation of genes
>5 fold. The differences between these two clusters are, in cluster 9 (Fig. 38h, i) the
genes in A3 are not down regulated, but the same in A4 these are down-regulated > 5
folds, whereas the genes in cluster 8 (Fig. 38f, g) continue to down regulate as stress
increases, these are shown to down-regulate > 2.5. The group of genes (187) in cluster 4
(Fig. 38j, k) show that 187 genes in A3 are down regulated by 2 folds but the same set of
genes show no expressional changes in A4, contrarily 20 genes in cluster 6 (Fig. 38l, m)
up-regulate in A3 and show no expressional changes in A4. The lists of interesting genes
from this part of the analysis are given in S-67L section of accompanied CD-ROM as
supplementary tables.

4.5.3.3.7 Expression Pattern of field leaf samples across stress regimes on day 71

The expression profile plots (Fig. 39a) show very interesting patterns of gene
regulation in A4, the phenotype at this stage in A4 as observed was fully lodged (Fig. 16)
The clusters 10_1 (Fig. 39b, c) and 10_2 (Fig. 39d, e) show few genes (17 and 6) which
are highly up-regulated (>50 and 300 folds). This stage is represented by most of the genes which are highly up-regulated, among the other highly expressed genes, are observed in A4 are represented in cluster 8 (86 genes) (Fig. 39f, g) which are up-regulating (> 10 fold), cluster 4 (sub-clusters 4_1 (98 genes) (Fig. 39 h, i), 4_2 (91 genes) (Fig. 39 j, k), and 4_3 (52 genes) (Fig. 39 l, m) which are expressed > 4 fold. The genes in cluster 4_1 and cluster 4_2 do not show altered regulation in A3 but are only observed in A4, since the plants are fully lodged during this stage in A4, these genes may have a major role during the death of the plant henceforth these are over expressed to such high levels. The only clusters where, 197 and 72 genes are up-regulated >3 and 4 fold are observed in cluster 7 (Fig. 39n, o), and cluster 9 (Fig. 39p, q), the same set of genes in A4, show no expressional changes in cluster 7, whereas they are 20 fold down-regulated in cluster 9. The only cluster where 173 genes (Fig. 39r, s) continue to down-regulate across A3 and A4 are observed in cluster 6. The lists of interesting genes from this part of the analysis are given in S-71L section of accompanied CD-ROM as supplementary tables.

4.5.3.3.8 Expression Pattern of field leaf samples across stress regimes on day 75

This stage is the end of seed setting stage and the rice plant completes all its developmental stages, the drought stress phenotypic response as observed under different field capacities show that the plants in A4 and A3 have fully lodged (Fig. 17). The expression profiles of all genes are shown in Fig. 40a. These contain 3 clusters which are up-regulated from 2 to 4 fold, and 3 cluster which are down-regulated from 2 to 5 fold. Among the clusters which are up-regulated > 4 fold, 165 genes in cluster 9 (Fig. 40b, c) continue to over express across all stress regimes. Cluster 6 (Fig. 40d, e) represent a large set of genes (241) which are up-regulated >2 fold in both A3 and A4. Cluster 3 (Fig. 40f, g) represent genes (373) that are over expressed > 3 fold in A4 whereas they remain unaltered in A3. The clusters which are down-regulated include clusters 5, 8 and 10, the genes in cluster 10 (77 genes) (Fig. 40l, m) and cluster 5 (237 genes) (Fig. 40h, i) continue to down regulate by > 5 fold in the former and >2 fold in the latter. The cluster 8 (Fig. 40j, k) represent genes (171 genes) which are down-regulated only in A4, and show no altered regulation in A3. The lists of interesting genes from this part of the analysis are given in S-75L section of accompanied CD-ROM as supplementary tables.
4.5.3.4 Cluster analysis of co-regulated genes in panicle tissue across field drought stress regimes from day 59 to day 75

4.5.3.4.1 Expression Pattern of field panicle samples across stress regimes on day 59

The phenotypic response of panicle booting varied from day 52 till day 58, panicle samples were taken at the pre-emergence stage, on day 59. The cluster profiles (Fig. 41a) of the co-regulated genes showed clusters 5, 6 and 10 which were up-regulated > 2 fold. In cluster 5 (Fig. 41b) 6 genes continue to up-regulated across the stresses, wherein in cluster 6 (14 genes) (Fig. 41c) the genes are up-regulated only in A3 and show no expressional changes in A4. A large group of genes (82 genes) are over expressed > 2 fold up in A4 (cluster 10, Fig. 41d), besides these 12 genes show >3 fold up-regulation. The genes which are down regulated all along the stress gradients are observed in cluster 7 and 9 (17 genes and 10 genes) which continue to be down-regulated by 2 fold (Fig. 41e, f). Though these are only a small sub-set of genes, these are shown to down-regulate > 2 fold. The lists of interesting genes from this part of the analysis are given in S-59P section of accompanied CD-ROM as supplementary tables.

4.5.3.4.2 Expression Pattern of field panicle samples across stress regimes on day 63

The k-means clustering of co-regulated genes into 10 clusters shows that there are not much complex expressional patterns at this data-point under study. The expression profile plots of the clusters (Fig. 42a) reveal 3 clusters which are up-only by 2 fold, and 3 clusters which are down regulated by 2 fold. The clusters which are up-regulated are clusters 5, 9 and 10. A large number of genes (201 genes) are over expressed >2 fold in cluster 5 (Fig-42b) across all stress regimes. The genes (133 genes) in cluster 10 (Fig. 42d) continue to up-regulate across stress regimes. The only other cluster which is up-regulated > 2 fold is cluster 9 (12 genes) cluster 9 shows 2 fold up-regulation in A3 and the same set of genes (92 genes) seems to be unaltered in A4 (Fig. 42c) contrarily 8 genes in cluster 8 show >2 fold up-regulation in A3 and remain unchanged in A4 (Fig. 42e). These genes are up-regulated only in A3, and show no expressional changes in A4. Cluster 4, 6 and 7 show down regulation of 15, 214, and 166 genes, to the level of >2 fold respectively. The expression plots of these down regulated genes are shown in Fig. 42f, g,
The genes in cluster 4 show altered regulation only in A4, whereas in cluster 6 they show down-regulation in both A3 and A4. In cluster 7 all the genes continue to down-regulate across all the stresses. The lists of interesting genes from this part of the analysis are given in S-63P section of accompanied CD-ROM as supplementary tables.

4.5.3.4.3 Expression Pattern of field panicle samples across stress regimes on day 67

The cluster analysis revealed that there are 4 clusters which show up-regulation, 3 clusters which are down-regulated (Fig. 43a), the cluster 2 (Fig. 43b) consisting of genes ~300, which are shown to be up-regulated >3 fold. These genes in cluster 2 continuously up-regulate across all stress regimes, whereas in other clusters they are up-regulated by >2 fold and they have different patterns of expression at different stress levels. In cluster 4 (Fig. 43c) the genes (295) in A3 do not shown any expressional changes and the same in A4 are up-regulated, contrarily genes (154 genes) in cluster 7 (Fig. 43e) are over expressed >3fold only in A4. The only cluster where, the genes (255 genes) are up-regulated both in A3 and A4 is in cluster 6 (Fig. 43d). The clusters which show down regulation >2 fold are 146, 103, and 127 genes in cluster 8, 9 and 10. In cluster 9 (Fig. 43g) shows genes that are down regulated across all the stress regimes by > 3 fold. In cluster 8 and 10 (Fig. 43f, h) genes are down-regulated only in A3 by >4 fold. The lists of interesting genes from this part of the analysis are given in S-67P section of accompanied CD-ROM as supplementary tables.

4.5.3.4.4 Expression Pattern of field panicle samples across stress regimes on day 71

The expression profile plots (Fig. 44a), show a large number of genes are up-regulated in various patterns. As observed in leaf expression profiles, during this stage is represented by most of the genes which are highly up-regulated, in panicles though these are not expressed to that levels as that of leafs however, a large proportion of genes are showing up-regulated. Among the clusters which are up-regulated >2 fold are clusters 5, 6, 8 and 10. The expression profile plots show that in cluster 5 (Fig. 44b) genes (229) are seen over expressed only in A4, whereas the genes (136) in cluster 10 (Fig. 44e) show up-regulation only in A3 (>2 fold) and as stress increases the same set of genes in A4 are down-regulated (>3fold). Cluster 6 (Fig. 44c) is the only cluster which contain a large proportion of genes (133) that are seen to up-regulated >3 fold, both in A3 and A4.
cluster 8 (Fig. 44d) the genes (212) continue to up-regulate across all the stresses another subset of these genes 4 genes cluster are over expressed >4 fold. The genes which are down-regulated at this stage is not more than 2 fold and only 2 clusters show down-regulation of genes in different patterns. The cluster in which the genes (183) continue to down-regulate across A3 and A4 are observed in cluster 9 (Fig. 44g). In cluster 7 (Fig. 44f) the genes (238) show down-regulation both in A3 and A4. The lists of interesting genes from this part of the analysis are given in S-71P section of accompanied CD-ROM as supplementary tables.

4.5.3.4.5 Expression Pattern of field panicle samples across stress regimes on day 75

This stage is the end of seed setting stage and the phenotypic response in A4 was relevant with no sign of seed-filling, and in A3 the seed did not set at all, in A2 the overall yield loss was 20% as compared to control. The k-means hierarchical clustering revealed (Fig. 45a) there, were at least 4 clusters which showed up-regulation > 2 and very few genes (5 genes) which showed expressional changes >10fold (Cluster 6_1; Fig. 45c). The other clusters which showed >2 fold up-regulation are clusters 3, 6_2 and 8. In cluster 8 (Fig. 45e) 125 genes were over expressed in A3 but did not show any altered regulation in A4, whereas in cluster 3 (Fig. 45b) 343, genes were up-regulated only in A4. The only cluster in which the genes (240 genes) continued to up-regulate are in cluster 6_2 (Fig. 45d). The clusters which are down-regulated include clusters 7, 9 and 10, the genes in cluster 10 (109 genes) (Fig. 45h) continued to down regulate by > 4 fold across all the stresses. In cluster 9 (Fig. 45g) 92 genes were down regulated in both A3 and A4, whereas in cluster 7 (Fig. 45f), 251 genes were down regulated in A4 by >3 fold and in A3 there was no considerable change in the expression level. The lists of interesting genes from this part of the analysis are given in S-75P section of accompanied CD-ROM as supplementary tables.