Functional Genomics of Drought Tolerance: Large Scale EST Generation, Annotation, Physical Mapping, and Gene Expression Profiling in Rice *(Oryza sativa* sub sp., *indica cv. Nagina 22)*

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

By

Markandeya Gorantla

Department of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad 500 046
Andhra Pradesh, India

Enrollment No. 01LPPH06

December 2005

Department of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad-500 046
INDIA

CERTIFICATE

This is to certify that **Markandeya Gorantla** has carried out the research work embodied in the present thesis entitled "**Functional Genomics of Drought Tolerance: Large Scale EST Generation, Annotation, Physical Mapping and Gene Expression Profiling in Rice *(Oryza sativa* sub sp., *indica cv. Nagina 22)"*** for the degree of **Doctor of Philosophy** under my supervision in the Department of Plant Sciences, School of Life Sciences, University of Hyderabad.

This work has not been submitted for the award of any degree or diploma of any other University or Institute.

Prof. Arjula Ramachandra Reddy
Supervisor

Head
Department of Plant Sciences

HEAD
Dept. of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad - 500 046. INDIA

Dean
School of Life Sciences

30/12/05

Dean, School of Life Sciences
University of Hyderabad,
Hyderabad-500 134. (India)

DECLARATION

I hereby declare that the work presented in this thesis entitled "**Functional Genomics of Drought Tolerance: Large Scale EST Generation, Annotation, Physical Mapping and Gene Expression Profiling in Rice *(Oryza sativa sub sp., indica cv. Nagina 22)* "** has been carried out by me under the supervision of **Prof. Arjula Ramachandra Reddy** in the Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad- 500 046, and that this work has not been submitted for any degree or diploma of any other University or Institute. All the assistance and help received during the course of the investigation have been duly acknowledged.

G. Markandeya

Markandeya Gorantla

ARReddy

Prof Arjula. Ramachandra Reddy
Supervisor
Department of Plant Sciences

This thesis is dedicated to

Sri Arjula Ramachandra Reddy

Who foresaw a bright future in me, bought me recognition, respect, and made my life at its best!

Acknowledgements

I express my sincere gratitude to my advisor, ***Professor Arjula Ramachandra Reddy,*** *for his constant guidance, team spirit, friendship, and moral support throughout this research and supervising this thesis with such great interest. I was always welcome to discuss my work and future direction with him. Though sometimes my views did not match his, he always let go free and kept my spirits high. I'm indebted to him for my life time and promise to carry his sense of attitude in my future scientific endeavors. He has been an excellent mentor, introducing me as a scientist, encouraging and getting me funds to work in his lab and in United States. The scientific knowledge and training in advanced technological innovations I gained during my stay with Prof Reddy and his research group has helped me in my scientific endeavors at its best.*

I would like to express my sincere thanks to ***Dr. John C O'Toole, Associate Director, the Rockefeller Foundation NY., USA.*** *For encouraging and helping me to participate in a collaborative research work at University of Georgia through extending financial assistance under* ***RF project grant*** *to* ***Prof Arjula R. Reddy****. I sincerely acknowledge the financial assistance from the* ***Rockefeller Foundation Projects*** *to* ***Prof. A. R Reddy****.*

I am thankful to our collaborators ***Prof Andrew Paterson****,* ***Dr. Alex Feltus*** *for extending a collaborative research support for gene expression profiling using cDNA microarrays at* ***University of Georgia*** *and making my research a success and also all the members of* ***plant genome mapping laboratory*** *for their support and making my stay a pleasant and memorable one at* ***UGA****.*

I am thankful to our other collaborators ***Prof Jeffrey L. Bennetzen, Dr Ramakrishna, Dr. M.K. Reddy*** *for assisting and giving me an opportunity to develop genetic resources for the present study on functional genomics of drought tolerance in rice.*

I am thankful to the former Deans, School of Life Sciences, ***Prof. T. Suryanarayana*** *and the present Dean,* ***Prof. A. S. Raghavendra****, for extending facilities available in the School for my research work. I would also like to thank the former Head, Department of Plant Sciences,* ***Prof. A. S. Raghavendra****,* ***Prof. M.N.V. Prasad*** *and the present Head,* ***Prof. P.B. Kirti****, for their help and cooperation in executing the work. I am grateful to* ***Prof Appa Rao Podile*** *and his student* ***Rajesh Chavli*** *for their support and help during my research at JNTU and UGA. I thank all the* ***non-teaching staff*** *of the School for their help and assistance.*

I am thankful to ***Dr N Seetharama*** *former senior scientist* ***ICRISAT****, for introducing me to* ***Prof Arjula R Reddy*** *and helping me join his research team and my former teachers* ***Prof. C.Manohara Chary, Prof K.V.B.R. Tilak, Prof Seenaiah, Prof Radhakrishniah, Prof Durbash Karan, Prof Laxmi Narasu, and all my teachers*** *for their constant encouragement and guidance during my academic career.*

I wish to thank my colleagues ***Vijaya Bhasker Lachagari*** *for his support and help during microarray analysis in developing in house perl scripts, databases, and solving all critical issues of microarray analysis and also for assisting me during field drought experiment. I thank my colleagues* ***Dr. B. Chandra Shekar*** *for brushing me up the basic molecular biology,* ***Dr. Nagabushana Ithal*** *for his*

suggestions and help me trouble shoot for high quality EST generation, ***Dr Ravindrababu Peram*** *for utilization of EST datasets,* ***L.Shivaram Prasad*** *, and* ***B.Venkat Reddy*** *for extending their working hands whenever needed,* ***A.Madan Mohan Reddy*** *for his working hands and hospitality during my stay at the campus hostel,* ***A.Chandrasekar***, ***V.Kiran Kumar***, ***G.Linga Reddy***, ***Maheedhar,*** *for their support during my research. I also wish to thank our past research group* ***Dr. B. Karunasree*** *,* ***Dr G. Madhuri*** *for their encouragement. I would like to thank the assistance received from* ***Raju, Ravinder, Poorna, Mallesh,*** *and* ***Narasimha*** *in the lab whenever needed.*

I sincerely acknowledge the ***Rockefeller Foundation,*** *for Research Assistantship, Visiting Scientist fellowship,* ***CSIR,*** *for Senior Research Fellowship and* ***DBT*** *for Research Associateship,* ***University of Hyderabad*** *for UPE travel grant award for attending PAG conference at USA, Facilities through* ***DST-FIST*** *funding to the Department of Plant Sciences and the* ***UGC-SAP*** *to School of Life Sciences are gratefully acknowledged.*

I thank all my ***friends*** *for their constant support and great time spent outside my laboratory with fond memories.*

I thank my ***parents, sister, and brother*** *for their patience and support. I also thank my near and dear* ***Chavli Rao and family, Pattabhi Rama Rao and family, Hari Singh and family, Lohithaswa and family****. I thank my Father-in-law* ***Sankarlingam Natarajan*** *who believes my potential to excel, his family and my wife* ***Sowmya S Gorantla*** *who joins me at the end to race ahead of time and I am glad that she is always there for me with lots of love and affection!*

Last but not the least I wish to express my sincere thanks and respect to ***Shri Kaukuttla Dharma Reddy & Sons.*** *I owe a debt of gratitude for their love, supporting spiritually, emotionally and financially, and understanding me along the path of my longest academic pursuits. The eldest* ***K, Anantha Reddy*** *who never waited to make me think that I have to ask god something,* ***Rajender*** *who never stepped back to support me in the hour of need, and all my* ***Bhabhis*** *for their 15 years of warm hospitality, my nieces and nephews* ***likitha, sushma, sunny, Chikki, Vicky****, and* ***Nanu*** *for their love.*

Special thanks *to my little loving niece* ***Akshitha*** *for being such an understanding and loving child. Who made me understand the word called* ***"smile"*** *brings in smiles in me every weekend and during the times when my thesis felt overwhelming and makes me forget all my worries to begin a new week with great success. I couldn't have asked for more.*

Date: Monday, Dec 26th, 2005 ***Markandeya Gorantla***

Contents

List of Tables

Supplementary Tables available in CD-ROM

Abbreviations

EST	: Expressed Sequence Tags
ABA	: Abscissic acid
SAGE	: Serial analysis of gene expression
SRG	: Stress responsive genes
MAS	: Marker-assisted selection
GSTs	: Glutathione-S-transferase
G X E	: Genetic and Environmental effects
BAC	: Bacterial artificial chromosome
PAC	: P1 derived artificial chromosomes
SNP	: Single nucleotide polymorphisms
PHRED	: Phil's Read edit data
CAP3	: Contig Assembly program version 3
TIGR	: The Institute of genomic research
QTL	: Quantitative tract loci
MPSS	: Massive Parallel Signature Sequencing
SDS	: Sodium dodecyl sulphate
BSA	: Bovine serum albumin
TSA	: Tyramide signal amplification
SVD	: Single value decomposition
MGED	: Microarray gene expression data
MIAME	: Minimum information about microarray experiment
N22	: Nagina 22
IRGSP	: International Rice Genome Sequencing Project
BGI	: Beijing genomics institute.
$CaCl_2$	: Calcium chloride
DREB	: Dehydration responsive element binding factor
CRT/DRE	: C repeat dehydration responsive element
PEG	: Poly ethylene glycol
DMSO	: Dimethyl sulphoxide
HCL	: Hydrochloric acid
NaOH	: Sodium hydroxide
ATP	: Adenosine tri phosphate
ROS	: Reactive oxygen species
APX	: Ascorbate peroxidase
H_2O_2	: Hydrogen peroxide
MAPK	: Mitogen activated protein kinase
EREBP/AP2	: Ethylene responsive element binding factor

1. INTRODUCTION

1. Introduction

Rice is the most important food crop as it is consumed by more than half of the world's population. It is predicted that production of 40% more rice is needed by the year 2030 to meet the demand of 5.0 billion rice consumers without affecting the natural resources, that is, cultivable land and water. This needs to be achieved by utilizing limited cultivable land and improved varieties which consume less water and less fertilizer (Khush, 2005). Dating back to 10,000 yrs since its cultivation began (Normile, 1997), rice has emerged as the most important cereal food crop for humans. The genus *oryza* includes two domesticated species *O.sativa* and *O.glaberima*, the former being the most cultivated species across the world while the latter is mostly grown in West Africa. Diverse climates, seasons, soils and varied cultural practices have all contributed to an enormous ecological diversity among rice cultivars. Rice growing areas span across tropics, sub tropics, semiarid tropics, and also the temperate regions of the world.

The predominantly rice growing areas in Asia (~130 million hectares) are often threatened by severe abiotic stresses, the most common being drought. Abiotic stresses such as drought, cold and high salt significantly limit cereal productivity thereby reducing average yields of most of the major crops (Bray et al., 2000). Due to inconsistent and erratic rainfall, drought is spreading across all crop growing regions and is expected to lead to serious salinization of more than 50% of arable lands by the year 2050 (Wang, 2003). Due to such a water limitation abiotic stresses cause crop losses as much as 50% (Boyer, 1982; Bray *et al.*, 2000) in some years, and drought alone may cause yield losses of as much as 15% (Dey *et al.*, 1996). Drought spells across Asia have become more frequent and severe, leading to irregular and insufficient irrigation of rice crop and depletion of ground water resources leading to 100% yield losses in certain areas.

Drought tolerance is a complex trait in plants controlled by many genes that are characterized by incomplete penetrance, epistatic interactions, and quantitative inheritance and therefore fits well as a model trait for global gene expression studies. Developing drought tolerant rice lines by conventional breeding was the most

commonly used approach to combat the problem of drought stress induced yield losses. However, breeding for drought tolerance in rice has been rather slow because of lack of information on mechanisms through which plants develop tolerance, limited knowledge of inheritance of resistance or tolerance, low heritability of the trait, and lack of extensive information on associated genes and their regulation. Genomic tools have provided new avenues to investigate and unravel allelic variation in target genes or genomic segments associated with drought tolerance. However, there is about 1% increase on an average in yield potential since last 35 years through conventional hybridization and selection procedures (Peng et al., 2000), and a large number of cultivars and hybrids have been developed to cope up with demand by an ever increasing population. There are over 100,000, accessions of traditional rice varieties collected from a range of geoclimates (Khush, 2005). Thus, these diverse species are a rich source of genes for wide environmental adaptability under different abiotic and biotic stress factors. Some of these genes even might have been lost or modified beyond recognition during the long period of cultivation (IRRI 1997). This enormous wealth of rice germplasm has largely remained unexploited for a host of agronomically important genes owing to difficulty in availability of technologies to unravel the allelic variations existing among these cultivars.

A crucial step in that direction is to decipher molecular genetic basis of drought tolerance. Identification of genes and their biological function in drought response processes require large-scale genomic and genetic resources. Rice improvement for better abiotic stress adaptation along with high yielding properties will require integrated approaches encompassing diverse germplasm, traditional breeding, modern technologies and emerging knowledge from comparative genomics (Khush, 2005).

The past decade has revolutionized plant genomics perspective and has greatly improved our ability to understand the genetic makeup of several plant species. The completion of whole genome sequence of Arabidopsis (The Arabidopsis Genome Initiative, 2000) and near completion of genome sequences of two rice subspecies indica and japonica. (Feng et al., 2002; Goff et al., 2002; Sasaki et al., 2002; Yu et al., 2002; The Rice Chromosome 10 Sequencing Consortium, 2003; IRGSP, 2005) spectacularly improved the tools and precision of genetic analysis in those plants. This

exciting new genetic information enabled plant biologists to take multidisciplinary integrative approaches instead of the usual analytical approaches to discover gene organization and functional regulation of biological processes. This will eventually lead to precise association of a phenotype to a genotype.

Rice is a model crop for cereal genome research in lieu of its small genome size (430Mb), and availability of dense physical maps, genetic maps and rich EST resources. Rice naturally offers several advantages for molecular genetic analysis and has now emerged as the natural choice of experimental system to understand the genetic basis of complex traits such as yield, hybrid vigor, biotic, and abiotic stress tolerance to study syntenic relations with other cereals (Devos and Gale, 1997). Complex traits like drought tolerance are now amenable to a detailed molecular analysis using such genomic tools and high-throughput genomic technologies. To date the major chunk of the data is on Expressed Sequence Tags (ESTs) generated through large-scale cDNA sequencing projects. This has led to the identification of many novel genes associated with tissue specific expression (Adams et al., 1995). The EST resources have been extensively used to analyze changes in gene expression controlling physiological processes, such as responses to biotic or abiotic stresses (Jiang et al., 2000). Expressed Sequence Tags have been generated on a large scale in rice which are a valuable tool to catalogue eventually all genes (Uchimiya *et al.*, 1992; Umeda et al., 1994; Yamamoto and Sasaki, 1997; Reddy et al., 2002; Markandeya et al., 2002, 2003; Zhang et al., 2005) and in deciphering the role of transcriptionally regulated genes in different tissues (Ewing *et al.*, 1999). However, only a few studies focused on the analysis of transcriptome profiles of rice seedlings subjected to abiotic stress (Umeda *et al*., 1994; Matsumura *et al*., 1999; Kawasaki *et al*., 2001) or drought (Babu et al., 2002; Markandeya et al., 2005;). Expressed Sequence Tags provide the most direct approach for discovering genes associated with stress response. This has been demonstrated in several plant systems (Michale et al., 2002; Fernandes et al., 2002; Echenique et al., 2002; Reddy et al., 2002; Markandeya et al., 2005).

Rice has a deep EST coverage in general, and relatively a large collection of ESTs generated from drought stressed plants have become available (Reddy et al 2002). These resources are valuable for carrying further experiments using microarray and also

for SNP analysis for discovering the specific alleles of target genes associated with drought tolerance phenotype. A comparison of sequences of expressed genes in different plant species would allow identification of functional genes that are related by descent, though they might have become orthologous and have divergent functions. Transcript mapping of candidate genes is a powerful tool to elucidate the underlying mechanisms of drought tolerance in diverse crop species. In rice and other crops, localization of trait-associated ESTs will lead to molecular dissection of QTLs through functional characterization of genes (Yano et al., 2001). In fact localizing rice ESTs to rice physical map provides a direct route for gene discovery (as opposed to *in silico* gene prediction) and elucidation of gene structure (Kan et al, 2001).

Numerous putative drought responsive genes have been uncovered by genome wide expression analysis strategies in rice (Matsumura et al., 1999, 2003; Kawasaki et al., 2001; Rabbani et al., 2003; Gibbings et al., 2003; Gowda et al., 2004 and Markandeya et al., 2005). Most of these are basically dehydration associated expression profiles of rice conducted under laboratory conditions. However, since the experiments were rigorously controlled and environmental variables are kept at a minimum, the expression analysis provided valuable information about stress response associated genes in rice. Gene profiling through generation of HTP EST sequences is rapid and economical, and can be efficiently used to identify differentially regulated mRNAs (Zhu et al., 2001). Microarray technology has become a powerful tool for genome wide gene expression studies (Eisen and Brown, 1999; Schena et al., 1995). It is an efficient way of bridging the gap between sequence information and functional genomics (Bohnert et al 2003). First report on using ESTs for genome wide expression was the analysis of 48 arabidopsis genes with differential expression in roots and shoots (Schena et al., 1995). Apart from applications in the field of genetic and physical mapping, ESTs are the central resource for the analysis of gene expression with the help of high-density arrays, as demonstrated for *Arabidopsis* (Schena *et al*., 1995; Girke *et al*., 2000; Schenk *et al*., 2000), barley (Ozturk *et al*., 2002) maize (Wang *et al*., 2003), and rice (Kawasaki *et al*., 2001; Rabbani *et al*., 2003).

The present study is aimed at discovering genes associated with drought stress, their sequence and expression in rice. In an effort to catalogue and categorize the

expression of genes associated with stress response, and to identify a suite of promising candidate genes involved in drought response, I have generated Expressed Sequence Tags from a normalized cDNA library previously constructed from drought stressed indica rice (Nagina 22) seedlings (Reddy et al., 2002). This library served as rich source of non-redundant cDNA clones and therefore was used on a continuous basis to generate ESTs in my gene discovery research. I have analyzed and annotated the ESTs thus generated, and deciphered the drought stress transcriptome, and constructed a unigene set, and a physical map. Expression studies were carried out using cDNA microarrays on field drought stress conditions to decipher the role of novel genes identified to define them as candidate genes. This study to our knowledge is the first elaborately conducted field drought stress gene expression profiling experiment on an indica rice line. I have also used pearl millet ESTs on the microarray for discovering orthologous genes associated with drought tolerance. Through this study, I have identified several genes, including novel ones that responded differentially to drought, which might be candidate genes for drought stress response and tolerance in rice. The specific objectives of my study are described below.

Specific objectives

1. **Identification of genes associated with drought tolerance through large scale EST generation from a drought tolerant indica cultivar Nagina 22.**
2. ***In silico* physical mapping of unigene sets / candidate genes.**
3. **Conducting a field drought experiment for investigating gene expression**
4. **Gene expression profiling under field drought stress through construction of MIAME compliant cDNA microarray.**
5. **Discovery of candidate genes for drought tolerance using data obtained from EST generation, Physical Maps, comparative analysis, cDNA Microarrays.**

2. LITERATURE REVIEW

2. Review of literature

2.1 Rice: Importance and the present status

Rice is irrefutably the only plant species that can feed almost half of the world's population and accounts for as much as 35-75 % of the calories in some Asian countries. About 90% of rice crop is grown in Asia alone (about 154 million hectares) which is about 10% of the world's cultivable land (Khush, 2005). With the implementation of technological advances in rice crop production, productivity has increased from 257 million tons in 1966 to 600 million tons in 2000, an increase of about 130%. Though rice production has increased more than the population growth rate, we need to produce 40% more rice by 2030, with more and more people depending on rice as a staple food. Further more, the last decade has witnessed a noticeable decrease in rice production due to limitation of natural resources and environmental factors (Brown, 1996; 1997).

Rice is cultivated across different agro-climatic areas, which range from regions of high rainfall upto 5,000 mm annually to the margins of the deserts where the wet spells account to less than 100mm (Catlin, 1992). Rice growing areas span across tropics, sub tropics, semiarid tropics where indica cultivars are the most cultivated sub-species. Rice is also grown in temperate regions of the world where the most cultivated subspecies are japonica lines (IRRI, 1997). In fact rice is grown in almost every part of the world, ranging from 300m above sea level as in Bhutan and Nepal to as 3m below mean sea level in Kerala, India. Rice is grown in all climatic regions except Antarctica (Chang, 1976; Grist, 1986; Khush, 1997, 2005). Rice germplasm is rich with 150,000 different accessions identified so far. Of these, about 100,000 have been collected and stored to preserve the germplasm in various centers. The genus, *oryza* includes two domesticated wild species, *O.sativa* and *O.glaberima*, the later is mostly grown in West Africa while *O.sativa* is the most cultivated species across the world (Kush, 1997). Diverse climates, seasons, soils and varied cultural practices have led to tremendous ecological diversity among rice cultivars. These are classified into three ecological groups (Sauter, 2000); the upland rice, which is grown in non-irrigated fields contributing to only 4 % of the total rice production; the low land rice which is grown in

rain-fed, irrigated fields representing the most cultivated group producing more than 500 million tones of rice annually; the floating or deepwater rice varieties that account to about 10% of the total rice production and serve as a major of food in regions where recurring floods occur.

2.2 Plant response to environmental stress

Plants are exposed to various environmental stresses during their life cycle. These are grouped into seven major classes (Elstener and Oswald, 1994), light (high and low), radiation (UV-B, UV-A etc.), temperature (high, low, chilling and freezing), hydration (drought and flooding), chemical factors (salt, heavy metals, pH etc.), and mechanical factors. Of these, abiotic stresses such as cold, salinity, and drought cause heavy yield losses across the world (Boyer, 1992; Epstein *et al.,* 1980). In order to escape or avoid stress, plants developed a myriad of strategies and mechanisms to adapt to a variety of stresses (Bohnert *et al.,* 1995; Thomashow, 1999; Zhu *et al.,* 1997). These include stress avoidance, escape and tolerance responses. Experimental evidence established that plants adapt to the surrounding environmental conditions on a daily, or even on an hourly basis (Etherington, 1998). The early events by which plants respond and adapt to environmental stresses include sensing of stress and subsequent signal transduction events that activate various physiological and metabolic responses. Such responses are mediated via numerous physiological alterations involving numerous genes and regulatory circuits resulting in changes such as, alteration of plasma membrane composition, changes in phytohormone levels and changes in relative water content in the leaves (Bohnert and Sheveleva., 1998; Lichtenthaler, 1998; Matters and Scandalios, 1986; Vierling and Kimpel., 1992) all of which eventually lead to phenotypic changes. These adaptive changes are indeed correlated with a range of adaptive strategies in plants that are by and large tolerant to such stresses (McKersie and Leshem, 1994).

Plants, when exposed to repeated and severe biotic and abiotic factors are reported to respond by both preexisting and induced defenses (Alvarez et al., 1998; Feye et al., 2001). Though the precise cellular and molecular mechanisms involved are not yet deciphered, recent physiological and biochemical studies have described

memory effects of stress in plants. Such stress recognition in plant systems is exemplified by the reported development of immunological memory of induced nicotine accumulation in tobacco (Baldwin and Schmelz, 1996). Others include cold memory in arabidopsis and acclimation in response to pretreatment with cold stress (Knight et al., 1996, 1998). Similarly, memory effects in conifers in response to ozone damage (Langebartels et al., 1998). Recently, it was shown that chemical signaling through repeated exposure of arabidopsis with phytohormone ABA impairs light induced stomatal opening or inhibit the response to a light stimulus after entrainment of arabidopsis with ABA under dark light cycles. Also a transient expression of rd22 gene (Chang-Hyo Goh et al., 2003,) was observed in those plants.

Among the different classes of abiotic stress factors mentioned above, drought, salinity and freezing are the major stress factors that affect plant growth and productivity by disturbing cellular water balance. Since our studies are mainly focused on drought, this literature review covers largely drought tolerance studies on rice and other crop plants. Excellent reviews are available on abiotic stress response of a number of plant species (Shinozaki and Yamaguchi-Shinozaki, 2000, 2005)

2.3 Plant response to drought

Depending on the severity of water stress, plants respond through a cascade of molecular, biochemical and physiological processes which include stress perception, signal transduction, and regulation of gene expression (Bray 1997). Major research efforts are currently focused on identification of gene products that confer tolerance to water deficit stress and understanding the mechanisms of adaptation under water deficit stress (Bray *et al.*, 2004). Drought tolerance in plants is a complex trait, controlled by numerous dispersed genes across the genome associated with a multitude of interconnected pathways and processes. Past efforts to improve plant tolerance to drought stress had a limited success owing largely to the genetic complexity of the trait. Rapid advances in functional genomics in analysis of the transcriptome are beginning to provide leads in understanding the global gene expression under stress. The identification of water stress responsive genes is possible through a range of molecular and genetic resources (Leung *et al.*, 2001). The key molecular tools required include cDNA sequencing strategies through which ESTs are generated through stress specific

libraries, genome wide physical maps saturated with candidate genes identified through large scale expression profiling experiments such as microarrays and SAGE (seki et al, 2001, 2002, 2003, 2004; Kawasaki *et al.,* 2003; Rabani *et al.,* 2003), proteomics (Bohnert *et al.,* 2003), saturated genetic maps and conversion of candidate SRGs into molecular markers for MAS to utilize in molecular breeding. The array of genetic resources available for trait improvement includes germplasm collections, near-isogenic lines, mapping and mutant lines, which constitute a remarkable genetic variations. Among these genetic resources, mutant stocks with discrete genetic lesions are essential to determining gene function and dissecting biochemical and metabolic pathways. Many national and international collaborative projects are active in the production of rice mutants (Hirochika *et al.,* 2004; Leung and An, 2004). The most popular approach is insertional mutagenesis using T-DNA and Ac/Ds insertions (Krysan *et al.,* 1999; Jeon *et al.,* 2000; Jeong *et al.,* 2002; Upadhyaya *et al.,* 2002; Greco *et al.,* 2003; Wu *et al.,* 2003) and transposons (Altmann *et al.,* 1995; Hirochika, 2001; Miyao *et al.,* 2003).

Drought stress in plants is initiated primarily through dehydration of cell components and subsequently perturbation of electron transport leading to the production of toxic active oxygen radicals. The major components of drought tolerance are desiccation tolerance, osmotic adjustment, and antioxidant capacity (Zhang *et al.,* 1996). Osmotic adjustment, which helps to maintain osmotic potential and turgor within the cell, has been associated with crop yield under drought stress, has reported first in wheat (Blum and Pnuel, 1990; Morgan, 1995) and later in sorghum (Ludlow, 1993). The gene products associated with osmotic adjustment have a functional role in protecting cell membrane integrity, preventing inactivation of enzymes and alleviate protein denaturation and aggregation (Bohnert and Jensen, 1996).

Researchers have been focusing for long on the cellular signaling mechanisms that are activated by water deficit stress (reviewed by Shinozaki *et al.,* 2003; Xiong *et al.,* 2002). Several model plants, with the greatest emphasis on arabidopsis for dicots and rice for cereals, are being exploited to understand the molecular mechanisms that underlie plant responses to abiotic stress. The outcome of the findings will eventually translate to applications in the improvement of crop growth and production (Zhang *et al.,* 2004). Expression of several major classes of genes have been reported, which show

altered responses to water-deficit stress; genes involved in signaling and gene regulation, and gene products that are proposed to support cellular adaptation to water-deficit stress are among the most frequently altered in gene expression. Yet, the functions of a majority of genes with altered expression remain unknown and there are presumably more genes yet to be discovered. The functional characterization and regulation of all the genes in model plant genomes, including rice, is expected to be completed by about 2010 (Konez, 2003)

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 the function of stress responses. These include the genes that regulate osmotic adjustment inside a cell which is otherwise disturbed due to water loss. Primarily such genes control the synthesis and accumulation of osmolytes without disturbing cellular functions (Ishitani *et al.,* 1997). These encode enzymes required for the biosynthesis of various osmoprotectant molecules such as sugars, proline, glycine-betaine etc. Also included in this class are proteins that protect macromolecules and membranes such as LEA proteins, osmotin, antifreeze protein, chaperon, mRNA binding proteins etc. Late embryonic abundant proteins such as dehydrins are abundantly expressed under stress and these are known to protect cellular machinery (Lisse *et al.,* 1996). Other important stress induced gene products include aquaporins, which are trans-membrane water channel proteins involved in the regulation of water movement across membranes and thus play an important role in osmoregulation and avoidance of water deficit (Fray *et al.,* 1994; Ruiter *et al.,* 1997). Stress induced enzymes include detoxifying enzymes (Ingram and Bartels., 1996) Such as GSTs, catalases, super oxide dismutase, ascorbate peroxidase etc. chaperonins and protienases that may destroy inactive proteins (Williams *et al.,* 1994), and enzymes involved in ATP production pathways (Riccardi et al 1998).

The second group largely contains protein factors involved in signal transduction and regulation of stress responsive gene expression. Examples include protein kinases, transcriptional factors, and phospholipases. The transcripts for genes encoding several of these proteins are shown to accumulate under drought conditions.

The roles of products of these genes are extensively reviewed (Shinozaki and Yamaguchi-Shinozaki, 2000, 2005).

2.4 Impact of drought on rice

Drought has become the most significant constraint to realizing the yield potential of rice across almost all agro-climatic zones. In some years, abiotic stresses caused crop losses as much as 50% (Boyer, 1982, Bray, et al. 2000) and drought alone may cause yield losses of as much as 15% (Dey *et al.*, 1996). Drought and salinity are the major factors limiting rice production in rainfed ecosystems, which lack infrastructure, crop and resource management practices, and robust rice breeding exercises (Bennett, 2002). Rice being a water loving plant consumes almost 5000lts of fresh water to produce one kilogram of rice. This places a heavy demand for water which is in short supply in rice growing semi arid and arid tropics. Till date, 28% of the land on earth is dry soils, which are not conducive for crop cultivation under such conditions (Bray, 2004). With an increasing population at a rate of 80 millions per annum (U.S. Bureau of Census, 1998), there will be a struggle for land, food, and water. It is projected that to meet the growing population demand we need to produce 40% more rice with less cultivable land, less water and low fertilizers (Khush, 2005).

Developing drought tolerant lines through breeding is rather slow due to the genetic complexity of the trait and variations in environment, and strong Genotype x environmental interactions (O'toole and Chang., 1997; Blum 1998; Fukai and Cooper 1995). Further, this apparent cause of slow progress has been due to lack of sufficient genetic information about genes that govern this complex trait, drought tolerance and its components, called secondary traits. Gene discovery has been greatly enhanced in the recent years due to rapid progress in structural and functional genomics (Pandey and Mann, 2000). The availability of complete genome sequences of two indica sub-species and the availability of gene expression profiles through ESTs generated through stress-specific libraries, have helped in identification of genes involved in drought tolerance and has led breeder's access to a large number of genes (Barry., 2001). This has become a key factor helping to accelerate breeding for improved tolerance (Cushman and Bohnert, 2000).

2.5 Rice functional genomics

Functional genomics tools have emerged in the post genomic era as a major means to elucidate gene functions. Functional genomics basically encompasses functional analysis of genes and their temporal and spatial regulation. The ESTs serve as a basic genomic resource for functional analysis of genes (Richmond and Somerville 2000), and are useful in analysis of the exact sequence, location and function of genes in a genome of a particular plant. This is the first step towards understanding how biological pathways work in a cell. Functional genomics plays a vital role in transforming huge nucleotide data into meaningful functional data. It's a general approach to understand how genes work together by assigning both known and new functions to unknown genes. Identification of a gene to phenotype is the first step in elucidation of physiological and biochemical function of the gene product or protein.

Rice functional genomics has entered a golden era with the availability of the whole genome sequences of both the subspecies. Determining the function of genes in rice is now under tremendous focus. Rice genome was estimated to have around 40,000-60,000 genes (Goff et al 2002, Yu et al 2002, Yu et al 2005). The IRGSP estimates are about 37,500 genes (non-TE related) (IRGSP, 2005). One reason for the variation in gene estimation is lack of supporting evidences from deep EST coverage of gene space. Further, 40% of the annotated genes have no apparent functions. The challenge therefore is to determine functions of these genes and their relationships.

The functional genomics approaches allow one to simultaneously study large groups of related genes affecting the phenotypes and decipher gene networks and biochemical pathways underlying complex phenotypes at molecular level (Zhi-Kang Li 2005). The ultimate aim of genome analysis is to predict a phenotype from a genotype. The question is weather is it necessary that transcript profiling is still required, when the whole genome sequence is available now, However, predicting genes is still an imperfect science (Hogenesch, 2001., Mathe *et al.*, 2002; Rogic *et al.*, 2002: Bennetzen 2004), and all the existing algorithms proved to be inadequate in accurate gene prediction. The rice genome annotation largely relied on existing gene information either in the form of ESTs or cDNA resources. Genetics plays a major role in predicting a phenotype since there are significant G X E effects for each and every trait. Genome

wide effects of epistasis will greatly limit such predictions. Even a paralouge of a same gene may end up with different function altogether, which is contributed by even a small factor eg missing or new protein binding domains, mutations, alternative splicing, post transcriptional/ translational changes, glycosylation all of which contribute to a very large diversity among phenotypes.

Functional genomic research can be carried out on a large scale through the application of high throughput genomic technologies and bioinformatics. Such genomic technologies include large scale Expressed Sequence Tags generation, gene expression profiling by various types of microarray and gene chips for differential gene expression, genotyping and proteomic studies (Yazaki *et al.*, 2002, 2004a, b; Sun *et al.*, 2003). The bioinformatics tools for rice functional genomic analysis are being regularly updated (for e.g http://www.grammene.org).

Thus, rice functional genomics studies are expected to provide deeper insights into the biochemical pathways and a wealth of information on cereal-specific proteins. In our efforts towards this, we have been continuously generating Expressed Sequence Tags from a drought stressed normalized cDNA library in our gene discovery program since the year 2000 and are making these invaluable resources available for public accesses through depositing in public domains (www.ncbi.nlm.nih.gov/dbest) (Reddy *et al.*, 2001; Markandeya *et al.*, 2002, 2003).

2.6 Rice genome sequencing and its aftermath

Due to the compact size of the rice genome, it's more amenable to genetic analysis than any other crop plants. Also, its close relationship with other cereal species makes rice an ideal model system to study grass physiology, development, agronomics, and genomics of the grass family (Gale and Devos, 1998b; Goff, 1999; Shimamoto and Kyozuka, 2002; Paterson *et al.*, 2003). Draft genome sequences of the two major subspecies of rice, indica, and japonica, were published in 2002 (Goff *et al.*, 2002; Yu *et al.*, 2002). The continuous and anchored (finished phase) sequences for japonica chromosomes 1, 4 and 10 were later released and published (Feng *et al.*, 2002; Sasaki *et al.*, 2002; The Rice Chromosome 10 Sequencing Consortium, 2003). Finished whole genome sequences for japonica and indica rice have been recently published by the

International Rice Genome Sequencing Project (IRGSP, http://rgp.dna.affrc.go.jp/ IRGSP/Build3/build3.html) and the Beijing Genome Institute (Yu *et al.,* 2005), respectively. Extensive homology has been detected between the rice genome and sequences or expressed sequence tags from other cereals (Feng *et al.,* 2002; Goff *et al.,* 2002; Kikuchi *et al.,* 2003; Bedell *et al.,* 2005; Yu *et al.,* 2005), indicating the presence of a large number of cereal-specific genes. The rice sequence data gains its significance not only because it is an important crop plant, but it is also expected to play a major role in understanding the grass family through comparative analysis. Such a comparison at genome level was earlier hindered due to the lack of long-range contiguity in sequence information, except for the existing sequence information for arabidopsis. This, coupled with the available extensive genetic and physical maps in rice, provides a foundation for organizing information on diverse cereals which will lead to identification of orthologues genes. This will also facilitate the genome sequencing of other cereals thereby throwing light on cereal evolutionary history.

The rice genome structure has been extensively analyzed through international collaborative efforts. Availability of a near completed draft sequence of a japonica cultivar Nipponbare, physical maps of the genome by BAC/PAC clones and genome analysis revealed that the genome size is nearly 400Mb (Sasaki, 2005). The filtered IRGSP finished rice genome sequence (IRGSP, 2005) generated 370Mb representing 95% of the genome. A total of 37,544 non-transposable element protein coding sequences have been detected with a gene density of one gene per 9.9kb, about 2,859 of these genes seem to be unique to rice and other cereals, some of which may have differentiated in monocots and dicot lineages (IRGSP, 2005). Around 3,243 genes of rice contain 11,487 retrotransposon insertion sites, which accounts to 35%, and consists all transposon families. Organellar DNA fragments accounts to 0.38% to 0.43% in the nuclear genome indicating repeated and ongoing transfer of organellar DNA to nuclear genome.

The comparison of draft sequences with the only other fully sequenced plant genome; arabidopsis thaliana “a popular laboratory model” reveals that 90% of the arabidopsis proteins occur in rice but only 71% of the rice proteins occur in arabidopsis. The gene predictions project that around 2859 genes are specific to rice and cereals and

do not posses any homologue in arabidopsis. The monocots lineage appears to be fast evolving as evidenced by the SNP frequency between two subspecies of rice, indica and japonica, which varies from 0.53% to 0.78% and this is 20 times more than the SNP frequency found between two ecotypes of Arabidopsis, Columbia and Landsberg .

2.7 Rice genome annotation

Though rice genome sequence annotation gives approximate gene index, it does not throw light on functional role of all genes and the tissue specific and environment influenced regulation. The availability of more than 300 000 public EST resources (www.ncbi.nlm.nih.gov/dbest) generated from various libraries from different tissues subjected to abiotic and biotic stress provide information on gene expression abundance, tissue specific and developmental expression of genes. These along with a large set of full-length cDNA clones (The Rice Full-Length cDNA Consortium, 2003) will be an important resource for functional genomics and proteomics of rice, both of which facilitate functional annotation of the rice genome. (Yunbi Xu., 2005). Gene prediction of the rice genome was done by FGENESH which was proved to be the best gene prediction algorithm for rice (Yu J *et al.,* 2002), and the estimates ranged from 49, 088 in Beijing *indica* to 45, 824 Syngenta *japonica* to 43,635 (IRGSP *japonica*). Of these, the percentage of ESTs represented for the subspecies are 31.9%, 32.6%, and 34.7% respectively. Using the above EST adjustments, the number of predicted genes in Beijing *indica* that are not found in either *japonica* presumably is 1,064. Conversely the Syngenta Japonica has 1,517 predicted genes that are not in indica (IRGSP japonica is 1,479). As a fraction of the total 2.2% and 3.3% of indica and japonica genes, respectively, are unique to the subspecies.

Rice genome analysis identified GC-rich islands which were randomly distributed, and in the case of chromosome 1, the GC content of coding and non-coding regions are 58% and 41%, respectively. The overall GC content is 43% (Sasaki *et al.,* 2002). Large segmental duplications have been found in more than half of the genome. By genetic mapping using a limited number of DNA markers, only the highly mutually similar segments were found at the distal ends of the short arms of chromosomes 11 and 12 (Wu *et al.,* 1998). However, unlike the Arabidopsis genome (The Arabidopsis

Genome Initiative, 2000), the internal duplication of chromosomes is not so significant in the rice genome. Efforts to unravel whole genome duplication has complicated due to rapid loss of duplicated genes and also because of high rate of duplications of individual genes. Though polyploidy is common in plants but it's yet proved as and when how many duplications have occurred in arabidopsis (Blanc., 2000; The Arabidopsis Genome Initiative(2000); Vision., 2000; Simillion., 2000; Bowers., 2003;Dominguez., 2003), for rice segmental duplication was known since long (Kishimoto *et al.*, 1994;Nagamura *et al.*,1995; Wang *et al.*, 2000)

Summary of rice genome annotation (IRGSP 2005)

S. No	Description of annotation	Number
1.	Non-Transposon FGENESH gene predictions	37, 544
2.	Genes supported by the existing full length cDNA sequences	17, 016
3.	Genes matching ESTs or cDNAs	28, 840 (61%)
4.	ESTs for each predicted matched gene	10.7
5.	Genes aligned to other cereal ESTs	2,927
6.	Genes matched other than non-rice ESTs	330
7.	Rice genes with no homologue in arabidopsis	2,859
8.	Rice genes supported by expression evidence homologue with arabidopsis	20, 311 (88%)
9.	Predicted rice genes with homologue with arabidopsis	26, 839 (71%)
10.	Arabidopsis genes with homologue	26, 004 (89.8%)
11.	Proteins which had matches with entries in Swiss-prot database	19,675
12.	Proteins with no expressional evidence	4, 500
13.	Domains present in the predicted proteome	3,328
14.	Percentage range of SNP frequency	0.53% - 0.73%
15.	Polymorphic sites that differentiate two rice subspecies *indica* and *japonica*	80, 127
16.	SSRs	18, 828

2.8 Genomics

2.8.1 Historical biological perspective

Spectacular technological advances in DNA sequencing have revolutionized the way genetic analysis is carried out today. Development of high throughput and powerful technologies has enabled sequencing of large DNA fragments and even complete genomes. The development of powerful technologies for DNA sequencing and sequence analysis has been possible with automated procedures for sample preparation and new software tools for sequence analysis.

The introduction of sequencing methodologies such as chemical method (Maxam and Gilbert, 1997) and the enzymatic chain termination method of Sanger (Sanger et al, 1982). Since then, the sequencing techniques, and especially the enzymatic chain termination method of Sanger, have been further developed and adapted to different kinds of automation the developments which include laboratory innovations such as four-colour fluorescence-based sequence detection (Smith *et al.*, 1986). Improvement in developing fluorescent dyes (Ju et al ., 1995; Rosenblum et al 1997), dye-labelled terminators (Prober, 1987), and advancement in developing Taq polymerases specifically designed for sequencing Reeve, M.A., 1995; Tabor, S., 1990), cycle sequencing (Murray et al ., 1989) and capillary gel electrophoresis (Guttman *et al.*,1990; Meldrum *et al.*, 2000) were the major landmarks in the history of sequencing innovations. These innovations brought substantial improvement in the automation, quality and throughput of DNA sequencing. (Meldrum *et al.*, 2000).

The other advances include the development of software packages for the analysis of sequence data. The PHRED software package77, 78 introduced the concept of assigning a `base-quality score' to each base, on the basis of the probability of an erroneous call. These quality scores make it possible to monitor raw data quality and also assist in determining whether two similar sequences truly overlap. The assembly of the base called FASTA sequence posed a major problem until the availability of PHRAP computer package http://bozeman.mbt.washington.edu/phrap.docs/phrap.html), CAP3 (Madan and Xhuang., 1999), TIGR assembler, all of which systematically assemble the sequence data using the base-quality scores. Subsequently, automated methods, for example, preparation and creating new biochemical protocols suitable for

automation, lead to a dramatic increase in sequence throughput which eventually led to the complete sequencing of large genomes.

2.8.2 Towards uncovering the gene space of a genome

The past decade has witnessed extensive eukaryotic genome sequencing beginning with the human genome. The availability of the complete genomes enables us to have a much better overview of the organization of life and also provides extensive starting material for gene prediction. Identification of genes from the huge genome sequence repositories often require sophisticated bioinformatics software tools. Once identified the task thereafter is validation of the existence of the predicted gene. The coding gene space in a higher eukaryotic genome hardly accounts 5%. In the case of plants, it's less than 1% in some genus. This is due to large number of retrotransposons, large genome duplications, and other repetitive DNA elements. Thus the whole genome sequencing for cataloguing all the genes(less than 5%), is enormously expensive. In view of this, alternate cost effective strategies to decode the gene space have become a necessity. Several methods by which the coding genes can be isolated have been described (Horton *et al.*, 1989; Adams *et al.*, 1991; Venter., 1993).This review places emphasis on detailed description of Expressed Sequence Tags which complement the whole genome sequence information. These technologies decode gene space as cDNAs (ESTs) represent gene tags of an individual organism.

2.8.3 Expressed Sequence Tags a major genomic resources

The first application of high-throughput sequencing of cDNA clones was described in 1991(Adams *et al.*, 1991), where clones from human brain were randomly selected and partially sequenced. These partial sequences represented genes expressed in the tissue at a certain time point and were termed ESTs (Expressed Sequence Tags) (Adams *et al.*, 1991; Venter., 1993). Similar approaches were made for different tissues and organisms. A major advantage of EST sequencing methodology over genomic sequencing from the gene discovery perspective is that it is considerably cheaper and quick.

Another advantage of EST sequencing would be that although cDNA clones are sequenced randomly, it is possible to compare the data and assemble contigs after sequencing. Thus, despite the limitation that only 400-500 bp are sequenced for each clone, alignment of these EST sequences can yield the entire sequence for full length cDNA representing the entire coding sequence of the corresponding gene. In addition to being a rapid and economical method for gene discovery, EST sequencing has other uses as well. Firstly, it provides a rapid means of cataloguing the genes transcribed in a tissue or cell type from which the cDNA library is constructed (Sterky *et al.,* 1998; Allona *et al.,* 1998; Ewing *et al.,* 1999). Secondly, sequences obtained by random selection of cDNA clones will yield a statistical picture of the level and complexity of gene expression for the sample tissue, and the expression levels of the corresponding gene ("*in-silico* or digital northern") (Audic and Claverie, 1997). Thirdly, comparing the EST distribution of two or more cDNA libraries can provide useful information on differences between the transcriptomes of the specific tissues or developmental stages (Ewing *et al.,* 1999; Audic and Claverie, 1997; Bhalerao *et al.,* 2003; Markandeya *et al.,* 2005).

The ESTs represent the largest amount of information possible per sequenced base of eukaryotic genomes. EST sequences are not only valuable in identification and cataloguing of all and new genes, but they also represent a snapshot of gene expression under a defined set of environmental factors and tissue specific gene expression. (Rudd, 2003; Okubo *et al.,* 1992). The gene sequences obtained can be efficiently used for physical mapping by determining their chromosomal position. Moreover, they also contribute to the understanding of intron and exon boundaries, which will help to predict the gene structures. These facts confirm the need for cDNA sequencing as a complement to genome sequencing. Furthermore, in addition to being a source of a catalogue of genes, EST sequences can be useful in SNP discoveries (Picoult-Newberg *et al.,* 1999; Deutsch *et al.,* 2001; Kota *et al.,* 2003). ESTs are derived from a number of different genotypes, and significant polymorphism is expected to exist between multiple EST representations of a gene. Thus, it should be possible to acquire a large number of SNP and indel polymorphisms by bioinformatics analysis of EST data. A major difficulty with this approach is that EST sequences contain sequencing errors, which

have to be distinguished from true polymorphisms. Several software tools are available for distinguishing true polymorphisms (Beutow 1999). Positional cloning is greatly facilitated by the combination of EST sequencing and dense physical maps (Collins, 1995) and EST sequencing from different organisms will enable evolutionary studies as well as cloning of interesting genes by hopping across taxonomic boundaries (Marra *et al.,* 1998).

Apart from the above, ESTs are important component of functional genomics, and aid in genome analysis, especially annotation of genomic DNA by providing expressional evidence from hypothetical to unknown. These help as in-*silico* tools to identify a suite of promising candidate genes and unravel QTLs/gene networks that control important traits. Till today, mapping coding regions on a genome scale in rice have focused on EST and full-length cDNA analysis (Kikuchi *et al.,* 2003). However, the available EST and cDNA resources do not comprehensively reveal all genomic coding information as they are biased towards the highly expressed genes. Not surprisingly, exhaustive efforts to uncover the transcriptome have only represented only a fraction of the predicted total genes (Feng *et al.,* 2002; Sasaki *et al.,* 2002; Reddy *et al.,* 2002; Yu *et al.,* 2003; Zhao *et al.,* 2004; Markandeya *et al.,* 2005). Hence it's obvious that more ESTs needs to be generated from different tissues and across all available environmental conditions since ESTs provide the most direct approach for discovering genes associated with stress response and validating all predictions by gene finding program's. This has been demonstrated in several plant systems (Michale *et al.,* 2002; Fernandes *et al.,* 2002; Echenique *et al.,* 2002; Reddy *et al.,* 2002; Markandeya *et al.,* 2005).

The large scale EST sequencing projects in rice have generated huge data which served as a valuable tool to catalogue substantial number of genes (Uchimiya *et al.,* 1992; Umeda et al 1994, Yamamoto and Sasaki, 1997., Reddy et al 2002., Markandeya *et al.,* 2005, Zhang et al 2005) and in deciphering the role of transcriptionally regulated genes in different tissues (Ewing *et al.,* 1999). However, only a few studies focused on the analysis of transcriptome profiles of rice seedlings subjected to abiotic stress (Umeda *et al.,* 1994; Matsumura *et al.,* 1999; Kawasaki *et al.,* 2001). ESTs generated from a vast number of sequencing projects aided by homology searches have been

essential to the categorization of known expressed genes and discovery of new genes (Staden 1994, Gaasterland and Sensen, 1996, Schultz et al, 2000). However, the inherent inconsistency of EST data, coupled with inconsistent prediction results from different gene-finding programs, make gene prediction a difficult task as exemplified by the unusually large numbers of "unique" no homology EST sequences that are reported in the EST sequencing projects (Sterky *et al.,* 1998; Allona *et al.,* 1998). Since excellent resources are already available on this subject, we will focus only on bioinformatics tools which aid in EST analysis.

Currently, two widely–used computational methods exist to predict reading frames for DNA sequences (Fickett, 1996). These include the one which uses the reading frame detected from strong similarities found from a database similarity search program such as BLASTX (Altschul, *et al.,* 1990, 1997), BlastN (Atlschul and Gish, 1991) and the other which uses a coding region identification program (also called gene–finding or gene identification programs), such as GRAIL (Uberbacher and Mural, 1991), FGENESH, Genscan, GeneMark HMM, Glimmer R, Rice Genome Automated Annotation System (Rice GAAS) and the more recent Genie (Reese, 2000).

The ESTs generated are assembled using an array of assembly programs such as PHRAP computer package http://bozeman.mbt.washington.edu/phrap.docs/phrap.html), CAP3 (Madan and Xhuang., 1999), TIGR assembler. Assembly of the ESTs is performed to derive a unigene set, since a unigene has more sequence length than individual ESTs, these can be utilized to search for protein motifs using InterProScan v3.3 (http://www.ebi.ac.uk/InterProScan/). Physical mapping is now possible with the available of cMAP (Fang *et al.,* 2003) to anchor unigenes onto chromosomes thus facilitating gene structure and introns exon borders.

Many of the sequencing projects proved to be expensive due to high redundancy (Reddy *et al.,* 2002) and particularly transcript profiling under drought were not carried out in rice till we initially began our study to identify drought transcriptome through large scale EST generation. In our efforts towards these and to overcome all the limitations of ESTs described above, we have chosen a normalized cDNA library which was earlier constructed from drought stressed seedlings of an indica cultivar Nagina 22 (Reddy *et al.,* 2002). We were successful in overcoming all the limitations of ESTs, by

choosing a normalized cDNA library to reduce redundancy and improving high-throughput sequencing strategies to obtain high quality sequences and generated high density SRG physical maps (Markandeya *et al.,* 2005). These resources aided in utilizing the ESTs as RFLP markers (Shivram Prasad L., PhD thesis), converting them into PCR based markers (Chandrasekhar A., PhD thesis) and were used to identify SNPs (Vijayabhasker Reddy *et al.,* unpublished data from our lab).

2.8.4 Gene expression profiling

Traditional methods for gene expressional analysis were done by Northern blot (Alwine *et al.,* 1979) and in situ hybridization (Gall and Pardue, 1969). Northern blots are the most routinely used method for detection of transcripts. Another method is in situ hybridization which is valuable and well established method for localization of gene expression at tissue level (Dagerlind *et al.,* 1992). Cellular maps of gene expression within cell types and tissues can be provided by in situ hybridization (Bankfalvi and Schmid, 1994). The major advantage being the sensitivity at which it detects. Detection of down to tens of mRNA molecules (Harris *et al.,* 1996) and it's ability to detect activity of a gene independently of the final protein product. The technological advances in global gene expression profiling are described below.

2.8.5 Serial analysis of gene expression (SAGE)

Serial Analysis of Gene Expression (SAGE) was first described by Velculescu in the year (1995). Since then it has been widely used in various studies. The method has been applied in a number of cancer studies (Zhang *et al.,* 1997; Hibi *et al.,* 1998), immunological studies (Chen *et al.,* 1998; Hashimoto *et al.,* 1999) and for rice and yeast transcription profiling (Matsumura *et al.,* 1999; Velculescu *et al.,* 1997). The obvious advantage of the method is that the expression profiles generated will benefit from the high number of tags included, so that rare transcript are more likely to be detected and tag counts more representative of true expression levels. Having created expression profiles for one cell type, a statistical comparison with SAGE profiles from other cells can be performed for detection of similarities and differences in gene expression.

However, the method has also been associated with several drawbacks and limitations, some of which have been solved over the years (Yamamoto *et al.*, 2001). SAGE is a very powerful technique for transcript profiling and has provided a wealth of valuable expression data in a number of different studies (for a comprehensive review see (Yamamoto *et al.*, 2001). Tens of millions of SAGE tags from various organisms are currently deposited in the public SAGEmap database (http://www.ncbi.nlm.nih.gov/SAGE/) and the gene expression omnibus database (GEO) (http://www.ncbi.nlm.nih.gov/projects/geo/).

2.8.6 Massively parallel signature sequencing (MPSS)

Another technological innovation similar to the SAGE principle is MPSS (Massive Parallel Signature Sequencing). It was first described by Brenner (Brenner and Johnson *et al.*, 2000) (Reinartz *et al.*, 2002). As in SAGE, MPSS too involves counting 16–20 bp signature tags to establish a transcription profile for a cell or tissue type. The tags are sequenced using microbead arrays instead of cloning them into concatemers and sequencing them by standard sequencers.

2.8.7 Microarray technologies

2.8.7.1 Microarrays: An overview

The myriads of genome sequencing data generated in the past decade have opened new avenues for biologists to explore individual organisms at the genome level. Microarray technology has revolutionized the analysis of genome scale gene expression (Eisen and Brown, 1999; Schena *et al.*, 1995) and enables comprehensive and high through-put surveys of DNA or RNA molecules on a genome wide scale (Somerville *et al.*, 1999; Richmond *et al.*, 2000; Schaffer, 2000). DNA microarray technology is one of the most powerful techniques recently developed to bridge the gap between sequence information and functional genomics. The microarray field is a good example of the assembly and convergence of several technologies, including automated DNA sequencing, DNA amplification by PCR, highly efficient oligonucleotide synthesis,

nucleic acid labeling chemistries, and bioinformatics (Barrett, 2003). Though the initial origin of this approach is not clear (Ekins and Chu 1999; Weeraratna *et al.,* 2004), this technology was developed somewhere in mid 80's or early 90's (Ekins *et al.,* 1989; Ekins., 1980; Augenlicht *et al.,* 1987; Lennon and Lehrach,, 1991; Southern *et al.,* 1994; Zhao *et al.,* 1995). The first use of microarrays was published in 1995 by Schena using 48 Arabidopsis genes for differential gene expression.

Microarray technology allows the determination of transcript abundance for many or all transcripts in a genome by comparing control and experimental states. The RNAs from different treatments are distinguished by the incorporation of different fluorescent labels (Schuchardt *et al.,* 2000; Deyholos and Galbraith, 2001). By looking at the patterns of characterized genes, knowledge of their function can be used to indicate functions of uncharacterized genes with the same pattern of regulation (Iyer *et al.,* 1999). Microarray technologies were used to study expression profiles of inflammatory disease-related genes and were analyzed under various induction conditions by this chip-based method (Heller *et al.,* 1997). Furthermore, the yeast genome of 6000 coding sequences has been analyzed for dynamic expression by the use of microarrays (DeRisi *et al.,* 1997; Wodicka *et al.,* 1997). However, in plant, relatively few reports of microarray analyses have been published (Schena *et al.,* 1995; Ruan *et al.,* 1998; Aharoni *et al.,* 2000; Reymond *et al.,* 2000; Kawasaki *et al.,* 2001). 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 (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).

The major advantages of microarrays are derived from the fact that data can be collected for large numbers of genes in one experiment permitting true genome-scale sampling of gene expression patterns. It is important to note that microarrays, like northern blots, provide a measure of steady-state RNA levels. One important information derived from expression profile analyses is the elucidation of different

temporal patterns of gene regulation. A limitation with expression profile analysis by hybridization of microarrays is the prerequisite to know the genes, whose regulation one is attempting to study. This limitation is of course diminishing as the availability of sequence information is increasing. With the completion of the sequencing of several genomes, the analysis of global gene expression patterns is likely to play an important role in understanding biology.

Microarrays have been used in many ways for transcriptome studies, for example, special arrays have been designed to study alternative splicing (Hu *et al.,* 2001; Clark *et al.,* 2002; Johnson *et al.,* 2003; Yeakley *et al.,* 2002). Microarrays have also been used to monitor transcript degradation (Bernstein *et al.,* 2002) (Fan *et al.,* 2002) and evaluate the expression of noncoding RNAs (Cawley *et al.,* 2004). For genome studies exon and "tiling" arrays have been used to experimentally validate and refine computational gene predictions (Shoemaker *et al.,* 2001; Kapranov *et al.,* 2002). Microarrays have also been used for SNP genotyping (Patil *et al.,* 2001) and to detect changes in DNA sequence copy number (Kashiwagi and Uchida, 2000), for mapping the origins of replication in Saccharomyces cerevisiae (Raghuraman *et al.,* 2001) and Sulfolobus (Lundgren *et al.,* 2004), identifying the binding sites for transcription factors on a genome-wide level (Iyer *et al.,* 2001; Bulyk *et al.,* 2001), and characterizing epigenetic modifications of the chromatin (Van Steensel and Henikoff, 2003).

Several methods have been described for producing microarrays, and two basic types which are most commonly used are spotted arrays where pre-synthesized DNAs are printed onto glass slides, and high-density oligonucleotide arrays on which sets of oligomers are synthesized *in situ* on glass wafers (Harrington *et al.,* 2000). Spotted arrays can be produced in-house whereas high-density arrays require advanced instrumentation and are commercially available. We will describe briefly on the most used array platforms.

2.8.7.2 Oligonucleotide microarrays / GeneChip™ technology

High density GeneChips were designed for the first time using gene information, EST sequences and genome sequence information from public databases by Affymetrix.

Their GeneChip technology is based on chemical synthesis of oligonucleotides directly on a solid surface, using a photolithographic method similar to techniques used in the production of computer chips (Fodor *et al.,* 1991; Pease *et al.,* 1994; Lockhart *et al.,* 1996; Lipshutz *et al.,* 1999). There are many advantages and also some disadvantages in GeneChip oligonucleotide arrays. The major disadvantages being short probes and only the existing genes can be represented on the chip whereas novel and unidentified genes will be missing.

2.8.7.3 Long oligonucleotide arrays

To overcome the drawbacks in GeneChips, long oligonucleotides about 40-80 bases have been made and are used as probes. The first such arrays were synthesized by an ink-jet printing method (Hughes *et al.,* 2001). Recently, pre-fabricated oligonucleotides have been used and spotted similarly on to cDNA microarrays (Beaucage, 2001). The whole genome longmer oligonucleotide sets for printing are now available from commercial vendors. The long oligonucleotide arrays combine properties of both cDNA and GeneChip arrays and offer several advantages; the greater length of the probes enables higher specificity in the hybridization than the shorter GeneChip probes, and this makes them less sensitive to phenomena such as single nucleotide polymorphisms.

2.8.7.4 cDNA microarrays

The popularity of cDNA microarrays as cheaper and more reliable alternatives to the GeneChip and oligonucleotide arrays were developed by Patrick Brown and David Botstein at Stanford University. Since its inception in early 90's many variations in robotic platforms and standard protocols for printing, hybridization and data analysis were developed and made available freely by the Brown group. This is the first academic effort which made a revolution in the way gene expression was studied. (Schena *et al.,* 1995; Shalon *et al.,* 1996; DeRisi *et al.,* 1997). Today, a wide range of diverse protocols for array fabrication, sample labelling and hybridization are available (reviewed in detail Lockhart, 1999 Nature Genetics 1999; Cheung *et al.,* 1999; Hegde *et*

al., 2000; Lee *et al.,* 2000; Yue *et al.,* 2001; Zhang *et al.,* 2001; Schulze and Downward 2001; Shoemaker and Linsley 2002; Nature genetics 2002; Weeraratna et al. 2004). In the present study we have used cDNA microarrays which were fabricated using ESTs from drought stressed seedling library and carried out expression profiling across different stages of rice life cycle under field drought stress conditions. We will review in brief a few aspects of cDNA microarray experimental design, array fabrication issues, hybridization, and data analysis.

The design of an microarray relies on the arraying of cDNA clones in that the array may be focused on transcripts associated with a particular library or tissue as in the case of our approach having cDNA clones generated from drought specific library (Markandeya *et al.,* 2006 PAG-IX) or may be more global, representing all or most of the transcriptome (Forster *et al.,* 2003; Weeraratna *et al.,* 2004). All the clones should be printed in duplicate and in addition, a set of control clones should be included on the array. These should preferably include replicate spots printed at different positions on the array to ensure reproducibility and hybridization quality. Negative controls can include repetitive DNA, poly(A) DNA, intergenic DNA, empty spots and non-homologous sequences from other organisms, to measure possible cross-reactivity and non-specific fluorescence. Also, positive controls can be included, e.g. for measuring spiked RNA for normalization purposes. Microarray experiments are very sensitive to a number of parameters such as tissue sampling, RNA preparation, labelling, and hybridization which introduce variations in the data that could distort the measured levels of gene expression. In order to compensate for this, different levels of replication are necessary (Lee *et al.,* 2000; Pan *et al.,* 2002; Churchill, 2002; Dobbin *et al.,* 2003). Biological replicates will account for genetically and environmentally induced variations between individuals and samples. Technical replicates (repeated hybridizations with the same biological sample) average out variation introduced during RNA extraction, labelling, and hybridization. Duplicate spots of the same probe are also often included on the array to account for measurement errors caused by artifacts during printing or hybridization.

There are basically 3 main types of experimental designs which include direct comparison, reference design and loop design for the above are largely used and have been described elaborately (Churchill 2002; Yang and Speed 2002; Dobbin *et al.,* 2003; Kerr and Churchill 2001). Among these the most opted and commonly preferred type of design is the reference design, which offers great flexibility in that multiple samples can be compared and samples can be added or withdrawn from the study without disturbing the rest of the samples. The reference could be a biologically relevant sample or a completely different sample, preferably containing as many different transcripts as possible (Dudley *et al.,* 2002; Sterrenburg *et al.,* 2002). Loop designs are also an option if multiple samples are to be compared (Kerr and Churchill 2001). Many other types of design have also been proposed (Yang and Speed 2002).

Most of the standardized methodologies of DNA binding onto glass surfaces are reported for aminosilane glass chemistries (Wang *et al.,* 2003). Successful printing requires controlled environmental conditions, such as stable air humidity and temperature and careful elimination of disturbing particles from spotting pens and slides (Hegde *et al.,* 2000; Diehl *et al.,* 2001). The results of contact printing depend largely on the quality of the printing pins and spotting solutions, the effect of different spotting solutions in spot morphology has been reported (Hegde *et al.,* 2000).

The most commonly employed method for target Labelling is indirect protocol which uses aminoallyl labeled nucleotides for cDNA synthesis followed by coupling to Cy-dye esters (Randolph and Waggoner, 1997; Schroeder *et al.,* 2002; Hughes *et al.,* 2001). Many alternatives are used to improve labelling efficiency and increase signal strength include the use of alternative reverse transcriptases (SuperScript™ III and FluoroScript™ from Invitrogen and CyScribe™ from Amersham Biosciences), alternative fluorophores (Wildsmith *et al.,* 2001., www.probes.com) and signal amplification systems such as the dendrimer-based 3DNA™ Submicro system from Genisphere®, (Stears *et al.,* 2000) and the tyramide signal amplification (TSA) system from PerkinElmer® (Karsten *et al.,* 2002). Various methods of initial blocking of reactive groups with non-fluorescent biomolecules such as bovine serum albumin (BSA) (Hegde *et al.,* 2000) and chemical blocking by succnic anhydride (Diehl *et al.,*

2001) have been proposed. The hybridization buffers commonly include denaturing agents such as formamide and sodium dodecyl sulphate (SDS), as well as salts and agents such as COT1-DNA and poly(A)-DNA that block repetitive sequences.

2.8.7.5 cDNA microarray data analysis

2.8.7.5.1 Approaches for microarray data analysis

Image acquisition and processing of the slides are done through scanning arrays with a confocal laser scanner. The scanner settings and thus the obtained signal intensities could potentially influence the quality of the downstream data analysis. Different strategies have been proposed to obtain optimal settings (Forster *et al.,* 2003; Yang *et al.,* 2002). In general, the settings are adjusted so that the brightest pixels are just below the saturation level, since it has been found that appropriate normalization will correct for any bias introduced during scanning (Yang *et al.,* 2002). The intensities are usually adjusted using the PMT settings of the scanner and keeping the laser power to maximum.

A number of different image analysis programs using different algorithms are available for these procedures, most of which provide a combination of automatic and manual gridding options. The choice of segmentation algorithm can strongly influence the outcome of a microarray experiment, since it defines which pixels will be included in the foreground and background measures. The most commonly used segmentation methods can be divided into four groups: fixed circle segmentation, adaptive circle segmentation, adaptive shape segmentation, and histogram segmentation (Yang *et al.,* 2002). Apart from these several other methods have also been proposed (Glasbey and Ghazal, 2003). Signal intensities for the foreground and background calculation of spot intensities have also been proposed (Yang *et al.,* 2002; Glasbey and Ghazal, 2003; Brown *et al.,* 2001). In addition to calculating the foreground and background intensities, most image analysis software also provides quality measurements of the spots. These can include statistics on within-spot pixel-to-pixel intensities, spot size, signal-to-noise ratios, and degree of pixel saturation (Brown *et al.,* 2001; Wang *et al.,* 2001). They can also be used for subsequent data filtration and analysis.

2.8.7.5.2 Flaging and log transformation

There is a substantial amount of background noise within the microarray data which is the result of problems encountered during overall processing of microarrays. These processes can yield granular and doughnut shaped spots, fabrication inconsistencies, highly variable background fluorescence, and other artifacts. Filtration can be done using a number of different criteria, for instance, the spot's size, signal-to-noise ratio and level of saturation (Wang *et al.,* 2001), within-spot pixel-to-pixel variation (Brown *et al.,* 2001), variation between repeatedly spotted clones (Tseng *et al.,* 2001; Jenssen *et al.,* 2002), weak spots that are indistinguishable from background (Yang *et al.,* 2001) and the difference between a spot's mean and median intensities (Tran *et al.,* 2002) to name a few. In addition, since cDNA microarray experiments are used to investigate relationships between samples, the obtained data are often presented as ratios between the sample intensities. In order to present and treat up- and down-regulated genes equally these ratios are also log-transformed.

2.8.7.5.3 Normalization

In order to remove variations from the data, different strategies for normalization of data are employed depending on assumptions made while performing the experiment (Yang *et al.,* 2001; Quackenbush, 2001; Smyth and Speed , 2003; Quackenbush, 2002). For a majority of microarray experiments, all genes on the array can be included in the normalization, based on the assumptions that among the many thousands of genes present on the array only a small proportion will be differentially expressed (Yang *et al.,* 2001). When smaller or more focused arrays are used, or when the target samples strongly differ, these assumptions may not be valid. One may then use a smaller subset of genes that are known to be uniformly expressed for normalization. One option is to use so-called "housekeeping genes"; genes that are involved in basic metabolic functions of the cell which are believed to be constantly expressed across a variety of conditions. Another similar approach is to find genes that are constantly expressed across the material under investigation. One such approach is the rank-invariant selection of genes described by Tseng and co-workers (Tseng *et al.,* 2001). Yet other option is to use controls for normalization. These should ideally span

the entire intensity range and be present at equal levels in either samples or a number of controls from a non-cross reactive species for which controlled amounts of RNA have been spiked into the original RNA samples. One of the simplest methods is an iterative linear regression method proposed by Finkelstein and co-workers (Finkelstein *et al.*, 2001).

The most commonly used normalization method, taking non-linearity into account, is the Lowess normalization proposed by Yang and co-workers (Yang *et al.*, 2001; Yang *et al.,* 2002). This Locally Weighted Scatter plot Smoothing regression (LOWESS) normalization strategy is becoming increasingly popular and software for Lowess normalization (the R-package) (Ihaka and Gentleman, 1996) together with detailed user instructions can be downloaded at http://www.maths.lth.se/bioinformatics/software/. The Lowess normalization method has been shown to perform well in a number of studies (Quackenbush, 2002; Tseng *et al.*, 2001; Xie *et al.,* 2004). A similar approach using a different local regression method instead of Lowess, was also proposed by Kepler *et al.,* (2002). Other normalization strategies include methods based on ANOVA (Kerr *et al.,* 2000; Wolfinger *et al.,* 2001), single value decomposition (SVD) (Alter *et al.,* 2000) and Bayesian approaches (Newton *et al.,* 2001). normalization and subsequent data analysis can be performed using the freely available statistical environment R (R Development Core Team); http://www.R-project.org) and associated packages developed for microarray analysis, e.g. Bioconductor (Gentleman *et al.,* 2004), LIMMA (Smyth, 2004; Aroma and Bengtsson, 2004).

2.8.7.5.4 Identification of co-regulated genes

The expressional ratios obtained after intensity calculation, filtration, and normalization will be used to identify the differentially expressed genes. Since the microarray experiments deal with global analysis of gene expression, there will be more than one or a list of differentially expressed genes. The primary aim of the whole experiment is to extract meaningful biological information from these lists of genes. Therefore, data mining from these differentially expressed genes is a critical and the final task. A wide range of tools are available for data mining from a microarray

experiment. , the most widely used approach among them is the use of clustering (extensively reviewed by Quackenbush, 2001). This review discusses a range of methods for organizing the genes into groups with similar expressional pattern. These unsupervised methods requiring no other information than the expression data include hierarchical clustering (Eisen *et al.*, 1998; Spellman *et al.*, 1998), k-means clustering (Tavazoie *et al.*, 1999), self-organising-maps, SOM (Tamayo *et al.*, 1999; Toronen *et al.*, 1999) and the related principal components analysis (Raychaudhuri *et al.*, 2000). One of the purposes of all of these methods is to assign potential functions to genes that have not yet been fully characterized. The basic assumption is then that genes with similar expression patterns are likely to be functionally related. Although not a rigorous assumption, the approach has proven to be successful in many cases (Cho *et al.*, 1998; Hughes *et al.*, 2000). Cluster analysis can also lead to the identification of new transcription factors involved in the transcription of a co-regulated set of genes and their binding motifs within promoter sequences (Roth *et al.*, 1998; Bussemaker *et al.*, 2001).

A number of different software packages can link the genes within microarray gene lists to appropriate GO terms and use different statistics to calculate their possible over-representation within the data. These include EASE (http://apps1.niaid.nih.gov/david; Hosack *et al.*, 2003), MAPPfinder (http://www.genmapp.org ; Doniger *et al.*, 2003) and GoMiner (http://discover.nci.nih.gov/gominer/ (Zeeberg *et al.*, 2003). A related way to mine microarray data is to look for overrepresentation of genes involved in certain known or predicted cellular pathways. Programs such as GeneSpring (Silicon Genetics, CA, USA) map differentially expressed genes to known pathways, collected in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Ogata *et al.*, 1999), and calculate scores of their overrepresentation. Other programs, such as PathwayAssist (Ariadne Genomics, MD, USA) use advanced scientific text mining tools to automatically extract biological findings from scientific literature and to build networks of molecular interactions and the freely available packages ExpressionProfiler (EMBL-EBI, Hinxton, UK) and GeneCluster (the Broad institute, MA, USA). Less user friendly, but highly flexible and versatile tools for microarray data analysis have also

been developed for the freely available statistical platform R (http://www.r-project.org) (Gentleman *et al.,* 2004; Smyth, 2004).

2.8.7.6 MIAME compliance microarray data

The need for standardized and informative presentation and exchange of microarray data has thus gradually been recognized. In 1999, the Microarray Gene Expression Data (MGED) Society was formed with the aim of drafting proposals for such standards. The resulting proposals, the Minimum Information About a Microarray Experiment (MIAME) standards, were published in 2001 (Brazma *et al.,* 2001), describing the minimum information required to ensure that microarray data can be easily interpreted and that results derived from such analyses can be independently verified. The MIAME standards have also facilitated the development of databases for storage and exchange of microarray data, including the European database ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) and the American Gene Expression Omnibus, GEO (http://www.ncbi.nlm.nih.gov/geo/).

2.8.7.7 Approaches for validation of microarray data

Northern blot and quantitative real-time PCR are methods, traditionally used for analyzing gene expression on a small scale (i.e. a few genes at a time). These methods are now valuable tools for verifying the results obtained with large scale methods. Microarrays, for example, always produce a certain level of noise and the results warrant independent confirmation. The northern blot technique which is also based on hybridization of a sample to a probe is commonly used to verify results.

Another considerably more sensitive method is based on real-time quantitative PCR (Heid *et al.,* 1996; Gibson *et al.,* 1996). Real-time PCR requires careful template titrations, good experimental design with replicates and well chosen primers and reference genes to yield reliable results (Boeckman *et al.,* 2001; Brisson et al. 2001; Meijerink *et al.,* 2001).

3. MATERIALS AND METHODS

3. Materials and Methods

3.1 Materials

3.1.1 Chemicals, reagents, and consumables

Chemicals for Bacterial Media Preparations	Hi-Media Inc Ltd India
Plasmid Miniprep purification modules	Qiagen Inc GMBH, Machery Nagel GMBH, Novagen USA.
Dyenamic ET Terminator Mix for Automated sequencing	Amersham Pharmacia Biotech, Sweden
Sequencing Capillaries, Matrix, Loading buffer for automated capillary sequencing	Amersham Pharmacia USA.
Trizol	Life technologies USA
RNA ZAP, and RNA Latter	Ambion, USA
SuperScript™ Indirect cDNA Labeling System for DNA Microarrays	Life Technologies, USA
PCR product purification modules	Machery Nagel, GMBH.
cDNA post labelling dye kit for microarrays Alexa Fluor 488, 555, 594, 647 dyes, ARES DNA labeling Kits, TOTO-dyes.	Molecular probes, USA
Super amine coated Microarray substrates	Telechem, USA
Microarray CMP3 stealth pins	Telechem, USA
First strand cDNA purification GFX columns	Life Technologies, USA
Lucidea universal score card	Amersham, USA
Lambda DNA and RNA	Takara, Japan
Hybri-slips	Sigma, Schleicher & Schuell GMBH
Yeast-t-RNA, Poly DA	Sigma, USA
Succinic anhydride, Methyl pyrilidone, 200 pint absolute alcohol	Aldrich, USA
Anchored Primer	Life Technologies, USA
Sequencing Primers	MWG Biotech GMBH
Tri-sodium citrate, SDS, Sod.acetate, Agarose, Sephadex G25, G50, Iso-propanol, glycogen, BSA, Low Temperature microarray Hybridization buffer, NaoH, Sod. Borate.	Sigma-Aldrich. USA
DNA modifying and restriction enzymes	Gibco-BRL USA, MBI fermentas, Lithuvania, New England Biolabs, USA, Bangalore Genie Ltd, Bangalore
0.2 micron filters	Nalgene, GMBH
All plastic ware for microarray	Nalgene, Nunc, USA, eppendorf GMBH, Tarsons, India.
Tris, EDTA, Agarose, DEPC, and all molecular biology grade chemicals	USB chemicals, USA.

3.1.2 Plasmid vectors

The cloning vector used to clone normalized cDNA library is PT7T3 D Pac from Amersham Pharmacia. The super promoter cassette, pE1806, was from Dr. Stanton Gelvin, Purdue University, USA. Cloning vector pBluescript SK+ was from Stratagene, La Jolla, USA.

3.1.3 Description of rice lines

Nagina 22 is an early maturing *indica* rice cultivar with deep roots and also both drought and heat tolerant. It is used as a parent in crosses for developing drought tolerant cultivars. The IR64, is a high yielding cultivar but susceptible to drought. It is one of the most popular rice cultivar in India.

3.1.4 Normalized cDNA library source

Normalized cDNA library (Reddy et al., 2000) constructed from the drought stressed N22 seedlings is the primary source for generating ESTs in this study. The vectors and procedures of normalization were described (Reddy et al., 2002).

3.1.5 Sequence repositories

The sequence repositories used in the study are given below:

1. EST sequences from N22 libraries, both earlier and present study.
2. Rice genome sequence of the *O. sativa* sub sp *japonica* cv Nipponbare generated by the International Rice Genome Sequencing Project (IRGSP)
3. Draft sequences of the *O. sativa* sub sp. *indica* cultivar generated by the Beijing Genomics Institute (BGI) available in the GenBank.
4. TIGR *japonica* rice assembly (ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules /version_1.0).
5. The rice full-length cDNA consortium sequences of Nipponbare cultivar.
6. Full-length cDNA sequences of possible candidate genes derived from *Arabidopsis* expression profiling studies from “The *Arabidopsis* Information Resource” (TAIR).

7. The nucleotide, protein and EST databases at NCBI were utilized for homology search using BLAST program.
8. Protein motifs were searched through protein data bases at InterPro (http://www.ebi.ac.uk/InterProScan/) (R. Apweiler 2001).

3.1.6 Software used for annotation

Standard sequence processing tools Phred (Ewing and Green, 1998), Phrap, and Cross_match (Smith & Waterman 1981; Gotoh, 1982) were used with CodonCode InterPhace. Homology search in the NCBI database was carried out using network client software with the DNATools interface http://www.crc.dk/dnatools. CAP3 (Huang and Madan, 1999) assembly algorithms were used to assemble the individual ESTs into clusters of sequences deriving from the same transcript as tentative consensus sequences (TCs) and singletons representing unique transcripts. Transcript mapping and localization of ESTs onto rice genome was done using cMAP software (Fang et al 2003). The unigene set derived from CAP3 assembly were searched for Protein motifs using InterProScan v3.3 (http://www.ebi.ac.uk/InterProScan/). All computations were performed on a 32-node Linux Cluster running Red Hat 9.0.

Genscan, GeneMarkHMM, Glimmer R, FGENESH, Rice Genome Automated Annotation System (Rice GAAS) were used for accurate gene prediction. Sim4, LALNVIEW V2.0 were used for identifying gene structure. TIGR tm4, Avadis (Strand genomics, India) suite was used for microarray data analysis, TIGR spot finder and Quantarray (Genomic solutions, USA) for quantification of intensities.

3.2 Methods

3.2.1 Bacterial transformation

E.coli strain DH5 α was grown at 37°C either on solid (1.5% agar) or in liquid LB medium (1% tryptone, 1% NaCl, 0.5% yeast extract). Liquid cultures were grown initially in 2 ml of LB medium in a test tube, and later in 1-liter flasks for plasmid isolation. Competent cells of *E. coli* were prepared as follows. One ml of DH 5 α cells from an overnight grown culture was inoculated in 100 ml of LB medium without antibiotic. The cells were grown till they reached an A_{600} of 0.4-0.6. Cells were then harvested into pre-cooled 50 ml falcon tubes by centrifugation at 3000 rpm for 10 min at 4°C. All the operations were performed under sterile conditions at 4°C. After the centrifugation, the cells were resuspended in 15 ml 0.1 M $CaCl_2$ and incubated on ice for 10 min. This suspension was centrifuged at 3000 rpm for 10 min. The resultant pellet was resuspended in 4 ml of 0.1 M $CaCl_2$ (in 10% glycerol) for every 100 ml of original culture, dispensed into 200 µl aliquots, frozen and stored at –70°C for future use.

Transformation of the competent cells was done as follows: Frozen *E. coli* cells were thawed on ice to which 1 ng of plasmid DNA or 100 ng of ligation mix were added. The suspension was carefully mixed with pipette tip and incubated on ice for 30 min. A heat shock of 42°C for 45 sec was applied followed by incubation on ice for another 2 min. 800 µl of LB medium was added and the bacterial suspension was incubated at 37°C with shaking for 1 h. Aliquots of the suspension were spread evenly on LB supplemented with an appropriate antibiotic. The plates were incubated at 37°C overnight. Single colonies were picked up following day and inoculated for plasmid mini preparation.

3.2.2 Expressed Sequence Tags generation

3.2.2.1 Template preparation

The Normalized cDNA library was transformed in chemically competent *E.coli* DH5α by heat shock at 42°C and kept for expression for 30 min and then plated at low density on Luria-bertani medium containing 20µg/ml ampicillin, and were incubated at 37°C overnight. Individual colonies were selected randomly and inoculated in 5ml Luria-Bertani broth containing 20µg/ml ampicillin and incubated at 37°C overnight in an

orbital shaker at 250 RPM. The overnight grown culture was used in plasmid DNA preparation for nucleotide sequencing. Plasmid DNA was prepared using a DNA purification column (Qiagen Plasmid Isolation Kit). These plasmid preparations represent clones of amplified normalized cDNA library. The quality and concentration of the template plasmid DNA was controlled on 1% agarose (USB biochemicals) gels and was used directly for sequencing.

3.2.2.2 Sequencing

3.2.2.2.1 Preparing of sequencing reaction

For EST generation, 3' end single pass sequencing of 7740 cDNA clones using M_{13} (-40) Reverse primer 5' CGCCGAGGTTTTCCCAGTCACGAC 3', and M_{13} (-20) Reverse primer 5' GTAAAACGACGGCCAGTG 3', were performed on automated high throughput 48 capillary genetic analysis system, MegaBACE 500 (Amersham Pharmacia). DYEnamic ET terminator chemistry (Amersham Pharmacia) was used for sequencing reaction set-up, post-sequencing reaction clean-up and loading samples on MegaBACE was adjusted to suite our conditions and the rest were followed according to the manufacturer's instructions. We have standardized the optimal template concentration to be used, which ranged from 300 ng to 400 ng for a quarter strength reaction for a vector size in between 3kb to 6kb. The primer is diluted in water or in buffer containing no more than 0.1 mM EDTA. And for each reaction 2.5 mM of sequencing grade primer is used (PAGE purified or HPLC purified). For each template to be sequenced, we use a $1/4^{th}$ strength Dyenamic ET terminator reagent premix (Amersham Pharmacia), and 1/2 strength dilution buffer, and 2.5 mM primer and 400 ng of plasmid DNA or 15 to 60 ng of PCR product are added depending on the size of the template, and set for cycle PCR with the following cycling parameters: 95^oC for 20 seconds, 50^oC for 15 seconds and 60^oC for 1 minute for 30 cycles and stopped the thermal cycler with rapid thermal ramp to 4^oC and hold. In our studies, 30 cycles are found to be optimum for achieving low noise and good read length.

3.2.2.2.2 Post reaction clean-up

The post reaction clean-up is either done by ethanol precipitation or by passing the samples through sephadex G 25 (Amersham Pharmacia). Ethanol precipitation method worked effectively both in terms of cost and quality. Millipore filter membrane in a 96 well format was used to load appropriate bed volume of sephadex G 25 (Amersham Pharmacia) and was allowed for 2 days to swell in milliQ water. Before the samples were loaded, the excess water was drained down by centrifuging at 1000 RPM for 2 min in order to minimize the sample loss. The bed was pre-wetted with 10µl of sterile milliQ water and the samples were loaded on top and spun at 3100g for 5 min to retain the samples. Samples were directly injected into the MegaBace capillary sequencer at 3 Kva for 40 sec. Ethanol precipitation was carried out by adding 1µl of 7.5M Ammonium acetate and 2.5 volumes (27.5µl) of 100% ethanol for a final concentration of 75%. The samples were incubated at -20^{o}C for 30 min. The samples were centrifuged at either room temperature or 4° C in a micro centrifuge for 15 min at ~ 12000 rpm. The 96-well plates are centrifuged for 30 min at 2500 x g or more. If plates are centrifuged at a lower relative centrifugal force, the time of centrifugation was increased. We routinely use 3100 x g for 30 min. The supernatant was removed by aspiration or by a brief inverted spin. The pellet was washed in 100-200 µl of 70% ethanol to remove any leftover salt that can have a deleterious effect during injection in capillary electrophoresis. The supernatant was removed by aspiration or by a brief inverted spin. Further, the pellet was vacuum-dried or air-dried care was taken that the samples are not over-dried. Each pellet was resuspend in 10 µl of loading solution (70% formamide) and vortexed vigorously for 10-20 sec to ensure complete resuspension. The plates were briefly centrifuged to collect the sample at the bottom of the tube/well and to remove any bubbles.

3.2.2.2.3 Injection and run parameters

For the greatest reproducibility, we use a voltage of 2 KV for injection (2 to 3 KV is standard). Injection time can be varied widely during optimization with injections as short as 5 sec or as long as 400 sec being equally successful. In the present study an injection time of 50KV for 50 Sec was found to be optimal and therefore used routinely. Using our standard run conditions of 200 min at 6 KV, an average read-length of greater than 800 bases with 98% accuracy were reproducibly obtained.

3.2.2.3 Sequence processing and analysis

The raw data was exported from MegaBace in SCF format, and the raw sequences were base called using Phred. The low quality regions present at the beginning and end of each sequence were trimmed using a Phred 20 cutoff value. Vector screening was performed using the cross_match program with Codoncode InterPhace software. Sequences were edited for the removal of oligo dT tracks and other contaminants. A batch file of ESTs having greater than 100 bp length of sequence reads were submitted to the NCBI db EST division of GenBank. After the rice genome sequence was largely completed (IRGSP, 2005), all ESTs from this project were compared to the genomic sequence. All sequences that did not exhibit excellent nucleotide homology with the Nipponbare genomic sequence were removed from GenBank, with the assumption that they were most likely to be derived from microbial contaminants.

3.2.3 Annotation

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

3.2.4 EST clustering and transcript mapping

Phrap and CAP3 (Huang and Madan, 1999) assembly algorithms were used to assemble the individual ESTs into clusters of sequences derived from the same transcript as tentative consensus sequences (TCs) and singletons representing unique transcripts. Assembled N22 unigenes were aligned to the TIGR *japonica* rice assembly (ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules /version_1.0) using BLASTN (Altschul et al., 1990) genes were annotated if they met the following stringency criteria: E-value $\leq 1x10^{-10}$; percent identity $\geq$ 95%; and alignment length $\geq$50bp. Functional annotations were then associated to mapped unigenes. These

positions were displayed using CMAP software (Fang et al., 2003). Also the putative functions of the stress responsive genes were also displayed along with the unigenes.

3.2.5 Data mining, compilation, and identification of putative stress responsive ESTs

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

3.2.6 Isolation of drought inducible transcriptional factors DREB1B and CRT/DRE binding factor from Nagina 22

Through *in-silico* analysis of the ESTs generated the sequence showing homology to the Os-DREB 1b Accession Number AF300972 was sequenced from both end to get the full length 862 nucleotides. The 1128 bp sequence of CRT/DRE binding factor was amplified from genomic DNA of N_{22} using the forward primer carrying *Sal*I restriction site (5' ACG CGT CGA CCC ATC ATC ACC GAG ATC GAC TCG AC – 3') and the reverse primer with *Not*I restriction site (5'- ATA AGA ATG CGG CCG CTC ATT GTT CGC TCA CTG GGA G – 3'). PCR amplification was carried out in a MJ Research thermal cycler. The amplified product of 1.2 kb was gel eluted and cloned in TA vector pTZ57R using InsT/Aclone PCR product cloning kit (MBI Fermentas). The cloned product, 1128 nucleotides in length, was sequence confirmed on MegaBace 500 using M_{13} forward and reverses primers. The CRT/DRE fragment was gel eluted by restriction digestion with *Sal*I and *Not*I and was cloned directionally in pBluescript.

3.2.6.1 Construction of plant expression vector for DREB1B

The DREB1B fragment was released from pT_7T_3Pac vector by digesting the vector with *HindIII* and *EcoRI* present in the 3' UTR of the gene and was cloned

directionally in pBluescript at *Hind*III and *Eco*RI sites. Further the 840 nucleotide DREB1B cDNA was excised from pBluescript KS^+ by digesting with *Xho*I and *Xba*I. The fragment was gel excised and cloned in the same sites in pE 1806 under the super promoter sequences from mannopine synthase (mas) and the octopine synthase (ocs) genes. The restriction map of the recombinant binary vector, pE1806 and their identities were confirmed by restriction analysis using suitable restriction endonucleases.

3.2.6.2 Construction of the bacterial expression vector for DREB1B

The 862 bp sequence of DREB 1b was gel eluted from the pT_7T_3 vector using *Not* I and *Xho* I and cloned directionally in frame at *Not* I and *Xho* I sites of T_7 RNA polymerase expression vector pET 28a. The restriction map of the recombinant plasmid and their identities were confirmed by restriction analysis using suitable restriction endonucleases.

3.2.6.3 Construction of the bacterial expression vector for CRT/DRE binding factor

The 947 bp open reading frame from the Os CRT/DRE cDNA cloned in pBlue Script was PCR amplified using the forward primer carrying *Nde* I site (5'- GGG AAT TCC ATA TGG AGA AGA ACA CCG CCG C –3') and the reverse primer with *Not* I restriction site (5'- ATA AGA ATG CGG CCG CTC ATT GTT CGC TCA CTG GGA G – 3') PCR amplification was carried out in a MJ Research Thermal Cycler. The amplified product of 1.0 kb was gel eluted and cloned in TA vector pTZ57R using InsT/Aclone PCR product cloning kit (MBI Fermentas). The cloned product was confirmed by sequencing and a 947 bp length fragment was obtained (Fig-3). The CRT/DRE fragment was gel eluted by restriction digestion with *Nde* I and *Not* I and was cloned directionally in frame at *Nde* I and *Not* I sites of T_7 RNA polymerase expression vector, pET 28a. The restriction map of the recombinant plasmid and their identities were confirmed by restriction analysis using suitable restriction endonucleases.

3.2.6.4 Restriction digestion and ligation of fragments

Restriction digestion of the plasmid DNA and the amplicons was carried out in a total volume of 20 µl containing 0.5 µg of plasmid DNA, 2 µl of appropriate 10 X restriction enzyme buffer, 1 µl BSA (1 µg/µl, if necessary), 15 (or 16) µl of double distilled water and 1 Unit of restriction enzyme. The reaction mixture was incubated at

37º C (or at an appropriate temperature according to the restriction enzyme used) for 1 hour and the digestion pattern was analyzed on 1% agarose gel. The following principle was used to calculate the concentration of fragment and plasmid DNA needed (3:1 ratio of fragment to vector) for ligation reaction. Fragment size/vector size X 100 X 3 where "100" denotes the amount (ng) of vector DNA and "3" denotes the number of times of fragment DNA (ng) required for ligation. The ligation reaction was carried out in a total reaction volume of 20 µl containing 100 ng of restriction digested vector DNA, appropriate amount of fragment DNA, 2 µl 10 X ligase buffer, 1 µl of T4 DNA ligase (5 U/µl) and sterile double distilled water to make up the volume. The reaction was incubated at 16º C for 2 h or overnight where necessary. After completion of the reaction, an aliquot of 10 µl was used for transformation.

3.2.7 Simulated field drought stress

3.2.7.1 Experimental design

In order to conduct a field experiment for gene expression profiling, we have constructed a rain out shelter with specifications described below. A plot of 13 feet width and 16 feet length was first excavated and filled with soil up to 5 feet depth. A sand bed of 1 ft was laid across the length in a slanting manner such that the end of the plot has just 1 inch sand, so that seepage of water will be towards sand end. Also, all the sides were filled with sand so as to leach the water out of the experimental plot. This has allowed us to maintain drought conditions and avoid any accidental seepage. Eight plots of the size 3ft x 4ft were partitioned and constructed with concrete walls with water proof agents to avoid seepage of water across plots. The inner walls were of 5ft and the outer walls facing the exterior soil apart from the black clayey soil were of 3ft long. The plots are raised enough to avoid accidental overflow from the control plots to the experimental.

All plots were filled (5ft depth) with black clayey soil in which the rice crop is usually cultivated. To maintain fertility, recommended amount of farm yard manure was added. Rainout shelter was constructed to cover the plots during rainy days to keep the plants protected from rain during experiment. The rainout shelter was provided with free moving rail tracks to move forward and backward. Proper care was taken to avoid any

shade of rainout shelter on plants. Drought stress induction was monitored through high-precision digital irrigation controllers procured from Netafim Inc (Israel). Irrigation was given taking Plot A1 /AR1 as 100% field capacity, which is a control plot in this experiment, and irrigation was set up in the other plots as shown in the table below, with moisture probes/Timer controllers using digital controllers.

No.	Plot	Genotype	Irrigation Levels	Recovery
1	A1	N22	100%	NR
2	A2	N22	70%	NR
3	A3	N22	40%	NR
4	A4	N22	10%	NR
5	A1	IR 64	100%	NR
5	AR1	N22	100%	-
6	AR2	N22	70%	-
7	AR3	N22	40%	87 days
8	AR4	N22	10%	80 days

NR- Not subjected to Recovery
AR1 –Control Recovery

Table 1: Irrigation plan to for induction of different degrees of drought stress

3.2.7.2 Rice seedling culture, stress treatments, and transplantation

Rice seeds were imbibed in water, surface sterilized with 5% sodium hypochlorite (v/v) for five minutes, thoroughly washed with sterile water and germinated in water upon filter papers in dark. Two-day-old germinated seeds were transferred to growth chambers and were supplemented with Hogland's medium. The average temperature during seedling culture was 28 ± 1^{o} C and a photoperiod of 16 hours light and 8 hours dark was maintained. Seedling trays were examined twice daily to maintain constant moisture content. Fifteen day old seedlings were transplanted into field for drought stress induction. The transplanted seedlings were grown in 100% field capacity with 2cm standing water until day 30; from day 31 field drought stress was initiated at 100%, 60%, 40% and 10% field capacities in A1, A2, A3, and A4. It took nearly 8 days for the onset of the actual drought for the experimental plots to reach to their designated field capacities from 100% FC. Prior to this actual experiment a trail run was performed to

check whether the electronic digital irrigation controllers are functioning properly all through the experiment. Samples were collected to check differential gene expression in various drought stress regimes based on the developmental stages of the plant. The genotype under study, N22, is a short duration crop and completes within 95 to 100 days. The vegetative stage ends on 50 to 55 days, pre-bloom stage is between 55 to 60 days, milking stage is between 60 to 70 days and seed maturing stage is between 70 to 100 days. Samples were collected as mentioned in table 2.

Stage	**Developmen-tal Type**	**Tissue Type**	**Day**	**Sampling**							
I	Vegetative	Leaf	45	A1	A2	A3	A4	-	-	-	-
I	Vegetative	Leaf	52	A1	A2	A3	A4	-	-	-	-
I	Vegetative	Leaf	55	A1	A2	A3	A4	-	-	-	-
II	Pre-bloom stage	Leaf	59	A1	A2	A3	A4	-	-	-	-
II	Pre-bloom stage	Panicle	59	A1	A2	A3	A4	-	-	-	-
III	Milking stage	Leaf	63	A1	A2	A3	A4	-	-	-	-
III	Milking stage	Panicle	63	A1	A2	A3	A4	-	-	-	-
III	Milking stage	Leaf	67	A1	A2	A3	A4	-	-	-	-
III	Milking stage	Panicle	67	A1	A2	A3	A4	-	-	-	-
IV	Seed setting	Leaf	71	A1	A2	A3	A4	-	-	-	-
IV	Seed setting	Panicle	71	A1	A2	A3	A4	-	-	-	-
IV	Seed setting	Leaf	75	A1	A2	A3	A4	-	-	-	AR4
REC	Recovery	Leaf	71	-	-	-	-	-	-	-	AR4
REC	Recovery	Leaf	75	-	-	-	-	-	-	-	AR4
REC	Recovery	Panicle	75	-	-	-	-	-	-	-	AR4
REC	Recovery	Leaf	78	-	-	-	-	-	AR2	AR3	AR4
REC	Recovery	Panicle	78	-	-	-	-	-	AR2	AR3	AR4
REC	Recovery	Leaf	81	-	-	-	-	-	AR2	AR3	AR4
REC	Recovery	Panicle	81	-	-	-	-	-	AR2	AR3	AR4

Table 2: Sample collection of field drought stressed samples

Table 2 legend: A1- 100%FC, A2- 60%FC, A3- 10%, AR2- 60% FC Recovery, AR3- 40% FC Recovery, AR4- 10% FC Recovery, REC-Recovery.

3.2.7.3 Seedling stress experiments: lab experiments

N22 seedlings were grown as mentioned above and fifteen-day-old seedlings were treated with either 20% poly ethylene glycol (PEG 8000) or 150 mM sodium chloride or 100 μM ABA or dehydration, solutions prepared in Hogland's solution. Control plants received only Hogland's solution. The treatments were given three hours after beginning of light period. Samples were collected at 1, 2, 3, 6, 12, 24, and 48 hours after treatment. The collected samples were frozen in liquid nitrogen and stored in -70 ^{o}C for RNA isolation and microarray expressional analysis.

3.2.8 cDNA template amplification for array fabrication

The amplified clones of cDNA normalized library from which ESTs have been generated were used for amplification of cDNA inserts from the plasmid templates. 5664, plasmid templates have been diluted to an average 5ng/μl. The cDNA inserts were amplified using 15 ng of the plasmid template with M_{13} (-40) forward primer 5' CGCCGAGGTTTTCCCAGTCACGAC 3', and M_{13} (-20) Reverse primer 5' GTAAAACGACGGCCAGTG 3', in an 100μl reaction volume with 10 mM each of forward and reverse primer, 5 mM dNTPs, 5U Taq DNA polymerase and the PCR conditions were, an initial denaturation at 95^{o} C for 5 min followed by 94^{o}C, 1 min; 55^{o}C, 1 min; and 72^{o}C, 4 min for 35 cycles. The amplified products were screened on 1% agarose gel electrophoresis. The PCR products were scored for low conc., absence, and nonspecific amplifications and were documented accordingly. These were purified using silica membrane based bind, wash and elute columns procured from Machery Nagel using vacuum manifold in 96 well plate formats, according to the Manufacturer's instructions. The PCR products were eluted in MilliQ water and were reformatted into 384 well gridder compatible plates (Nunc 384 well square well plates) such that a 96 well plate clones are printed in series in 2 subarrays. 7μl of the PCR product was aliquoted in each well of 384 well plate to which 7μl of DMSO was added to make a final concentration of 50% DMSO (v/v). Likewise 1152 pearl millet cDNA clones amplified from differential cDNA λ phage libraries using PBSK^{+} forward and reverse primers were purified using GFX columns (Amersham Pharmacia) were also resuspended to a final concentration of 50% DMSO. In all 6816 cDNA clones of rice and pearl millet, 24

artificial genes designed from yeast genomic sequence information procured from Amersham Biosciences USA, and λ DNA procured from Takara Inc Japan, which were aliquoted in 48 different wells, were used for arraying on superamine coated glass microarray slides.

3.2.9 Array fabrication

3.2.9.1 Array design

cDNA microarray chip consisting of 15552 features with 6144 ESTs from drought stress induced cDNA library of *Oryza sativa* cv Nagina22, 1152 ESTs from drought stressed *Pennisetum glaucum,* 24 yeast artificial genes , λ DNA controls were printed in duplicate on superamine coated glass substrate using Omnigrid (GeneMachines). The machine was calibrated for 384 Nunc square well plates, CMP3 print head and stealth printing pins and 75mm x 25 mm microarray glass slides. We have used 16 pins altogether in 4 x 4 format, which gave us an option to print 4 subarrays along Y direction and 4 arrays in X direction. The origin offset values set were 3000 microns in X direction and 2000 microns in Y direction. We have used 225 μ spot to spot spacing in a subarray. The number of spots printed in a subarray is 18 clones in X direction and 18 clones in Y direction, the number of clones printed in one subarray is 18 x 18 = 324 stamps, in which 162 clones have been printed onto each subarray. The maximum number of arrays which would be printed were 4 in which we limited to 3 arrays, with a spacing of 500 microns each, in which a maximum number of 7776 clones in duplicate could be printed. The above format allowed us to print 7680 samples in duplicate including blanks.

For printing, the sample intake was for every 100 slides, the dip time being 1000 milliseconds, with each dip the sample is taken in by capillary action and is sufficient enough to print 200 stamps. The sample volume which is delivered in each stamping is around 0.5 to 0.7 ηl. After each printing the pins were sonicated for 2000 milliseconds, followed by washing in sterile water for 3000 milliseconds, and dried under vacuum for 2000 milliseconds which is repeated thrice, the final cycle has an extended drying time of 3000 milliseconds. After each round the 384 well plate and the pins were also sonicated

thrice for 3000 milliseconds. In order to avoid the possibility of sample sticking to the walls of the pins, during each dip, which can cause excess sample being deposited on slides resulting in run over of samples, the pins were initially blotted on a poly-L-Lysine coated blot pad 10 times with 150 micron spot to spot spacing. The time taken by the robotic gridder to print 7680 samples in duplicate is 36 hours. The cDNAs were printed at a relative humidity between 45 to 50%, which is very critical in printing good and reproducible microarrays. Also we found that before printing, calibration of wash station, dry station, sonicator, and slides across the platen of the slide holder is of utmost importance for uniform spot morphology in all the printed slides.

3.2.9.2 Array printing

Microarraying was by robotic systems built by Gene Machines, California (Genomic Solutions). We have used aminosilane coated slides from Telechem, SuperAmine, aminosilane coated glass microscope slides have been found to be the most consistent, when printed in 50% DMSO as a printing solution at 45% relative humidity and 72^{o}F (22^{o}C). We have used quill pens (Arrayit ChipMaker3 microspotting pins from Telechem International Inc.) which were very durable and could reproduce high-quality spots with good precision. All arrays were printed with the same set of ChipMaker 3 pins for more than six months. Following printing, the slides were rehydrated on a steam coming from a 500ml conical flask containing MilliQ water, which was kept boiling on a hot plate. Care was taken such that all slides were uniformly rehydrated; the slides were passed over the steam with the array side down 2 to 3 times until a thin layer of mist was formed on the printed side. Extreme care was taken not to rehydrate such that the spots run into each other. Following rehydration, the slides were immediately snap dried for 5 sec on a hot glass plate. The slides were allowed to dry and spotted DNA was fixed to slide by UV cross-linking at 320 mJ using a Stratalinker (Stratagene, Cat# 400071) and by baking at 80º C for 2 hrs. Printed slides were stored in a light-tight box in a bench-top dessicator, with desiccant and stored at 4^{o}C until they were used for hybridization.

3.2.10 Extraction of total cellular RNA

Total RNA was isolated from leaf and panicle samples collected at different time intervals after stress treatment using Trizol (Chirgwin *et al.*, 1979; Chomczynski and

Sacchi, 1987) from Life Technologies with minor modifications. The harvested tissues were stored immediately in RNA later (Ambion) and stored at 4 ^{o}C for 24 hours after which the RNA latter was removed and the samples were quick frozen in liquid nitrogen and stored at –70^{o}C. Five grams of freshly harvested or frozen tissues were ground in liquid nitrogen to a fine powder and transferred to 50ml tubes with 15ml of Trizol. The samples were suspended in Trizol and incubated at room temperature for 5 minutes. Cellular debris was removed by centrifugation at 2700*g* for 15min at 4^{o}C. The supernatant was transferred to a fresh RNAase free 50 ml oakridge tube. To it 3 ml of chloroform (0.2 ml/1 ml Trizol) was added and shaken vigorously for 1 minute, and incubated at room temperature for 5 min. The samples were centrifuged at 2700 *g* for 15min at 4^{o}C to remove proteins and high molecular weight DNA. The top supernatant layer containing RNA was carefully removed without disturbing the protein layer and was transferred to a fresh RNAase free Oakridge tube. To this, an equal volume of isopropanol was added to precipitate the RNA and incubated at RT for 15 min. The sample was centrifuged at 21,000*g* for 15 minutes, the supernatant discarded, and the pellet resuspended in 70% ethanol. The RNA pellet was resuspended in diethylpyrocarbonate (DEPC) treated water. The quality of the total RNA was determined through 1.2% denatured agarose formaldehyde gel and concentration estimated spectrophotometrically. For labeling at "University of Georgia" the samples were resuspended in 1/10th volume DEPC treated sodium acetate and 2.5 volumes ethanol. Prior to labeling the precipitated RNA was spun at 12000 RPM for 10 min and washed with 70% ethanol. The RNA was dried in a Savant rotary vapor trap and resuspended in RNAase free milliQ water for labeling.

3.2.11 RNA labeling

3.2.11.1 Labeling procedure

Probes for microarray analysis were prepared from RNA templates by first strand synthesis of cDNA containing amino-allyl-labeled nucleotide dCTP and amino-hexyl modified dATP for both control and stress samples, followed by a covalent coupling to the NHS-ester of the appropriate Cyanine fluor. This method has proven highly efficient for labeling RNAs from all kinds of tissues from rice for expression analysis. Among the reverse transcriptases used, including AMV and MMLV, we found the Superscripț III RT

(Life Technologies; Cat# L-1014-02) generates probes with significantly greater activity. Coupling of modified cDNA with fluorescent probes was done by photostable Alexafluor fluorescent dyes (we have used Alexafluor 488, and 555 for control labeling and Alexafluor 594, and 647 NHS ester dyes for coupling stressed samples). The photosensitivity of dyes was minimized by avoiding exposure to light during labeling, hybridization, washing, and scanning processes. Upon receipt, these NHS ester dyes were stored in light- proof tubes and stored at -20^{0}C until needed. All reactions were carried out in foil-wrapped tubes and all hybridizations and washes in foil-wrapped containers.

3.2.11.2 First strand amino modified cDNA synthesis recipe step 1

Prior to labeling, the quality of total RNA for both control and stressed experimental sample, were quantitated and suspended in 40 µg of total RNA in not more than 10 µl DEPC treated water. Anchored oligo DT primer was used; this oligo (20mer oligo DT) primed with equal concentrations of A, C, G, or T, allowed proper priming of the primer to the 5' end of the cDNA. This helped us in avoiding reverse transcription of long stretches of dT in cDNA. Five microgram of anchored primer was mixed with 40 µg of total RNA and the reaction was incubated for 5 min at 70° C, following which the reaction was snap cooled for 1 min on ice, to which the following components were added, 5X first strand buffer - 6.0 µl, 10 mM dNTP mix (contains modified aminoallyl dUTP and aminohexyl dATP - 1.5µl, RNAse out 20units/µl - 1.5µl Super script RT III 400 units/µl - 2.0µl). DEPC treated water was added to make-up a final volume of 30.0 µl. The above reaction mixture was incubated for 2 hours at 42 C in a thermocycler.

3.2.11.3 Alkaline hydrolysis and neutralization step 2

To the reaction from step 1, 15 µl 1N NaoH was added to denature the RNA through hydrolysis from RNA-DNA hybrid and was incubated for 10 min at 70 ^{o}C for complete RNA denaturation. To this 15µl of 1N HCl was added to neutralize the pH and mixed gently. To the above reaction, 20µl sodium acetate, pH 5.2, was added and mixed gently. Once the cDNA was generated with amino-modified nucleotides, it was purified to remove unincorporated dNTPs and hydrolyzed RNA, using bind wash elute SNAP columns (Life Technologies) as per manufacturers instructions following which, the

cDNA was ethanol precipitated by adding 1/10th volume sodium acetate and 2.5 volumes of ethanol.

3.2.11.4 First strand cDNA dye coupling step 3

The aminoallyl and aminohexyl modified cDNA precipitated in step2 was further centrifuged at 12000 RPM for 10 min to pellet it down and the pellet was washed with 70% ethanol and dried down in Speedvac (Savant). The pellet was dissolved in 6 µl sterile water. For coupling of the dye, the dye vials (Alexa fluor dyes) were allowed to come down to room temperature in order to avoid condensation of water inside the vial. The dyes were dissolved in 4 µl of DMSO. For coupling, 10 µl of 2X coupling buffer (sodium bicarbonate) was added to modified cDNA to bring the final concentration to 1X, to the above reaction 4µl of the dissolved dye was added and allowed to couple overnight in dark at room temperature. Further, the cDNA was purified by bind wash elute SNAP columns (Life Technologies) as per manufacturers instructions to remove any unreacted free dye molecules.

3.2.11.5 Quantification and assessing labeling efficiency of the cDNA

The relative efficiency of a labeling reaction was evaluated by calculating the approximate ratio of bases to dye molecules. This ratio has been determined by measuring the absorbance of the amino modified cDNA at 260 nm and the absorbance of the dye at its absorbance maximum (max) (λ_{max} for Alexa 488, 555, 594, and 647 are 492nm, 555nm, 588nm and 650 nm), and by using the Beer-Lambert law: $A = \epsilon \times$ path length (cm) $\times$ concentration (M), where ϵ, is the extinction coefficient in $cm^{-1}M^{-1}$. The absorbance measurements were used to determine the concentration of nucleic acid in the sample. Most fluorescent dyes absorb light at 260 nm as well as at their λ_{max}. To obtain an accurate absorbance measurement for the nucleic acid, it was therefore necessary to account for the dye absorbance using a correction factor (CF260). Hence using the CF_{260} values in the following equation: $A_{base} = A_{260} - (A_{dye} \times CF_{260})$ exact absorbance A_{base} was calculated.

3.2.11.6 Measurement of base:dye ratio

The ratio of bases to dye molecules was calculated using the following equation:

Base:Dye = (A_{base} × $Є_{dye}$) / (A_{dye} × $Є_{base}$) where A_{base} is the dye corrected absorbance calculated previously, $Є_{dye}$ is the extinction coefficient for the fluorescent dye and $Є_{base}$ is the average extinction coefficient for a base in single-stranded DNA (ssDNA). The values of $Є_{dye}$ for Alexa 488, 555, 594, and 647 are 62,000, 150,000, 80,400 and 239,000, whereas the values for $Є_{base}$ is 8919.

3.2.11.7 Measuring the concentration of nucleic acid

The absorbance values, A_{260} and A_{dye}, and the Beer-Lambert law were used to determine the concentration of cDNA. In order to obtain an accurate measurement for a dye-labeled nucleic acid, a dye-corrected absorbance value (A_{base}) was used, in the following formula [N.A.] (mg/mL) = (A_{base} × MWbase) / ($Є_{base}$ × path length), whereas A_{base} is the dye corrected absorbance calculated previously, MW_{base} is average molecular weight of base which is 330 g/mol, $Є_{base}$ is 8919, and path length is 1 cm. Using the above formulas, the amino modified cDNA concentrations and efficiency of dye incorporation was calculated for all the samples for hybridization to probes on the arrays with equal concentrations of cDNA, which helped in employing efficient normalization strategies.

3.2.12 Hybridization

The dye labeled cDNA once quantified, equal amounts of cDNA from tester and reference samples were pooled and was precipitate by adding $1/10^{th}$ of sodium acetate and 2.5 volumes of ethanol, the cDNA populations were spun at 8000 RPM for 10 min, to pellet down the cDNA. The pellet was washed with 70% ethanol, dried and the pellet was dissolved in 50µl of low temperature hybridization buffer from Sigma at room temperature.

3.2.12.1 Preparation of slides for pre-hybridization/hybridization

Our protocols employed yielded reproducible and high-quality hybridizations while maximizing the measured fluorescence from the array. Aminosilane coated slides bind DNA with high efficiency. Prior to hybridization, the free amine groups on the slide were blocked or inactivated to avoid nonspecific binding of labeled cDNA to the slide. The slides were blocked in sodium borate and succinic anhydride blocking reagent.

Microarray slides stored at 4 ^{o}C were allowed to come to room temperature. The required number of slides were dipped in the above solution and kept for shaking on an orbital shaker for 15 min. Slides were placed in opposite direction such that the non printed side touches each other. This was scaled up if more than 2 slides were processed using a Lipshaw slide holder and a jar. After 15 min, the slides were washed in 0.1% SDS for 2 min in a clean glass jar (Lipshaw dish staining assembly), the SDS solution was removed and a brief 10sec wash was performed with milliQ water Slides were transferred to a fresh dish jar containing 2XSSC, washed twice with 2X SSC for 2 min each.

3.2.12.2 Denaturing and fixing

Denaturation was performed by dipping the slides in boiling hot water at 95 ^{o}C and incubating them for 2 min. The denatured DNA was fixed by transferring the slides in to ice cold 200ml 95% ethanol (Aldrich). Slides were retained in ice cold ethanol for 2 to 3 min before being transferred to a 50ml falcon tube with a tissue paper at the bottom. The slides were spun without cap at 1000 RPM for 5 min at RT After spinning the slides were immediately used for hybridization.

3.2.12.3 Target hybridization

The labeled cDNA dissolved in low temperature hybridization buffer from Sigma. Two blocking agents, yeast tRNA (Sigma) at a concentration of 20 μg per hybridization, and 2 μg poly dA (Sigma) per hybridization were used to block non-specific hybridization. The pre-hybridized slides were heated to bring them to 45 ^{o}C for hybridization. After denaturation the labeled probe was applied to a pre-hybridized microarray slide and covered with 22mm x 60mm plastic hybri-slips (Sigma, Schleicher, and Schuell Gmbh). Once the coverslip is properly placed without any air bubble trapped between the slide and the coverslip, the slides were inverted (array side down) and kept inside a 50ml falcon tube, which were used as a makeshift hybridization chamber. A small piece of tissue paper was soaked with 1ml of 3x SSC and was kept inside the bottom of falcon tube to maintain humidity during hybridization and to avoid drying of the target. The Falcon tube was placed in hybridization bottle (Amersham Pharmacia) and dipped in hot water bath set at 50 ^{o}C. Weight rings were used between the bottles to

avoid floating of the hybridization bottles. Hybridization was allowed for 16 hours in dark at 50 Deg C.

The slides from the water bath were taken out after 16 hours of hybridization and dipped in wash buffer 1 (1x SSC, 0.03% SDS). They were shaken gently until the coverslip come off by itself, Once the coverslips come off, the slides were transferred to wash buffer 2 (0.2x SSC) and washed for 2 min by gently shaking and then transferred to wash buffer 3 (0.05x SSC). After gentle agitation for 5 min, the slides were then transferred to a 50ml Falcon tube and with its lid open were spun at 1000 RPM for 5 min to dry out completely for scanning.

3.2.13 Data collection, normalization, and analysis

Scanning of the hybridized slide was done using ScanArray 3000. This scanner uses 4 Helium-Neon lasers operating at 488 nm, 543 nm, 594 nm and 633 nm to excite Alexa 488, 555, 594 and 647, respectively. Hybridized slides were scanned first in the high wave length channel (594 nm and 647 nm), and then the lower wavelength (488 nm and 555 nm). Data from each fluorescence channel was collected and stored as a separate 16-bit TIFF image. To quantify the right intensities slides were scanned as soon as possible after washing. The slides were scanned at appropriate PMT gain and 100% laser power, that is, low to high at 30 microns scan resolution, then appropriate PMT settings for both the dye wavelength was standardized. Scanning was done at PMTs where we could observe 20 to 30 white spots, such that the overall intensities are good. This was determined by observing the RI plot of intensities. The slides were scanned at 10micron. During quantification we used the highest threshold limits for spot intensity calculation and ANOVA performed for these low PMT and High PMT technical replicates.

Scanned images generated from microarray experiment were processed with TIGR Spotfinder and Quantarray, to identify and quantify the spots on the array. The two images were loaded on to the two different channels of Spotfinder and Quantarray. Grids were generated taking printing style into consideration. Overlay of the two channel intensities revealed the hybridization was good and more or less similar intensities were observed in adjacent in-slide replicas. A total of 48 blocks with 324 features in each were generated with 12 rows and 4 columns in each covering all the 15552 features including controls. A sum of 15552 cells were generated with 5 - 23 pixel spot size adjusted as to

cover all the features of the superimposed slide. In house custom perl scripts were written to facilitate the annotation to the features on the array. Annotation including gene function, chromosome BAC/PAC alignment, and map position of the ESTs, unigenes cluster, functional classification, and redundancy were provided to facilitate further analysis with ease. Otsu threshold algorithm has been adopted to estimate the spot size since it does not require close estimate of spot size and requires just one value reasonably close. All the grids were processed and spots were quantified in the cells eliminating local background. The spots with poor quality in either of the slide were flagged off and were eliminated during the normalization. If the intensity values approached to a maximum (2^{16}-1) per pixel for a 16 bit scanner comparison, those are no longer meaningful as the array elements become saturated and all these were excluded. Quantified spots from each analysis were exported for further analysis.

3.2.14 Data normalization

Total intensity normalization was first applied to all the data points assuming the initial RNA taken for the hybridization is equal in both the samples. It has been observed based on pre and post normalized RI plot (log ratio intensity vs log intensity product) the factor calculation for scaling up the data was not appropriate. 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 of normalized plots. LOWESS normalization was performed which considered all the genes on the array, considered house keeping genes and internal control in calculating normalization factor.

3.2.15 Clustering of co-regulated genes

3.2.15.1 Analysis across the stress regimes

The genes were clustered with Hierarchical classification after normalizing expression data and unsupervised classification was followed. In all the data points from day 45 to day 81 of the development stages A2, A3, A4, vectors were taken for the (A2 denotes 60% FC, A3 denotes 40 % FC, A4 denote 15% FC) stress regime analysis of the expression data. Hierarchical clustering was carried out for either genes or experiments at an instance on whole slide data, as it is a memory intensive process. Average linkage

clustering technique was followed for hierarchical clustering. Significant gene clusters were identified by k-means clustering to ~10-20 groups based on hierarchical clustering results.

3.2.15.2 Analysis across the developmental stages:

The intensities captured at each developmental stage from 45-75 days with in the stress regime were considered for time series analysis and carried out at all the stress regimes viz. A2 (60%FC),A3 (40%FC),A4 (15%FC). Panicle sample from day 59 to 75 were analyzed at each stress regime for expression patterns in panicle tissue in stress across the developmental stages.

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 PT_7T_3Pac 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 green house 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

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 *Os*DREB1B, *Os*DREB1A, *Os*DREB2 and *Os*CRT/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 pT_7T_3Pac 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 30th, 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 64th 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 through out 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. 27 b, c) reveals expression pattern of 162 genes showing atleast 3 fold up-regulation from day 45 to day 59 and their after these genes show atleast 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 there after these subset of genes show atleast 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 atleast 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 atleast 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 upto day 55 but shows atleast 3

fold up-regulation by day 59 and thereon shows atleast 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 atleast 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 atleast 498 genes showing altered genes expression. These genes fall under 4 clusters represented in cluster 7, 8, 9 and 10. Cluster 7 (Fig. 29b, c) represents 156 genes show 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 atleast 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 atleast 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 atleast 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 >10 fold up-regulation by day 71. These genes drop suddenly in the same proportion (>10 fold down regulation). In cluster 7_3, which represent 40 genes, show atleast 3 fold changes on day 52, 59 and 71 and these genes show atleast 4 fold down regulation on day 63 and 67 compared to day 59 or day 7, as expected this set of genes show a sudden drop by > 4 fold by day 75.

Cluster 8 (Fig. 29d, e) represents 187 genes showing no expressional difference till day 55 and shows atleast 3 fold up-regulation by day 59 and continue to be up to day 63, the same set of genes starts down-regulating >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 >10 fold up-regulation by day 71 and drops by atleast 10 fold by day 75. Cluster 10 (Fig. 29h, i)

represents 81 genes showing no considerable change till day 55 and shows atleast 7 fold up-regulation on day 59 and continue to be same till day 63 and thereafter starts down-regulating by atleast 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 atleast 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 there after 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 atleast 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 atleast 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 atleast 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 atleast 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 atleast 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 atleast 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 atleast 192 genes which are expressed

differentially in stress regimes. These include a set of 154 genes showing up-regulation by atleast 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 atleast 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 atleast 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 101genes, shows a atleast 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 atleast 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.

4.5.3.3.3 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 cluster10) that are up-regulated differently across stress regimes, and 6 clusters (also fall into 3 broad clusters cluster5, cluster6 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. cluster5, cluster6 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. The only

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 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 shown 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,

& h). 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. In

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.

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 high light 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 green house 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 green house 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 N22unigene 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 protienases 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, H_2O_2 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*

al., 1997). 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 H_2O_2 to water: catalase (BU673091, BU673392), ascorbate peroxidase (APX) (BU673288) showing homology to tomato APX (A3251882) and manganese superoxide dismutase (MnSOD) (BU673715) which is 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, Mc Kersie *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 H_2O_2, 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/CBF-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 reveled 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 atleast 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 atleast 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 DNAses, 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-transglycoxylase, 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 through out 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 atleast 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, trans membrane 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 1 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.

6. SUMMARY

6. Summary

In the work described in this thesis, an EST based strategy has been successfully adapted to identify putative candidate genes of drought response in rice. As a powerful extension of this strategy, a cDNA microarray has been constructed and gene expression profiling under field drought stress has been carried out. The results from these original experiments and interpretations are summarized below. The microarray experiments described here may well constitute the first report on field drought stress expression profiles in indica rice.

Utilizing a high quality, normalized cDNA library constructed from drought stressed seedlings of an indica cultivar, N22, we have generated large scale ESTs in a very cost effective manner and analyzed. More than 6000 high quality Expressed Sequence Tags of rice have been deposited in the public domain (NCBI dbEST division; www.ncbi.nlm.nih.gov: Accession Numbers BU672765 to BU673915 & CB964418 to CB967504). The EST data presented in this thesis is probably the first global overview of the transcripts that are expressed in rice under drought stress. The ESTs generated here by an efficient HTP single pass sequencing with a success rate of more than 85%, are of high Phred score of > 30 with an average read length of more than 500bp, and also include rare transcripts. A N22 unigene set of 1677 sequences has been constructed and physically mapped onto rice chromosomes. Annotation of the N22 unigene sets from these ESTs led to the identification of 589 genes putatively associated with stress response functions and 334 novel genes. We have taken a comparative genomic approach to identify these stress responsive genes utilizing the earlier validated SRGs through published information from various sources of rice, arabidopsis, and other model plants, including data from microarray experiments, in addition to our own. Among these, only 57% of the genes had known function. The function of the remaining 43% of genes are yet be defined in rice and also in other plants. Classification of these annotated EST on the basis of metabolic function revealed that transcription factors were effectively captured, as they constitute 17% the stress responsive genes. Among the total annotated ESTs, TFs represents the third largest group. Genes associated with protein synthesis,

cellular metabolism, signal transduction are other major classes are uncovered in this study. Relative abundance of transcripts was analyzed and genes encoding metallothioniens, anti-oxidative enzymes and stress responsive proteins were found to be predominant. Expression profiling of unknown category of genes has led to preliminary knowledge of how these genes are co-regulated with other classes of genes through which it is possible to predict the nature of these unknown genes.

A high quality cDNA microarray was constructed with 15552 features representing both rice N22 (6144) and pearl millet (1152) cDNA differential clones made from high salt and drought stressed leaf tissue. By using a well trained dataset of controls, we could achieve excellent normalization across experiments (52 data points), thus showcasing the efficacy of overall strategies employed in this highly reliable microarray experiment. We achieved in generating MIAME compliance microarray data for all the data points studied. Expression profiling was carried out through a drought stress experiment under field conditions across selected developmental stages and different stress regimes. This is probably the first report on gene expression profiling under field drought conditions carried out through out the development and growth of the plant.

Microarray data analysis revealed interesting patterns of gene expression uncovering clusters of co-regulated genes during different drought stress regimes. Our results high light the fact that the changes in gene expression, as reported in earlier such experiments are not limited to up and down-regulation. The expression patterns are rather complex as a function of stage of development and growth, and severity of drought stress. The gene expression profiles revealed ~ 90 novel genes which are co-regulated along with SRGs which may play a major role in drought stress response. These novel genes represent a major proportion of the highly up-regulated genes. The present study enabled us to identify more than 150 genes of unknown function and of hypothetical nature. The data revealed there are atleast 200 to 250 functionally characterized genes which show varying patterns of expression when analyzed across developmental stages and different stress regimes. This data will be an invaluable resource in deciphering new or unknown pathways of drought stress response in rice. Considerable number (about

17% of the arrayed) cDNAs encoding known transcription factors and their expressional changes have been identified. Further, the expression profiling studies uncovered a number of rice orthologous candidate stress responsive genes on the basis of rice-pearl millet cross hybridization on the array.

This study used a comparative genomics approach to define the genetic basis of drought stress response among cereals and thus represents with a modest beginning using rice and pearl millet. The time series gene expression profiles revealed that most of the genes which are overexpressed under severe stress were through heterologous hybridization of rice RNA with that of pearl millet clones, indicating that there seem to be many rice genes whose role in conferring drought tolerance in rice is yet to be elucidated. Nevertheless, the extensive results presented in this thesis will be a valuable resources for discovering genes and defining their role in drought tolerance in rice. Currently, these resources have been extensively used in SNP discovery and development of EST based gene specific polymorphic markers in rice, sorghum, and pearl millet. This information will also be useful in dissecting well characterized QTLs and identifying candidate genes for drought tolerance.

7. REFERENCES

7. References

Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D,Shinozaki K (1997) Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell 9:1859−1868.

Abe S, Katagiri T, Saito-Hisaminato A, Usami S, Inoue Y, Tsunoda T, Nakamura Y. (2003). Identification of CRYM as a candidate responsible for non syndromic deafness, through cDNA microarray analysis of human cochlear and vestibular tissues. Am. J. Hum. Genet.72 (1):73-82.

Adams MD, Kelley JM, Gocayne JD, et al., (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science. 21; 252:1651-1656.

Adams MD, Kerlavage AR,Fleischmann RD, Fuldner RA, Bult CJ, et al, 1995, Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 377:3-174.

Adams, M.D., Kerlavage, A.R., Fields, C., Venter, J.C., 1993. 3400 new expressed sequence tags identify diversity of transcripts in human brain. Nat. Genet. 4, 256–267.

Adams, M.D., Kerlavage, A.R., Fields, C., Venter, J.C., 1993.3400 new expressed sequence tags identify diversity of transcripts in human brain. Nat. Genet. 4, 256–267.vierling and Kimpel., 1992.

Aharoni, A., et al. (2000). Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell 12, 647–661.

Allona I, Quinn M, Shoop E, Swope K, St Cyr S, Carlis J, Riedl J, Retzel E, Campbell MM, Sederoff R, Whetten RW (1998) Analysis of xylem formation in pine by cDNA sequencing. Proceedings of the National Academy of Sciences of the United States of America 95: 9693-9698.

Alter, O., P. O. Brown, et al. (2000). Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci U S A 97(18): 10101-6.

Altmann T, Felix G, Jessop A, Kauschmann A, Uwer U, Pena Cortes H, Wilmitzer L: (1995) Ac/Ds transposon mutagenesis in *Arabidopsis thaliana*: mutant spectrum and frequency of Ds insertion mutants. Mol Gen Genet 247: 646–652.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. Journal of Molecular Biology 215: 403-410.

Altschul, Stephen, F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Alwine, J. C., D. J. Kemp, et al. (1979). Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzyloxymethyl paper. Methods Enzymol 68: 220-42.

Audic S, Claverie JM. The significance of digital gene expression profiles (1997). Genome Research 7, 986-995

Augenlicht, L. H., M. Z. Wahrman, et al. (1987). Expression of cloned sequences in biopsies of human colonic tissue and in colonic carcinoma cells induced to differentiate in vitro. Cancer Res 47(22): 6017-21.

Babu et al., 2003 R. C. Babu, B. D. Nguyen, V. Chamarerk, P. Shanmugasundaram, P. Chezhian, P.Jeyaprakash, S. K. Ganesh, A. Palchamy, S. Sadasivam, S. Sarkarung, L. J. Wade, and H. T. Nguyen,Genetic Analysis of Drought Resistance in Rice by Molecular Markers: Association between Secondary Traits and Field Performance Crop Sci., July 1, 2003; 43(4): 1457 – 1469.

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

Babu PR., PhD thesis 2005. Ph. D thesis. Development of gene resources in rice (*Oryza. sativa cv* N22) University of Hyderabad.

Bankfalvi, A. and Schmid, K.W. (1994) [In situ hybridization. Theoretical principles and practical applications]. Pathologe 15, 76-84.
Barrett, J. C. and E. S. Kawasaki (2003). Microarrays: the use of oligonucleotides and cDNA for the analysis of gene expression. Drug Discov Today 8(3): 134-41.
Barry GF (2001) Plant Physiol. 125: 1164-1165.
Beaucage, S. L. (2001). Strategies in the preparation of DNA oligonucleotide arrays for diagnostic applications. Curr Med Chem 8(10): 1213-44.
Bengtsson, H. aroma (2004)- An R Object-oriented Microarray Analysis environment. [http://www.maths.lth.se/publications/].
Bennett J, Arumugam K, Lafitte R, Wen J, Rudd S, Bruskiewich RM. 2002. IRRI Drought Stress Panicle cDNA Library. www.ncbi.nlm.nih.gov/dbest.
Bennetzen JL, Coleman C, Liu R, Ma J, Ramakrishna W. 2004. Consistent over-estimation of gene number in complex plant genomes. Curr Opin Plant Biol 7, 732–736.
Bernstein, J. A., A. B. Khodursky, et al. (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci U S A 99(15): 9697-702.
Bevan et al., (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature: 391, 485 – 488
Bhalerao R, Keskitalo J, Sterky F, Erlandsson R, Bjorkbacka H, Birve SJ, Karlsson J, Gardestrom P, Gustafsson P, Lundeberg J, Jansson S (2003) Gene expression in autumn leaves. Plant Physiology 131: 430-442
Blanc., 2000 Blanc G., Barakat A., Guyot R., Cooke R., Delseny M., Extensive duplication and reshuffling in the *Arabidopsis thaliana* genome, Plant Cell 12 (2000) 1093–1101.
Blum A (1988) Plant breeding for stress environments. Boca Raton, FL. CRC Press.
Blum, A. and Pnuel, Y. 1990. Physiological attributes associated with drought resistance of wheat cultivars in a Mediterranean environment. Aust. Jour. Agric. Res. 41:799-810.
Boeckman, F., M. Brisson, et al. "Real-Time PCR: General consideration." Technote 2593, BioRad Laboratories Inc. Boeuf, S., M. Klingenspor, et al. (2001). "Differential gene expression in white and brown preadipocytes." Physiol Genomics 7(1): 15-25
Bohnert HJ., Jensen, RG., Metabolic engineering for increased salt tolerance-the next step Aust.J plant physiol 1996. 23:661-87.
Bohnert, H.J., Nelson, D.E., and Jensen, R.G. (1995). Adaptations to environmental stresses. Plant Cell 7, 1099–1111.
Bonaldo, M.F. et al., (1996) Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res 6: 791–806.
Bovy A, de Vos R, Kemper M, Schijlen E, Almenar Pertejo M, Muir S, Collins G, Robinson S, Verhoeyen M, Hughes S et al.: High- flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. Plant Cell 2002, 14:2509-2526.
Bowers, J.E., Chapman, B.A., Rong, J.K. and Paterson, A.H. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422: 433–438.
Boyer JS. 1982. Plant productivity and environment. Science 218, 443.
Bray E A et al., 2004. Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana* Journal of Experimental Botany, Vol. 55, No. 407,pp. 2331–2341, November 2004 doi:10.1093/jxb/erh270
Bray E, Bailey-Serres J, Weretilnyk E. 2000. Responses to abiotic stresses. In Biochemistry and Molecular Biology of Plants. Edited by Buchanan B, Gruissem W, Jones R Rockville: American Society of Plant Biologists pp.1158-1203.
Bray E, Bailey-Serres J, Weretilnyk E. 2000. Responses to abiotic stresses. In Biochemistry and Molecular Biology of Plants. Edited by Buchanan B, Gruissem W, Rockville JR: American Society of Plant Biologists pp.1158-1203.
Bray, E.A. (1997). Plant responses to water deficit. Trends in Plant Sci 2: 48-54.
Brazma, A., P. Hingamp, et al. (2001). "Minimum information about a microarray experiment (MIAME)-toward standards for microarray data." Nat Genet 29(4): 365-71.

Brenner, S., M. Johnson, et al. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18(6): 630-4.
Brown, C. S., P. C. Goodwin, et al. (2001). Image metrics in the statistical analysis of DNA microarray data. Proc Natl Acad Sci U S A 98(16): 8944-9.
Bulyk, M. L., X. Huang, et al. (2001). Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. Proc Natl Acad Sci U S A 98(13): 7158-63.
Bussemaker, H. J., H. Li, et al. (2001). Regulatory element detection using correlation with expression. Nat Genet 27(2): 167-71.
Catlin D (1992) Rice in deep water. Macmillan, London
Cawley, S., S. Bekiranov, et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116(4): 499-509.
Chandrasekhar A., 2005. Ph. D thesis. Molecular Mapping Using SSRs and EST-PCR Based Markers, and Identification of Quantitative Trait Loci (QTLs) for Drought Tolerance in a Doubled Haploid Population of Rice *(Oryza sativa L.*).
Chang-Hyo Goh, Hong Gil Nam, Yu Shin Park (2003) Stress memory in plants: a negative regulation of stomatal response and transient induction of rd22 gene to light in abscisic acid-entrained Arabidopsis plants; Volume 36: Issue 2 Plant Journal
Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B. et al., (2002) An integrated physical and genetic map of the rice genome. Plant Cell. 14: 537-45.
Chen, H., M. Centola, et al. (1998). "Characterization of gene expression in resting and activated mast cells." J Exp Med 188(9): 1657-68.
Cheong YH, Chang HS, Gupta R, Wang X, Zhu T, Luan S (2002)Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses inArabidopsis. Plant Physiol 129:661–677
Cheung, V.G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R. and Childs, G. (1999) Making and reading microarrays. Nat. Genet. 21, 15-19.
Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., Rutter, W.J., 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294–5299.
Cho, R. J., M. J. Campbell, et al. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2(1): 65-73.
Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. Anal. Biochem. 162, 156–159.
Churchill,G. (2002) Fundamentals of experimental designs for cDNA microarray. Nat.Genet., 32, 490–495.
Clark, T. A., C. W. Sugnet, et al. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 296(5569): 907-10.
Cohen, S.N., Chang, A.C., Boyer, H.W. and Helling, R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. Proc. Natl. Acad. Sci. USA 70, 3240-3244
Collins, F.S., 1995. Positional cloning moves from perditional to traditional. Nat. Genet. 9, 347–350.
Cushman JC, Bohnert HJ (2000) Genomic approaches to plant stress tolerance. Curr Opin Plant Biol 3:117–124.
Dagerlind, A., Friberg, K., Bean, A.J. and Hökfelt, T. (1992) Sensitive mRNA detection using unfixed tissue: combined radioactive and non-radioactive in situ hybridization histochemistry.Histochemistry 98, 39-49.
DeRisi, J.L., Iyer, V.R., Brown, P.O., 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680–686.
DeRisi, J.L., Iyer, V.R., Brown, P.O., 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680–686.
Desikan R, Mackerness S, Hancock JT, Neill SJ (2001) Regulation of the Arabidopsis transcriptome by oxidative stress. Plant Physiol 127:159–172
Deutsch S, Iseli C, Bucher P, Antonarakis SE, Scott HS (2001) A cSNP map and database for human chromosome 21. Genome Res 11: 300-307
Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol35:3-15.

Dey, M.M., and Upadhyaya, H.K. (1996) Yield loss due to drought, cold, and submergence in Asia. In Rice research in Asia, progress and priorities. Edited by R.E. Evenson, R.W. Herdt, and M. Hossain. Oxford University Press, Cary, N.C. pp. 231–242.

Deyholos M, Galbraith DW (2001) High-density microarrays for gene expression analysis. Cytometry 43: 229–238

Diehl, F., S. Grahlmann, et al. (2001). "Manufacturing DNA microarrays of high spot homogeneity and reduced background signal." Nucleic Acids Res 29(7): E38.54.

Doniger, S. W., N. Salomonis, et al. (2003). MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. Genome Biol 4(1): R7.

Dudley, A. M., J. Aach, et al. (2002). Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. Proc Natl Acad Sci U S A 99(11): 7554-9.

Dudoit, S., Y. H. Yang, et al. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Statistica Sinica 12: 111-140.

Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. and Trent, J.M. (1999) Expression profiling using cDNA microarrays. Nat. Genet. 21, 10-14.

E M Southern, S C Case-Green, J K Elder, M Johnson, K U Mir, L Wang, and J C Williams. Arrays of complementary oligonucleotides for analysing the hybridisation behaviour of nucleic acids. Nucleic Acids Research, 22(8):1368{73, April 1994

Echenique V, Stamova B, Wolters P, Lazo G, Carollo L, Dubcovsky J.Frequencies of Ty1- copia and Ty3- gypsy retroelements within the Triticeae EST databases. Theor Appl Genet. 2002 Apr;104(5):840-844. Epub 2002 Feb 08.

Edwards, D. and Batley, J. 2004. Plant bioinformatics: from genome to phenome. Trends Biotech. 22: 232–237.

Eisen and Brown, 1999 Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C., Trent, J.M., Staudt, L.M., Hudson, J., Jr., Boguski, M.S., Lashkari, D., Shalon, D., Botstein, D. and Brown, P.O. (1999) Thetranscriptional program in the response of human fibroblasts to serum. Science 283, 83-87.

Eisen, Spellman et al. 1998 Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D. and Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell 9, 3273-3297.

Ekins, R. and Chu, F.W. (1999) Microarrays: their origins and applications. Trends Biotechnol. 17, 217–218

Ekins, R., F. Chu, et al. (1989). High specific activity chemiluminescent and fluorescent markers: their potential application to high sensitivity and 'multi-analyte' immunoassays. J Biolumin Chemilumin 4(1): 59-78.

Eliana Drenkard, Brent G. Richter, Steve Rozen, Lisa M. Stutius, Nathaniel A. Angell, Michael Mindrinos, Raymond J. Cho, Peter J. Oefner, Ronald W. Davis and Frederick M. Ausubel A Simple Procedure for the Analysis of Single Nucleotide Polymorphisms facilitates Map Based Cloning in Arabidopsis. (2000) Plant Physiology, Vol 124; 1483-1492.

Etherington JR (1988) The aims and development of plant ecology. In Environment and plant ecology, 2nd ed., (Manchester: John Wiley and sons, Ltd.), pp. 1-4.

Ewing B, Green P: Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Research 8:186-194 (1998).

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

Fan, J., X. Yang, et al. (2002). Global analysis of stress-regulated mRNA turnover by using cDNA arrays. Proc Natl Acad Sci U S A 99(16): 10611-6.

Fang Z, Polacco M, Chen S, Schroeder S, Hancock D, Sanchez H, Coe E. (2003), cMap: the comparative genetic map viewer, Bioinformatics. 2003 Feb 12; 19(3): 416-7.

Feltus FA, Wan J, Schulze SR, Estill JC, Paterson AH, A SNP Resource for Rice Genetics and Breeding Based on subspecies indica and japonica Genome Alignments, Genome Research 14:1812-1819, 2004.

Feng Q, Zhang Y, Hao P, Wang S, Fu G, et al. (2002) Sequence and analysis of rice chromosome 4. Nature 420, 316–320.

Fernandes J, Brendel V, Gai X, Lal S, Chandler VL, Elumalai RP, et al. 2002. Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and microarray hybridization. Plant Physiol 1283, 896–910

Fickett,J. (1996) Finding genes by computer: the state of the art. Trends in Genetics 12(8), 316–320.

Finkelstein, D. B., J. Gollub, et al. (2001). Iterative linear regression by sector. Methods of microarray data analysis. Papers from CAMDA 2000. S. M. Lin and K. F. Johnson, Kluwer Academic: 57-68.

Fodor, S. P., J. L. Read, et al. (1991). Light-directed, spatially addressable parallel chemical synthesis. Science 251(4995): 767-73.

Forster, T., D. Roy, et al. (2003). Experiments using microarray technology: limitations and standard operating procedures. J Endocrinol 178(2): 195-204.

Fowler S, Thomashow MF (2002) Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF response pathway. Plant Cell 14:1675–1690.

Fray RG, Wallace A, Grierson D , Lycett GW (1994) Nucleotide sequence and expression of a ripening and water stress-related cDNA from tomato with homology to the MIP class membrane channel proteins. Plant Mol. Biol. 24: 539-543

Fukai, S., and M. Cooper. (1995). Development of drought-resistant cultivars using physiomorphological traits in rice. Field Crops Res. 40:67–86.

Gaasterland,T. and Sensen ,C.W. (1996) MAGPIE: automated genome interpretation. Trends in Genetics 12(2), 76–78

Gale, M.D. and Devos, K.M. 1998. Comparative genetics in the grasses. Proc. Natl. Acad. Sci. USA 95: 1971–1974.

Gall, J. G. and M. L. Pardue (1969). Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proc Natl Acad Sci U S A 63(2): 378-83.Gall and Pardue, 1969

Gentleman, R., V. Carey, et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biology 5(10): R80

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

Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J (2000) Microarray analysis of developing Arabidopsis seeds. Plant Physiol 124:1570–1581.

Glasbey, C. A. and P. Ghazal (2003). Combinatorial image analysis of DNA microarray features. Bioinformatics 19(2): 194-203.

Goda H, Shimada Y, Asami T, Fujioka S, Yoshida S (2002) Microarray analysis of brassinosteroid-regulated genes in Arabidopsis. Plant Physiol130: 1–16

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

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

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

Gu YQ, Wildermuth MC, Chakravarthy S, Loh YT, Yang C, He X, Han Y, Martin GB: Tomato transcription factors pti4, pti5, and pti6 activate defense responses when expressed in Arabidopsis. Plant Cell 2002, 4:817-831.

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

Gygi, S. P., Y. Rochon, et al. (1999). Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19(3): 1720-30.

Harrington, C.A., Rosenow, C. and Retief, J. (2000) Monitoring gene expression using DNA microarrays. Curr. Opin. Microbiol. 3, 285-291.

Harris, D.W., Kenrick, M.K., Pither, R.J., Anson, J.G. and Jones, D.A. (1996) Development of a high-volume in situ mRNA hybridization assay for the quantification of gene expression utilizing scintillating microplates. Anal. Biochem. 243, 249-256.

Harushima, Y., M. Yano, A. Shomura, M. Sato, T. Shimano, Y. Kuboki, T. Yamamoto, S.Y. Lin, B.A. Antonio, A. Parco, H. Kajiya, N. Huang, K. Yamamoto, Y. Nagamura, N. Kurata, G.S. Khush, and T. Sasaki. (1998). A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148:479–494

Hashimoto, S., T. Suzuki, et al. (1999). "Serial analysis of gene expression in human monocyte-derived dendritic cells." Blood 94(3): 845-52

Haynes, P.A., Gygi, S.P., Figeys, D. and Aebersold, R. (1998) Proteome analysis: biological assay or data archive? Electrophoresis 19, 1862-1871.

Hegde, P., R. Qi, et al. (2000). "A concise guide to cDNA microarray analysis." Biotechniques 29(3): 548-50, 552-4, 556

Heid, C. A., J. Stevens, et al. (1996). "Real time quantitative PCR." Genome Res 6(10): 986-94

Heller, R.A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E., and Davis, R.W. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc. Natl. Acad. Sci. USA 94, 2150–2155.72 The Plant Cell

Hibi, K., Q. Liu, et al. (1998). "Serial analysis of gene expression in non-small cell lung cancer." Cancer Res 58(24): 5690-4

Hillier, L.D et al., 1996. Generation and analysis of 280 000 human expressed sequence tags. Genome Res. 6, 807–828.

Hirochika, H., Guiderdoni, E., An, G., Hsing, Y.I., Eun, M.Y., Han, C.D., Upadhyaya, N., Ramachandran, S., Zhang, Q., Pereira, A., Sundaresan, V. and Leung, H. 2004. Rice mutant resources for gene discovery. Plant Mol. Biol. 54: 325–334.

Hogenesch JB, Ching KA, Batalov S, Su AI,Walker JR, Zhou Y, Kay SA, Schultz PG, Cooke MP: A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. Cell 2001, 106:413-415.

Holmberg, N. and Bu¨ low, L. (1998) Improving stress tolerance in plants by gene transfer. Trends Plant Sci. 3, 61–66. Ingram, J. and Bartels, D. (1996) The molecular basis of dehydration tolerance in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47, 377–403.

Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. and Pease, L.R. (1989) Engineering hybrid geneswithout the use of restriction enzymes: gene splicing by overlap extension. Gene 77, 61-68.

Hosack, D. A., G. Dennis, Jr., et al. (2003). Identifying biological themes within lists of genes with EASE. Genome Biol 4(10): R70.

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

Hu, G. K., S. J. Madore, et al. (2001). Predicting splice variant from DNA chip expression data. Genome Res 11(7): 1237-45.

Huang X, Madan A. 1999. CAP3: A DNA Sequence Assembly Program. Genome Research 9, 868-877.

Huang X, Madan A. 1999. CAP3: A DNA Sequence Assembly Program. Genome Research 9, 868-877.

Hughes, T. R., M. J. Marton, et al. (2000). Functional discovery via a compendium of expression profiles. Cell 102(1): 109-26.

Hughes, T. R., M. Mao, et al. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 19(4): 342-7.

Ihaka, R. and R. Gentleman (1996). "R: A language for data analysis and graphics." Journal of Computational and Graphical statistics 5: 299-314.

Ingram, J. and Bartels, D. 1996. The molecular basis of dehydration tolerance in plants. Annual Review Plant Physiology Plant Mol Biol. 47:377-403

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

Ishii, Hashimoto et al. 2000 Ishii, M., S. Hashimoto, et al. (2000). "Direct comparison of GeneChip and SAGE on the quantitative accuracy in transcript profiling analysis." Genomics 68(2): 136-43.

Ishitani M, Xiong L, Stevenson B, and Zhu JK (1997) Genetic analysis of osmotic and cold stress signal transduction in Arabidopsis: interaction and convergence of abscisic acid dependent and abscisic adic-independent pathways. Plant Cell 9: 1935-1949.

Iyer, Horak et al. 2001 Iyer, V. R., C. E. Horak, et al. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 409(6819): 533-8.

Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C., Trent, J.M., Staudt, L.M., Hudson, J., Jr., Boguski, M.S., Lashkari, D., Shalon, D., Botstein, D. and Brown, P.O. (1999) The transcriptional program in the response of human fibroblasts to serum. Science 283, 83-87.

J. Carl Barrett 2003 J. Carl Barrett and Ernest S. Kawasaki Microarrays: the use of oligonucleotides and cDNA for the analysis of gene expression DDT Vol. 8, No. 3 February 2003 PII: S1359-6446(02)02578-3

Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T, Thomashow MF: Components of the Arabidopsis C-repeat/dehydration-responsive element binding factorcold-response pathway are conserved in Brassica napus and other plant species. Plant Physiol 2001, 127:910-917.

Jaglo-Ottosen, K.R., Gilmour, S.J., Zarka, D.G., Schabenberger, O., and Thomashow, M.F. (1998). Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. Science, 280: 104–106.

Jenssen, T. K., M. Langaas, et al. (2002). Analysis of repeatability in spotted cDNA microarrays. Nucleic Acids Res 30(14): 3235-44.

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.H., Jeong, D.H., Lee, J., Kim, C., Jang, S., Lee, S., Yang, K., Nam, J., An, J., Han, M.J., Sung, R.J., Choi, H.S., Yu, J.H., Choi, J.H., Cho, S.Y., Cha, S.S., Kim, S.I. and An, G. 2000. T-DNA insertional mutagenesis for functional genomics of rice. Plant J. 22: 561–570

John, I. et al. (1997) Cloning and characterization of tomato leaf senescence-related cDNAs. Plant Mol. Biol. 33, 641–651

Johnson, J. M., J. Castle, et al. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302(5653): 2141-4.

Ju, J., Ruan, C., Fuller, C.W., Glazer, A.N., Mathies, R.A., 1995. Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. Proc. Natl. Acad. Sci. USA 92, 4347–4351.

Kan, Z., E. C. Rouchka, W. R. Gish, et al. (2001). Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. Genome Res 11(5): 889-900.

Kapranov, P., S. E. Cawley, et al. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. Science 296(5569): 916-9.

Karsten, S. L., V. M. Van Deerlin, et al. (2002). An evaluation of tyramide signal amplification and archived fixed and frozen tissue in microarray gene expression analysis. Nucleic Acids Res 30(2): E4.

Kashiwagi, H. and K. Uchida (2000). Genome-wide profiling of gene amplification and deletion in cancer. Hum Cell 13(3): 135-41.

Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, and Shinozaki K (1999) Improving plant drought, salt and freezing tolerance by gene transfer of a single stress inducible transcription factor. Nature Biotechnol. 17:287-291.

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

Kepler, T. B., L. Crosby, et al. (2002). Normalization and analysis of DNA microarray data by self-consistency and local regression. Genome Biol 3(7): RESEARCH0037.

Kerr, M. K. and G. A. Churchill (2001). Experimental design for gene expression microarrays. Biostatistics 2(2): 183-201.

Kerr, M. K., M. Martin, et al. (2000). Analysis of variance for gene expression microarray data. J Comput Biol 7(6): 819-37.

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Molec Biol 35: 25–34.

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

Khush GS. 2001. Green revolution: the way forward.Nature Rev. Genetics 2:815-822.

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N ,Yazaki J, Ishikawa M, Yamada H, Ooka H. et al. 2003. Collection, mapping and annotation of over 28,000 cDNA clones from japonica rice. Science 301, 376-379.

Kim H, Snesrud EC, Haas B, Cheung F, Town CD, Quackenbush J (2003) Gene expression analyses of Arabidopsis chromosome 2 using a genomic DNA amplicon microarray. Genome Res 13:327–340.

Klok EJ, Wilson IW, Wilson D, Chapman SC, Ewing RM, Somerville SC, Peacock WJ, Dolferus R, Dennis ES (2002) Expression profile analysis of the low-oxygen response in Arabidopsis root cultures. Plant Cell 14:2481–2494.

Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphisms from large EST collections in barley (Hordeum vulgareL.). Mol Genet Genomics

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

Lee, M. L., F. C. Kuo, et al. (2000). "Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations." Proc Natl Acad Sci U S A 97(18): 9834-9

Lennon, G. G. and H. Lehrach (1991). Hybridization analyses of arrayed cDNA libraries. Trends Genet 7(10): 314-7.

Lichtenthaler HK. (1996) J. Plant Physiol. 148, 4-14.

Lilley JM, and Ludlow MM (1996) Expression of osmotic adjustment and desiccation tolerance in diverse rice lines. Field Crops Res. 48: 185-197.

Lipshutz, R. J., S. P. Fodor, et al. (1999). High density synthetic oligonucleotide arrays. Nat Genet 21(1 Suppl): 20-4.

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

Lockhart, D. J., H. Dong, et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14(13): 1675-80.

Lockhart., 1999 D J Lockhart and E A Winzeler. Genomics, gene expression and DNA arrays. Nature,

Lukowitz W, Gillmor CS, Scheible WR, 2000 Positional Cloning in Arabidopsis: Why it feels good to have a Genome Initiative working for you. Plant Physiology: 123: 795-805

Lundgren, M., A. Andersson, et al. (2004). Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. Proc Natl Acad Sci U S A 101(18): 7046-51.

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

Markandeya G, V.B.Reddy., Alex Feltus., A.H.Paterson., A.R.Reddy. 2005 Functional genomics of drought stress response in rice: transcript mapping of annotated unigenes of indica rice (*Oryza sativa* L. cv. Nagina 22) Vol 89, No 3 Aug 10, 2005, current science.

Markandeya G., A.R.Reddy et al., Expression Profiling of Drought Responsive Genes In An Indica Rice Using cDNA Microarrays, PAGE – XIV, San Diego CA, USA 2006.

Markandeya Gorantla, P. R. Babu, V. B. Reddy Lachagari, A. M. M. Reddy, Ramakrishna Wusirika, Jeffrey L. Bennetzen, Arjula R. Reddy., Identification of stress responsive genes in an indica rice (*Oryza sativa* L.) using ESTs generated from drought stressed seedlings, Communicated, Journal of Experimental Botany

Markandeya, G.; Ravindra Babu,P.; Venkat Reddy, B.; Nagabhushana Ithal.; Chandra Sekhar, A.; Bennetzen,J.L.; Ramakrishna,W.; Reddy, A.R. (2003) ESTs from a normalized cDNA library of drought stressed rice seedlings (*Oryza sativa* L.cv Nagina 22) Gene Bank Accession: CB964418 to CB967504 www.ncbi.nlm.nih.gov/dbest.

Markandeya,G., Reddy,A.R., Ramakrishna,W., Nagabhushana,I.,Ravindra Babu,P., Madana Mohan Reddy,A., Chandra Sekhar,A.,Bennetzen,J.L. (2002) Novel EST enrichment with normalized cDNA libraries from drought stressed rice (*Oryza sativa* L.cv Nagina 22) Gene Bank Accession: BU672765 to BU673915 www.ncbi.nlm.nih.gov/dbest**.**

Marra, M. A., L. Hillier, et al. (1998). Expressed sequence tags--ESTablishing bridges between genomes. Trends Genet 14(1): 4-7.

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

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

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

Matters GL, Scandalios J (1986) Effect of the free radical-generating herbicide paraquat on the expressing superoxide dismutase (SOD) genes in maize. Biochim Biophys Acta 882: 29-38

Maxam, A. M. and W. Gilbert (1977). A new method for sequencing DNA. Proc Natl Acad Sci U S A 74(2): 560-4.

McKersie B.D., Leshem Y.Y. Stress and stress coping in cultivated plants.(1994). London: Kluwer Academic Publishers,

McKersie BD, Bowley SR, Harjanto E, Leprince O. 1996. Water-Deficit Tolerance and Field Performance of Transgenic Alfalfa Overexpressing Superoxide Dismutase. Plant Physiology 111, 1177-1181.

Meijerink, J., C. Mandigers, et al. (2001). "A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR." J Mol Diagn 3(2): 55-61

Michalek W, Weschke W, Pleissner KP, Graner A. 2002. EST analysis in barley defines a unigene set comprising 4,000 genes. Theoretical and Applied Genetics. 104, 97–103.

Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K. and Hirochika, H. 2003. Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 16: 1771–1780

Murray, V., 1989. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res. 17, 8889.

Nagamura, Y., et al. (1995). Conservation of duplicated segments between rice chromosome-11 and chromosome-12. Breed. Sci. 45, 373–376

Nasu Shinobu, Junko Suzuki, Rieko Ohta, Kana Hasegawa, Rika YUI, Noriyuki Kitazawa, Lisa Monna, and Yuzo Minobe, Search for and Analysis of Single Nucleotide Polymorphisms (SNPs) in Rice (*Oryza sativa*, Oryza rufipogan) and Establishment of SNP Markers. 2002

Nature Genetics 1999 (1999). *Nature genetics* 21(suppl): 1-60.

Newton, M. A., C. M. Kendziorski, et al. (2001). On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. J Comput Biol 8(1): 37-52.

Nilsson O, Weigel D (1997) Modulating the timing of flowering. Curr Opin Biotechnol8: 195-199.

Noctor G, and Foyer C (1998) Annu. Rev. Plant Physiol. Mol. Biol. 49: 249-279.

Normile D (1997) Yangtze seen as earliest rice site. Science275:309

Ogata, H., S. Goto, et al. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 27(1): 29-34.

Okubo, K., N. Hori, et al. (1992). Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nat Genet 2(3): 173-9.

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

Pan, W., J. Lin, et al. (2002). How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach. Genome Biol 3(5): research0022.

Pandey A, Mann M. Proteomics to study genes and genomes. Nature 2000; 405:837–46.
Park JM, Park CJ, Lee SB, Ham BK, Shin R, and Paek KH (2001) Overexpression of the tobacco Tsi1 gene encoding an EREBP/AP2-type transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. Plant Cell 13: 1035-1046.
Patil, N., A. J. Berno, et al. (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294(5547): 1719-23.
Patrick J. Krysan, Jeffery C. Young,, and Michael R. Sussmana. T-DNA as an Insertional Mutagen in Arabidopsis .Plant Cell, Vol. 11, 2283-2290,
Pease, A. C., D. Solas, et al. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A 91(11): 5022-6.
Peng S, Laza RC, Visperas RM, Sanico AL, Cassman KG, Khush GS (2000) Grain yield of rice cultivars and lines developed in the Philippines since 1966. Crop Sci 40:307–314
Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST databases. Genome Res 9:167-174.
PNAS 102:6373-6378
Price et al. (2002) Mapping QTLs associated with drought avoidance in upland rice grown in the Philippines and West Africa. Plant Mol Biol; 48: 683−695.
Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A., Baumeister, K., 1987. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science 238, 336–341.
Quackenbush, J. (2001). Computational analysis of microarray data. Nat Rev Genet 2(6): 418-27.
Quackenbush, J. (2002). Microarray data normalization and transformation. Nat Genet 32 Suppl: 496-501.
R, Mackerness S, Hancock JT, Neill SJ (2001) Regulation of the Arabidopsis transcriptome by oxidative stress. Plant Physiol 127:159–172.
Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K. et al., 2003. Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiology 133, 1755-1767.
Raghuraman, M. K., E. A. Winzeler, et al. (2001). Replication dynamics of the yeast genome. Science 294(5540): 115-21.
Randolph, J. B. and A. S. Waggoner (1997). Stability, specificity and fluorescence brightness of multiply-labeled fluorescent DNA probes. Nucleic Acids Res 25(14): 2923-9.
Raychaudhuri, S., J. M. Stuart, et al. (2000). Principal components analysis to summarize microarray experiments: application to spoulation time series. Pac. Symp. Biocomput. 2000: 455-466.
Reddy AR, Ramakrishna W, Sekhar AC, Ithal N, Babu PR, Bonaldo MF, Soares MB, Bennetzen JL. 2002a. Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. indica cv Nagina 22). Genome 45, 204–211.
Reddy et al., 2001 Novel EST enrichment with normalized cDNA libraries from drought-stressed rice (*Oryza sativa* L. cv Nagina 22),. www.ncbi.nlm.nih.gov/dbest.
Reese,MG, Kulp,D, Tammana,H, Haussler,D (2000) Genie--Gene Finding in Drosophila Melanog in Genome Research 10(4) 529-538
Reeve, M.A., Fuller, C.W., 1995. A novel thermostable polymerase for DNA sequencing. Nature 376, 796–797.
Reinartz, J., E. Bruyns, et al. (2002). Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. Brief Funct Genomic Proteomic 1(1): 95-104.
Reymond, P., Weber, H., Damond, M. and Farmer, E.E. (2000) Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. Plant Cell, 12, 707–719.
Richmond, T.,Somerville, S.C. and Manners, J.M. (2000) Coordinated plant defense responses in Arabidopsis revealed by microarray analysis.Proc. Natl. Acad. Sci. USA, 97, 11655–11660.

Roth, F. P., J. D. Hughes, et al. (1998). Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat Biotechnol 16(10): 939-45.

Roxas VP, Smith RK Jr, Allen ER, and Allen RD (1997) Overexpression of glutathione-S-transferase/glutathione peroxidase enhances the growth of transgenic tobacco seedlings during stress. Nature Biotechnol. 15: 988-991.

Ruan, Y., Gilmore, J., and Conner, T. (1998). Towards Arabidopsis genome analysis: Monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J. 15, 821–833.Ruan et al., 1998

Rudd Stephen (2003) Expressed sequence tags:alternative or complement to whole genome sequences. TRENDS in Plant Science Vol.8 No.7 July 2003 Rudd., 2003

Sablowski RW, Meyerowitz EM: A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA. Cell 1998, 92:93-103.

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., Petersen, G.B., 1982. Nucleotide sequence of bacteriophage lambda DNA. J. Mol. Biol. 162, 729–773.

Sanja Rogic, Alan K. Mackworth, and Francis B.F. Ouellette. (2001) Evaluation of Gene-Finding Programs on Mammalian Sequences. Genome Research 11: 817-832.

Sasaki Takuji .Where does the accurate rice genome sequence lead us? Plant Molecular Biology (2005) 59:27–32. DOI 10.1007/s11103-005-1224-4. Sasaki., 2005

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

Sauter Margret. Rice in deep water: How to take heed against a sea of troubles. Naturwissenschaften (2000) 87:289–303

Schaffer R., Landgraf J., Perez-Amador M., Wisman E., Monitoring genome-wide expression in plants, Curr. Opin. Biotechnol. 11 (2000) 162–167.

Schena M. Genome analysis with gene expression microarrays. Bioassays 1995;18:427–31.

Schena M., Shalon D., Davis R.W., Brown P.O.,Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science 270 (1995) 467–470.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467–470.

Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM (2000) Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. Proc Natl Acad Sci USA 97:11655–11660.

Schroeder, B. G., L. M. Peterson, et al. (2002). "Improved quantitation and reproducibility in lycobacterium tuberculosis DNA microarrays." J Mol Microbiol Biotechnol 4(2): 123-6.

Schuchardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrach H, Herzel H (2000) Normalization strategies for cDNA microarrays. Nucleic Acids Res 28: E47

Schultz et al, 2000 Schultz, J, Doerks T, Ponting CP, Copley RR, Bork P (2000) More than1,000 Putative New Human Signaling Proteins Revealed by EST Data Mining, Nature Genetics 25(2), 201-204.

Schulze, A. and J. Downward (2001). Navigating gene expression using microarrays--a technology review. Nat Cell Biol 3(8): E190-5.

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

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

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T et al. 2002a. Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. The Plant Journal 31, 279–292.

Shalon, Smith et al. 1996 Shalon, D., S. J. Smith, et al. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 6(7): 639-45.

Sheveleva E, Chmara W, Bohnert HJ, and Jensen RG (1996) Increased salt and drought tolerance by D-ononitol production in transgenic Nicotiana tabacum L. Plant Physiol. 115: 1211-1219.

Shimamoto, K. and Kyozuka, J. 2002. Rice as a model for comparative genomics of plants. Annu. Rev. Plant Biol. 53:99–419.

Shinozaki and Yamaguchi-Shinozaki, 1997 Shinozaki, K. and Yamaguchi-Shinozaki, K. (1996). Molecular responses to drought and cold stress. Curr. Opin. Biotech. 7: 161-167.

Shinozaki and Yamaguchi-Shinozaki, 2000 Shinozaki K, and Yamaguchi Shinozaki Y (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Plant Biol. 3: 217-223.

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

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

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

Shoemaker, D. D. and P. S. Linsley (2002). "Recent developments in DNA microarrays." Curr Opin Microbiol 5(3): 334-7.59

Shoemaker, D. D., E. E. Schadt, et al. (2001). "Experimental annotation of the human genome using microarray technology." Nature 409(6822): 922-7.

Silva, A. P., J. E. De Souza, et al. (2004). The impact of SNPs on the interpretation of SAGE and MPSS experimental data. Nucleic Acids Res 32(20): 6104-10.

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

Smith et al., 2001 Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O. et al.(2001) The sequence of the human genome. Science 291, 1304-1351.

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

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B., Hood, L.E.,1986. Fluorescence detection in automated DNA sequence analysis. Nature 321, 674–679.

Smyth, G. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3(1): Article 3.

Smyth, G. K. and T. Speed (2003). Normalization of cDNA microarray data. Methods 31(4): 265-73.

Soares et al., 1994 Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L., Efstratiadis, A., 1994. Construction and characterization of a normalized cDNA library. Proc. Natl. Acad. Sci. USA 91, 9228–9232.

Southern, E M, Case-Green, S C., J K Elder, M Johnson, K U Mir, L Wang, and J C Williams. Arrays of complementary oligonucleotides for analysing the hybridisation behaviour of nucleic acids. Nucleic Acids Research, 22(8):1368{73, April 1994.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Spellman, P. T., G. Sherlock, et al. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell 9(12): 3273-97.

Staden 1994 Staden,R. (1994) Methods Mol Biol. 22, 9–170.

Stears, R. L., R. C. Getts, et al. (2000). "A novel, sensitive detection system for high-density microarrays using dendrimer technology." Physiol Genomics 3(2): 93-9.

Sterky et al., 1998 Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg, A., Amini, B., Bhalerao, R., Larsson, M.,Villarroel, R., Van Montagu, M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W., Gustafsson, P., Uhlen, M.,Sundberg, B., Lundeberg, J., 1998. Gene discovery in the wood-forming tissues of poplar: analysis of 5692 expressed sequence tags. Proc. Natl. Acad. Sci. USA 95, 13330–13335.

Sterrenburg, E., R. Turk, et al. (2002). A common reference for cDNA microarray hybridizations. Nucleic Acids Res 30(21): e116.

Sun, N., Ma, L., Pan, D., Zhao, H.Y. and Deng, X.W. 2003. Evaluation of light regulatory potential of Calvin cycle steps based on large-scale gene expression profiling data. Plant Mol. Biol. 53: 467–478.

Tabor, S., Richardson, C.C., 1989. Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis. J. Biol. Chem. 264,

Takahashi, S., Katagiri, T., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2000). An Arabidopsis gene encoding a Ca21-binding protein is induced by abscisic acid during dehydration. Plant Cell Physiol. 41, 898–903.

Talam`e V, Ozturk N, Bohnert HJ, et al. 2003. Microarray analysis of transcript abundance in barley under conditions of water deficit. Plant and Animal Genome Conference XI,11−15 January 2003, San Diego, CA, p. 15.

Tamayo, P. et al. (1999). \Interpreting patterns of gene expression with selforganising maps: Methods and applications to hematopoietic di_erentiation" PNAS. 96: 2907{2912.

Tavazoie, S. et al. (1999). \Systematic determination of genetic network architecture"Nature genetics. 22: 281-285.

Tepperman JM, and Dunsmuir P (1990) Transformed plants with elevated levels of chloroplast SOD are not more resistant to superoxide toxicity. Plant Mol. Biol. 14: 501-511.

The Rice Chromosome 10 Sequencing Consortium. 2003. In-depth view of structure, activity, and evolution of rice chromosome 10. Science 300: 1566–1569.

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

Thimm O, Essigmann B, Kloska S, Altmann T, Buckhout TJ (2001) Response of Arabidopsis to iron deficiency stress as revealed by microarray analysis. Plant Physiol 127: 1030–1043

Thomas JC, Sepahi M, Arendall B, and Bohnert HJ (1995) Bohnert et al., 2003

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

Thomashow MF. Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Physiol Plant Mol Biol 1999;50:571–99.

Toronen, P., M. Kolehmainen, et al. (1999). Analysis of gene expression data using self-organizing maps. FEBS Lett 451(2): 142-6.

Tran, P. H., D. A. Peiffer, et al. (2002). Microarray optimizations: increasing spot accuracy and automated identification of true microarray signals. Nucleic Acids Res 30(12): e54.

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

Tseng, G. C., M. K. Oh, et al. (2001). Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Nucleic Acids Res 29(12): 2549-57.

Tzeng TY, Chen HY, Yang CH: Ectopic expression of carpel-specific MADS box genes from lily and lisianthus causes similar homeotic conversion of sepal and petal in Arabidopsis. Plant Physiol 2002, 130:1827-1836.

U.S. Bureau of Census. 1998. International Database, Electronic Database, Suitland M.D. Updated 15 June 1998.

Uberbacher,E.C. and Mural,R.J. (1991) Locating protein–coding regions in human DNA sequences by a multiple sensor–neural network approach. Proc. Natl. Acad. Sci., USA 88, 11261–11265

Uchimiya H, Kidou S, Shimazaki T, Takamatsu S, Hashimoto H, Nishi R, Aotsuka S, Matsubayashi Y, Kidou N, Umeda M, Kato A. 1992. Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.).The Plant Journal 2, 1005–1009.

Umeda M, and Uchimiya H (1994) Differential transcript levels of genes associated with glycolysis and alcohol fermentation in rice plants (*Oryza sativa* L.) under submergence stress. Plant Physiol. 106: 1015-1022.

Urao T, Yamaguchi-Shinozaki K, Urao S, and Shinozaki K (1993) Plant Cell 5:1529-1539.

Van Camp W, Capiau K, Van Montagu M, Inze D, and Slooten L (1996) Enhancement of oxidative stress tolerance in transgenic tobacco plants overproducing Fe-superoxide dismutase in chloroplasts. Plant Physiol. 112: 1703-1714.
van Steensel, B. and S. Henikoff (2003). Epigenomic profiling using microarrays. Biotechniques 35(2): 346-50, 352-4, 356-7.
Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE Jr, Hieter P, Vogelstein B, Kinzler KW (1997) Characterization of the yeast transcriptome. Cell 88: 243–251 cells. Science 276, 1268–1272
Velculescu, V. E., L. Zhang, et al. (1995). Serial analysis of gene expression. Science 270(5235): 484-7.Zhang et al., 1996
Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. Science 270,
Veylder L, Beeckman T, Beemster GT, de Almeida Engler J, Ormenese S, Maes S, Naudts M, Van Der Schueren E, Jacqmard A, Engler G et al.: Control of proliferation, endoreduplication and differentiation by the Arabidopsis E2Fa-DPa transcription factor. EMBO J 2002, 21:1360-1368.
Vierling E, Kimpel JA (1992) Plant responses to environmental stress. Curr Opin Biotechnol 3: 164-170
Vision TJ, Brown DG, Tanksley SD: The origins of genomic duplications in Arabidopsis. Science 2000, 290:2114-2117.
Vroemen, C.W. et al. (2003) The CUP-SHAPED COTYLEDON3 gene is required for boundary and shoot meristem formation in Arabidopsis. Plant Cell 15, 1563–1577.
Wang X and B Seed. Selection of oligonucleotide probes for protein coding sequences. Bioinformatics,19(7):796 802, May 2003.
Wang, E., L. D. Miller, et al. (2000). High-fidelity mRNA amplification for gene profiling. Nat Biotechnol 18(4): 457-9.
Wang, X., S. Ghosh, et al. (2001). Quantitative quality control in microarray image processing and data acquisition. Nucleic Acids Res 29(15): E75-5.Wang et al, 1998
Weeraratna, A. T., J. E. Nagel, et al. (2004). Gene expression profiling: from microarrays to medicine. J Clin Immunol 24(3): 213-24.
Weig A, Deswarte C, Chrispeels MJ.1997. The major intrinsic protein family of Arabidopsis has 23 members that form three distinct groups with functional aquaporins in each group. Plant Physiology 114, 1347-57.Wildsmith and Archer et al., 2001
Wildsmith, S. E., G. E. Archer, et al. (2001). Maximization of signal derived from cDNA microarrays. Biotechniques 30(1): 202-6, 208.
Wodicka, L., Dong, H., Mittmann, M., Ho, M.H., and Lockhart, D.J. (1997). Genome-wide expression monitoring in Saccharomyces cerevisiae. Nat. Biotechnol. 15, 1359–1367.Wolfinger and Gibson et al., 2001
Wolfinger, R. D., G. Gibson, et al. (2001). Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 8(6): 625-37.
Wolford JK, Blunt D, Ballecer C, Prochazka M (2000) High Throughput SNP detection by DNA pooling and denaturing high performance liquid chromatography (DHPLC) Human Genetics 107, 483-487.
Wu K, Tian L, Hollingworth J, Brown DC, Miki B: Functional analysis of tomato Pti4 in Arabidopsis. Plant Physiol 2002, 128:30-37.
Xie, Jeong et al. 2004. Xie, Y., K. S. Jeong, et al. (2004). A case study on choosing normalization methods and test statistics for two-channel microarray data. Comparative and Functional Genomics 5: 432-444.
Xiong L, Schumaker K, and Zhu JK (2002) Plant Cell 14: S165-S183.
Yamaguchi-Shinozaki K, and Shinozaki K (1993) Mol. Gen. Genet. 238: 17.
Yamaguchi-Shinozaki K, and Shinozaki K (1994) Plant Cell 6: 251-264.
Yamamoto K, Sasaki T. 1997. Large scale EST sequencing in rice. Plant Molecular Biology 35, 135–144.
Yamamoto, M., T. Wakatsuki, et al. (2001). Use of serial analysis of gene expression (SAGE) technology. J Immunol Methods 250(1-2): 45-66.

Yang, Ruan et al. 2001 Yang, M. C., Q. G. Ruan, et al. (2001). A statistical method for flagging weak spots improves normalization and ratio estimates in microarrays. Physiol Genomics 7(1): 45-53.

Yang, Y. H., M. J. Buckley, et al. (2002). Comparison of methods for image analysis on cDNA microarray data. Journal of Computational and Graphical Statistics 11: 108-136.

Yang, Y. H., S. Dudoit, et al. (2001). Normalization for cDNA microarray data. Microarrays: Optical Technologies and Informatics. M. L. Bittner, Y. Chen, A. N. Dorsel and E. R. Dougherty. Volume 4266 of Proceedings of SPIE.

Yang, Y. H., S. Dudoit, et al. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30(4): e15.

Yang, Y.H. and Speed, T. (2002) Design issues for cDNA microarray experiments. Nat. Rev. Genet. 3, 579–588

Yano et al., xxxx Takahashi Y, Shomura A, Sasaki T, Yano M (2001). Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the subunit of protein kinase CK2. Proc Natl Acad Sci USA, 98:7922-7927Yazaki et al., 2002, 2004a, b

Yazaki, J., Kishimoto, N., Ishikawa, M. and Kikuchi, S. 2002. Rice expression database: the gateway for rice functional genomics. Trends Plant Sci. 7: 563–564.Yeakley and Fan et al., 2002

Yazaki, J., Kojima, K., Suzuki, K., Kishimoto, N. and Kikuchi, S. 2004b. The rice pipeline: a unification tool for plant functional genomics. Nucl. Acids Res. 32: D383–D387.

Yazaki, J., Shimatani, Z., Hishimoto, A., Nagata, Y., Fujii, F.et al. 2004a. Transcriptional profiling of genes responsive to abscisic acid and gibberellin in rice: phenotyping and comparative analysis between rice and Arabidopsis. Physiol.Genom. 17(2): 87–100.

Yeakley, Fan et al. 2002 Yeakley, J. M., J. B. Fan, et al. (2002). Profiling alternative splicing on fiber-optic arrays. Nat Biotechnol 20(4): 353-8.

Yu and Setter 2003 Yu LX, Setter TL (2003) Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water deficit. Plant Physiol 131:568–582.

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

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

Yue, H., P. S. Eastman, et al. (2001). "An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression." Nucleic Acids Res 29(8): E41-1.

Yunbi Xu, Susan R. McCouch and Qifa Zhang. How can we use genomics to improve cereals with rice as a reference genome?. Plant Molecular Biology (2005) 59:7–26 DOI 10.1007/s11103-004-4681-2.

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

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

Zhang K, Qin ZS, Liu JS, Chen T, Waterman MS, Sun F, Haplotype block partitioning and tag SNP selection using genotype data and their applications to association studies., Genome Res. 2004 May;14(5):908-16.

Zhang X, Zhang L, Dong F, Gao J, Galbraith DW, and Song CP (2001) Plant Physiol. 126: 1438-1448.

Zhang, Y., B. D. Price, et al. (2001). "Reproducible and inexpensive probe preparation for oligonucleotide arrays." Nucleic Acids Res 29(13): E66-6.

Zhang, Y., Zhao, X., Zhang, X., Zhang, Y., Han, B., Cheng, J. and Xue, Y. 2004. Monitoring of gene expression profiles and isolation of candidate genes involved in pollination and fertilization in rice (*Oryza sativa* L.) with a 10K cDNA microarray. Plant Mol. Biol. 54: 471–487.

Zhao W, Wang J, He X, Huang X, Jiao Y, et al. 2004. BGI-RIS: An integrated information resource and comparative analysis workbench for rice genomics. Nucleic Acids Research 32, 377–382.
Zhao W, Wang J, He X, Huang X, Jiao Y, et al. 2004. BGI-RIS: An integrated information resource and comparative analysis workbench for rice genomics. Nucleic Acids Research 32, 377–382.
Zhu B, Su J, Chang MC, Verma DPS, Fan YL, and Wu R (1998) Overexpression of a pyrroline-5-carboxylate synthetase gene and analysis of tolerance to water and salt stress in transgenic rice. Plant Sci. 139: 41-48.

Web sites referred

http://apps1.niaid.nih.gov/david
http://bozeman.mbt.washington.edu/phrap.docs/phrap.html
http://cmgm.stanford.edu/pbrown/protocols/index.html
http://discover.nci.nih.gov/gominer/
http://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules
http://rgp.dna.affrc.go.jp/IRGSP/Build3/build3.html
http://stress-genomics.org/stress.fls/expression/expression.html
http://www.arrayit.com
http://www.corning.com/lifesciences/pdf/gaps_ii_coated_slides_10_01_ss_cmt_gaps
http://www.crc.dk/dnatools
http://www.ebi.ac.uk/InterProScan/
http://www.genmapp.org
http://www.grammene.org.
http://www.irri.org
http://www.maths.lth.se/bioinformatics/software/
http://www.ncbi.nlm.nih.gov/dbest
http://www.ncbi.nlm.nih.gov/geo/
http://www.ncbi.nlm.nih.gov/geo/
http://www.ncbi.nlm.nih.gov/projects/geo/
http://www.ncbi.nlm.nih.gov/SAGE/
http://www.phrap.org
http://www.probes.com
http://www.R-project.org
http://www.sigmaaldrich.com

8. LIST OF PUBLICATIONS

Publications:

Markandeya Gorantla, P. R. Babu, V. B. Reddy Lachagari, A. M. M. Reddy, Ramakrishna Wusirika, Jeffrey L. Bennetzen, Arjula R. Reddy., Identification of stress responsive genes in an indica rice (*Oryza sativa* L.) using ESTs generated from drought stressed seedlings, Communicated, Journal of Experimental Botany

Markandeya G, V.B.Reddy., Alex Feltus., A.H.Paterson., A.R.Reddy. 2005 Functional genomics of drought stress response in rice: transcript mapping of annotated unigenes of indica rice (*Oryza sativa* L. cv. Nagina 22) Vol 89, No 3 Aug 10, 2005, current science.

Babu, P.R.,. Sekhar, A.C., Ithal, N., **Markandeya G**., Reddy, A. R, (2002) Annotation and BAC/PAC localization of non-redundant EST's from drought stressed seedlings of an indica rice J. Genet 81, 25-44

Rabi M, **Markandeya G**, A.R.Reddy, M. K. Reddy, Isolation and Characterization of Expressed Sequence Tag-Based Gene Expression analysis in response to salinity, osmotic and cold stress in Pennisetum seedling using subtractive cDNA library construction, communicated.

GenBank Submissions:

Markandeya,G., Reddy,A.R., Ramakrishna,W., Nagabhushana,I.,Ravindra Babu,P., Madana Mohan Reddy,A., Chandra Sekhar,A.,Bennetzen,J.L. (2002) Novel EST enrichment with normalized cDNA libraries from drought stressed rice (*Oryza sativa* L.cv Nagina 22) Gene Bank Accession: BU672765 to BU673915 www.ncbi.nlm.nih.gov/dbest**.**

Markandeya, G.; Ravindra Babu,P.; Venkat Reddy, B.; Nagabhushana Ithal.; Chandra Sekhar, A.; Bennetzen,J.L.; Ramakrishna,W.; Reddy, A.R. (2003) ESTs from a normalized cDNA library of drought stressed rice seedlings (*Oryza sativa* L.cv Nagina 22) Gene Bank Accession: CB964418 to CB967504 www.ncbi.nlm.nih.gov/dbest**.**

Ravindra Babu,P., **Markandeya,G**., Ramakrishna,W., Nagabhushana,I., Chandra Sekhar,A., Madana Mohan Reddy,A., Bennetzen,J.L. and Reddy,A.R. (2002). Isolation and characterization of mRNA of a DRE-binding protein 1B (DREB 1B) of indica rice. Gene Bank Accession: AY166833 www.ncbi.nlm.nih.gov.

Rabi M, **Markandeya G**, A.R.Reddy, M.K.Reddy et al Analysis of Expressed Sequence Tags from a subtracted cDNA library prepared from Pennisetum glaucum seedlings that were exposed to drought stress 2003. Gene Bank Accession: CD724312 to CD726805. www.ncbi.nlm.nih.gov/dbEST

Ravindra Babu, P., **Markandeya, G.,** A.R.Reddy., Isolation and characterization of CRT/DRE - binding protein of indica rice. (2004). Gene Bank Accession: AY502052 www.ncbi.nlm.nih.gov

International conferences:

Markandeya G., A.R. Reddy et al., Expression Profiling of Drought Responsive Genes In An Indica Rice Using cDNA Microarrays, PAGE – XIV, San Diego CA, USA 2006.

Markandeya G., A.R.Reddy et al., Identification of Stress Responsive Genes in Rice Using ESTs Generated from Drought Stressed Seedlings, PAGE – XII, San Diego CA, USA 2004.

Markandeya Gorantla., Rice and Pearl Millet EST Micorarrays for Comparative Genomic Analysis for candidate genes involved in drought stress through Expression Profiling, TIGR NSF Rice Oligonucleotide array workshop, Rockville MD, USA 2004.

A.R.Reddy., et al **Markandeya G**., 2nd International symposium on Functional Genomics of Rice Tucson, Arizona USA. 2004.

R. Reddy et al**, Markandeya G**., Development of functional markers for drought tolerance in rice: Identification and validation of candidate genes and SNPs, Resilient Crops for Water Limited Environments Resilient crops for water limited environments, Cuernavaca Mexico, May 2004

Arjula R. Reddy, et al, **Markandeya G**, Functional genomics of drought stress response in rice: gene discovery, marker development and mapping, and SNP identification at candidate gene loci. Third workshop on drought tolerance in rice, National Network Program, 22nd – 25th, July 2003.

Babu PR.,Markandeya G., A. R. Reddy et al., putative functions, BAC clone identification, and expression analysis of ESTs generated from cDNA libraries of drought-stressed seedlings of indica rice (*Oryza sativa* L.) International Rice Congress, Beijing, China 2002.

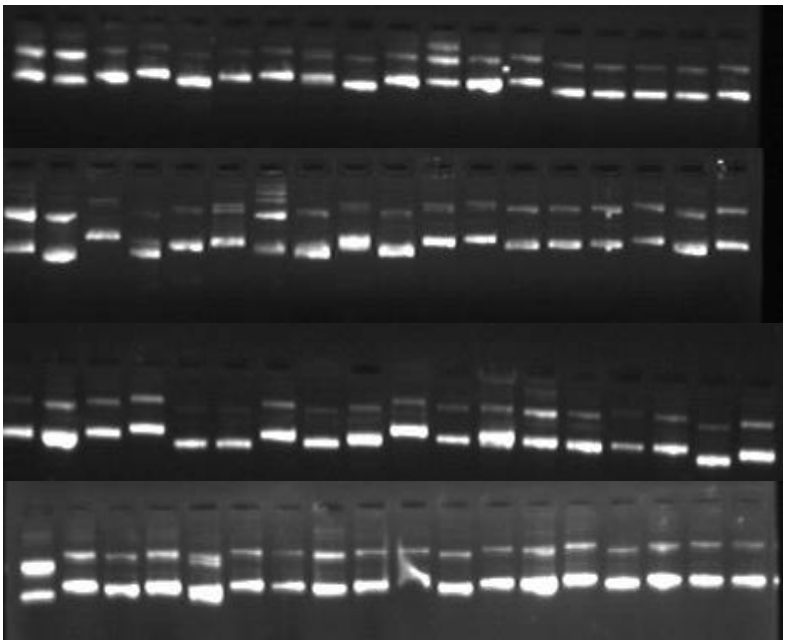

Fig 1**:** Quality and Uniform concentration of recombinant plasmid preps used for EST generation

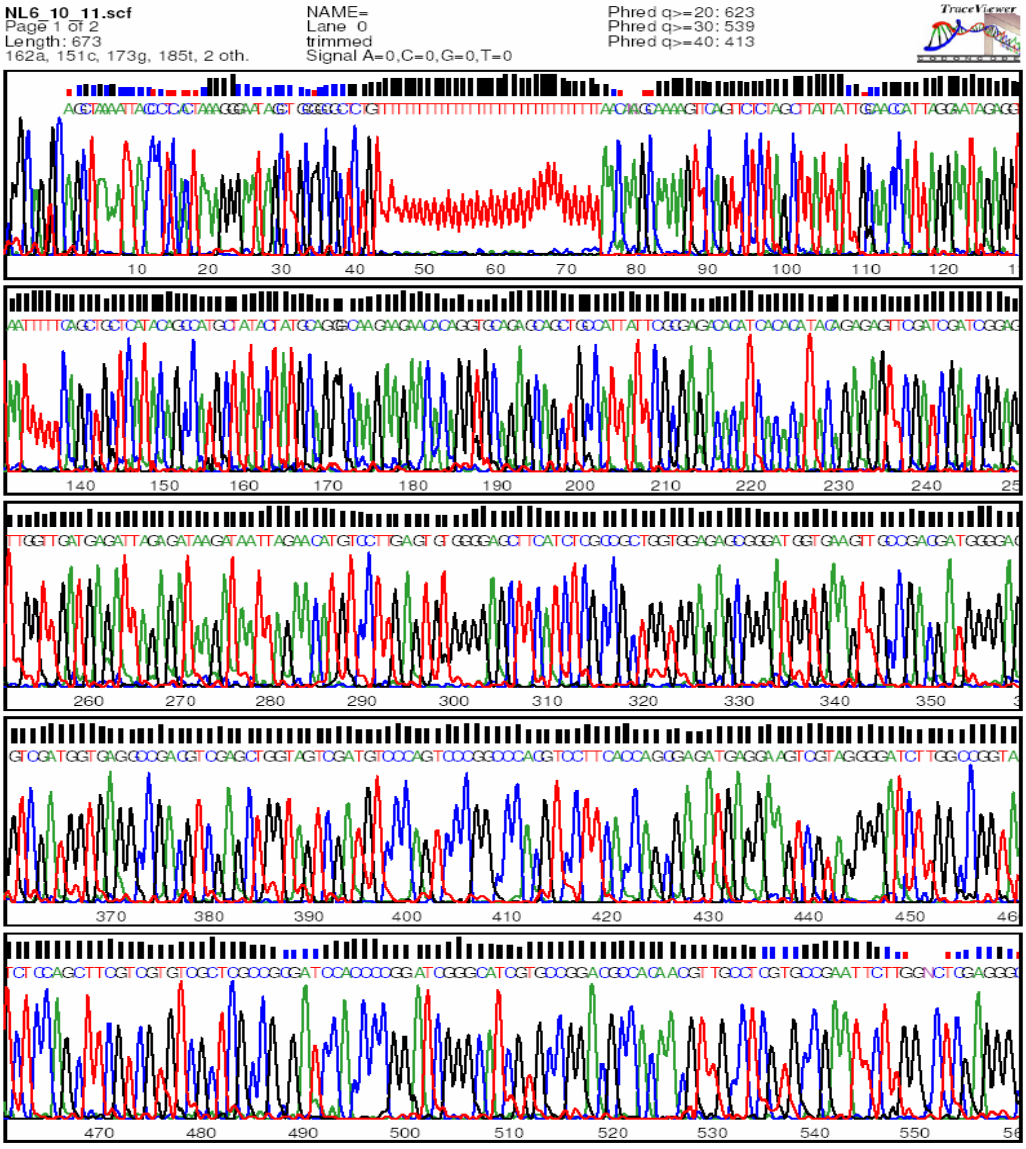

Fig 2a: Sequence chromatogram obtained through PHRED base calling from raw data obtained from MegaBace

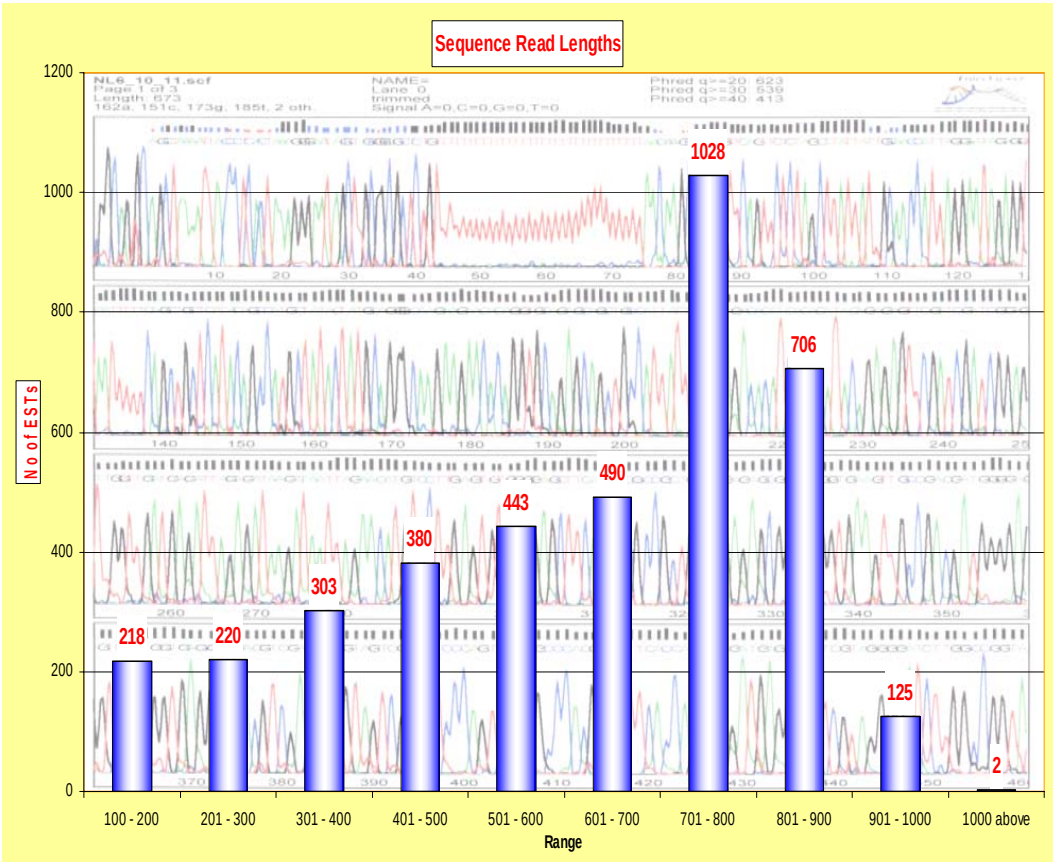

Fig 2b: Sequence read lengths obtained on single pass sequencing of 4000 clones from *O.sativa cv* Nagina 22 library

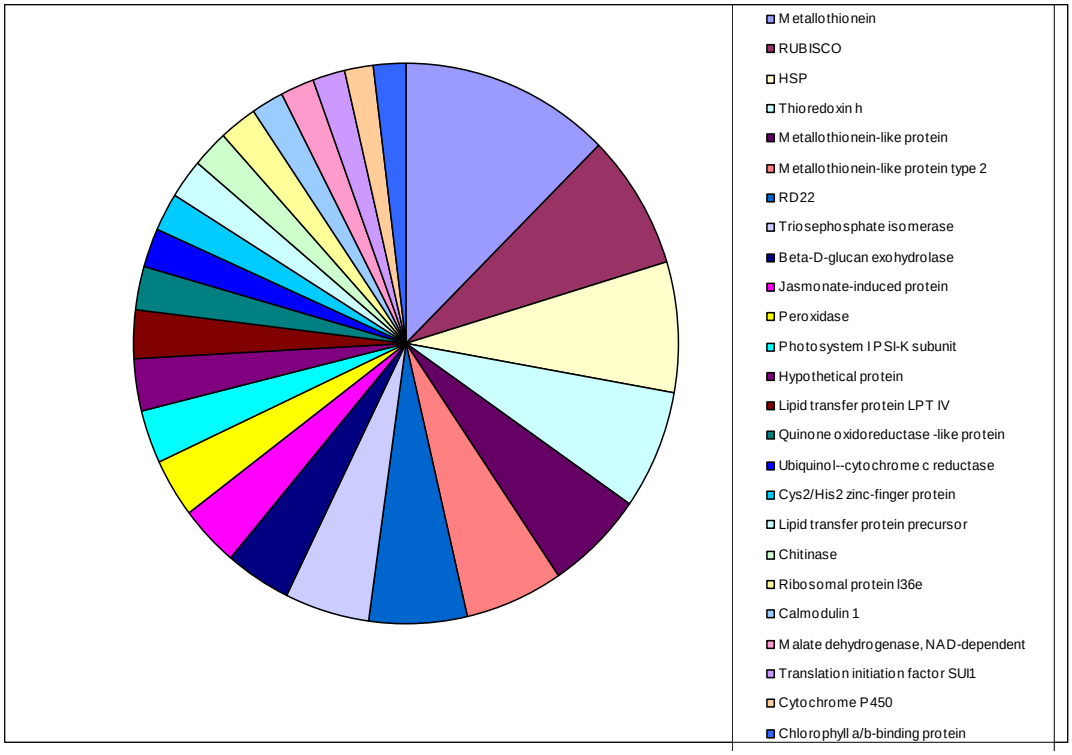

Fig 3: Transcript abundance of highly expressed genes from *O.sativa* cv Nagina 22 library

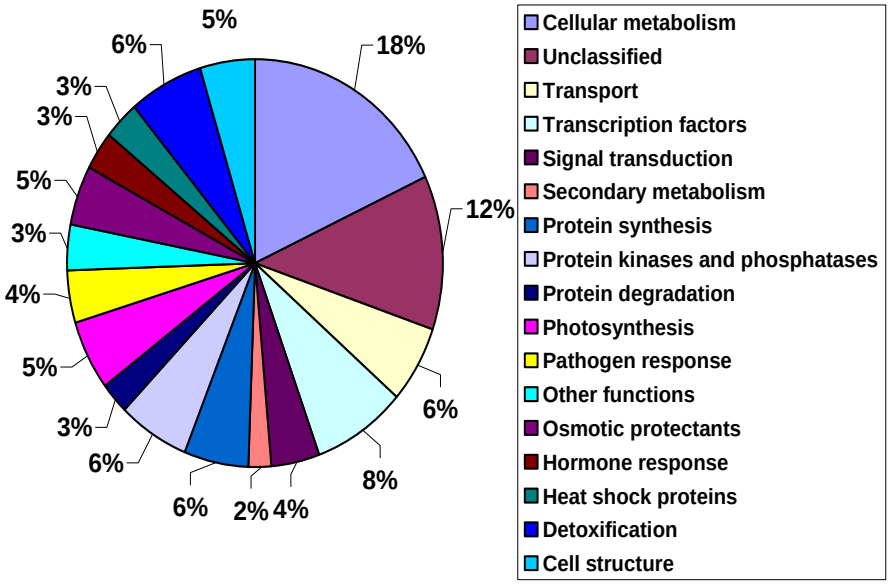

Fig 4: Functional classification of 589 putative stress responsive genes from *O.sativa* cv Nagina 22

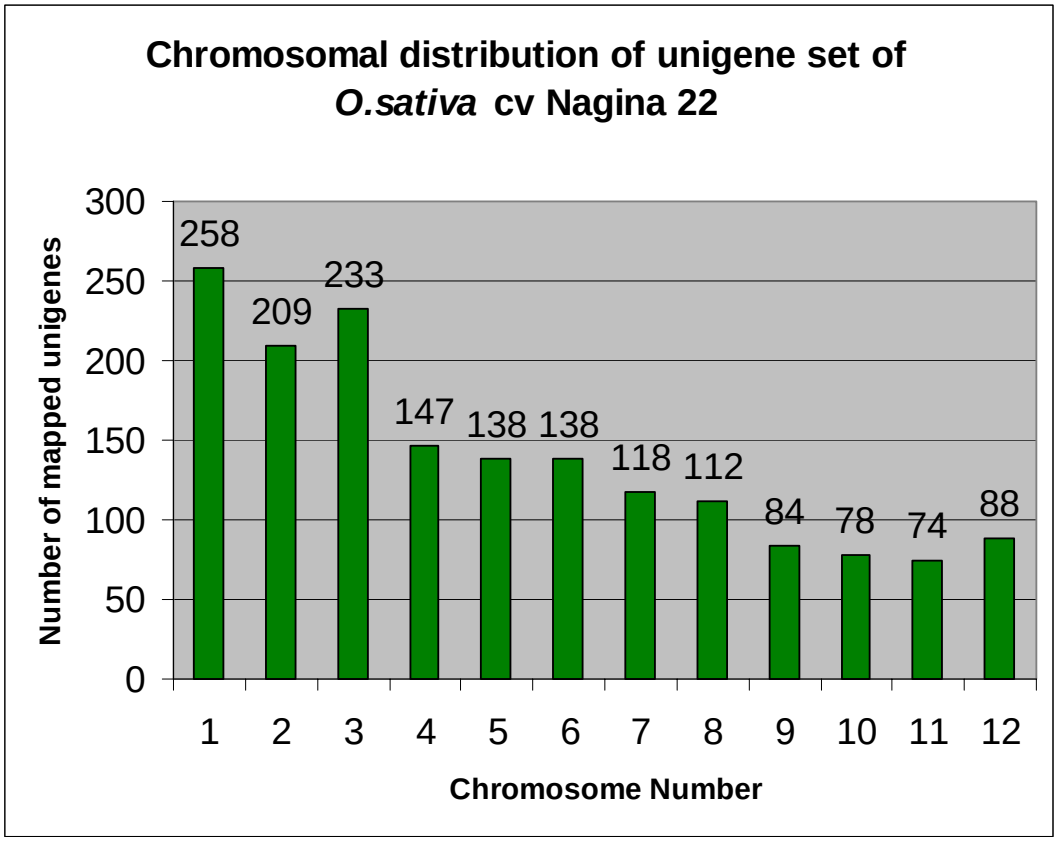

Fig 5: Distribution of N22 unigene set on rice genome

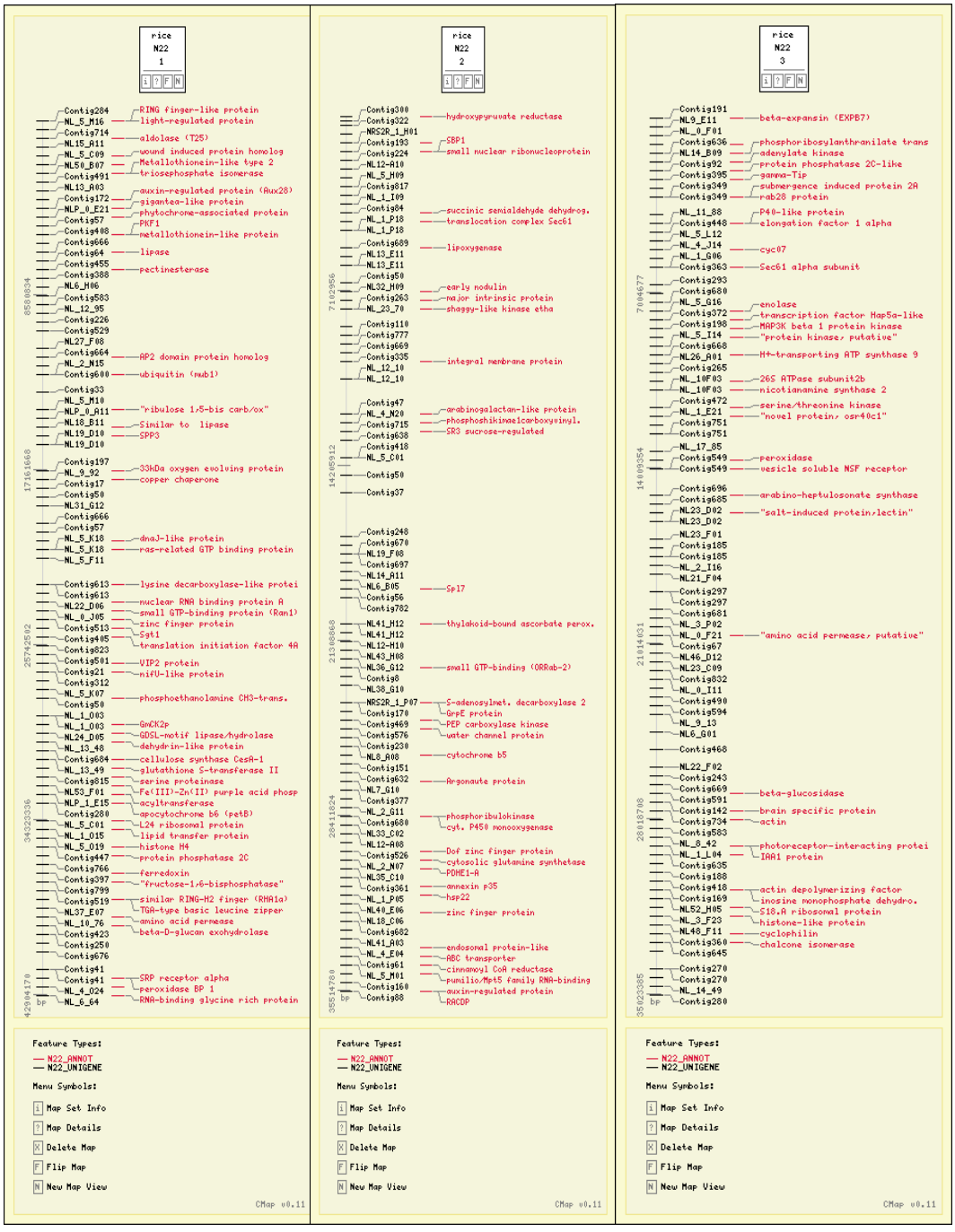

Fig 6: Physical maps of N22 unigene set with functional annotations of SRGs mapped on chromosome 1, 2 and 3

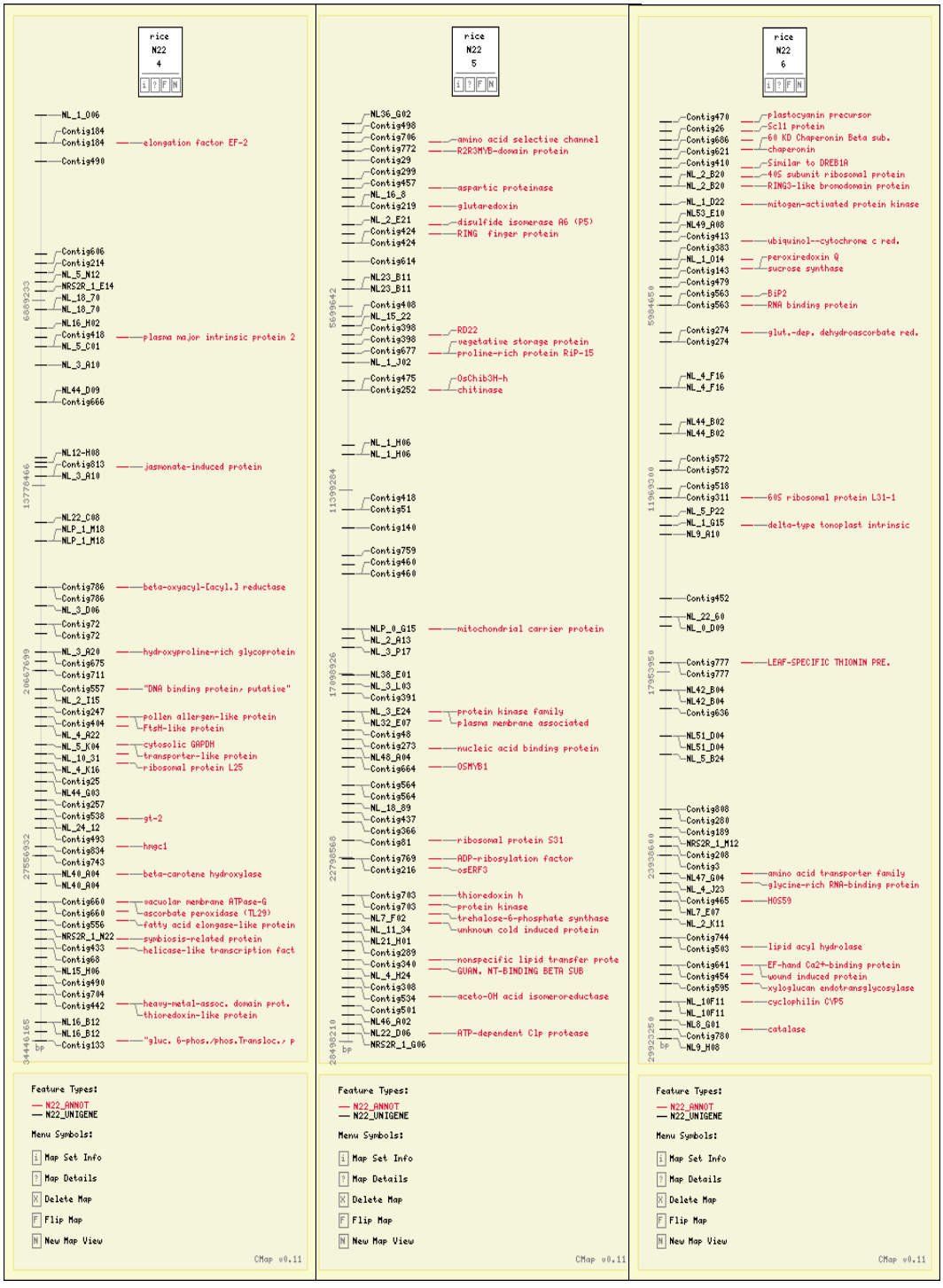

Fig 7 : Physical maps of N22 unigene set with functional annotations of SRGs mapped on chromosome 4, 5 and 6

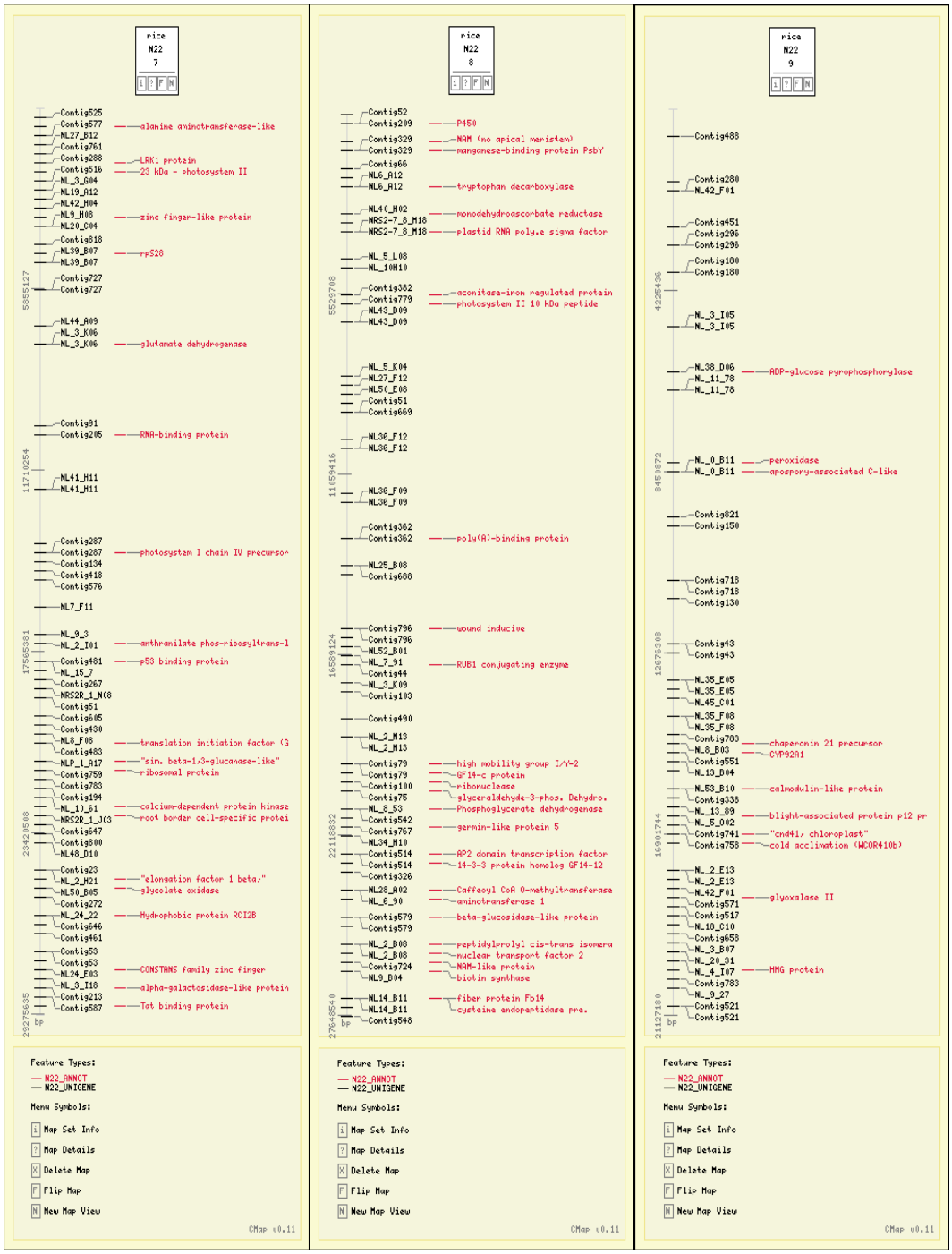

Fig 8: Physical maps of N22 unigene set with functional annotations of SRGs mapped on chromosome 7, 8 and 9

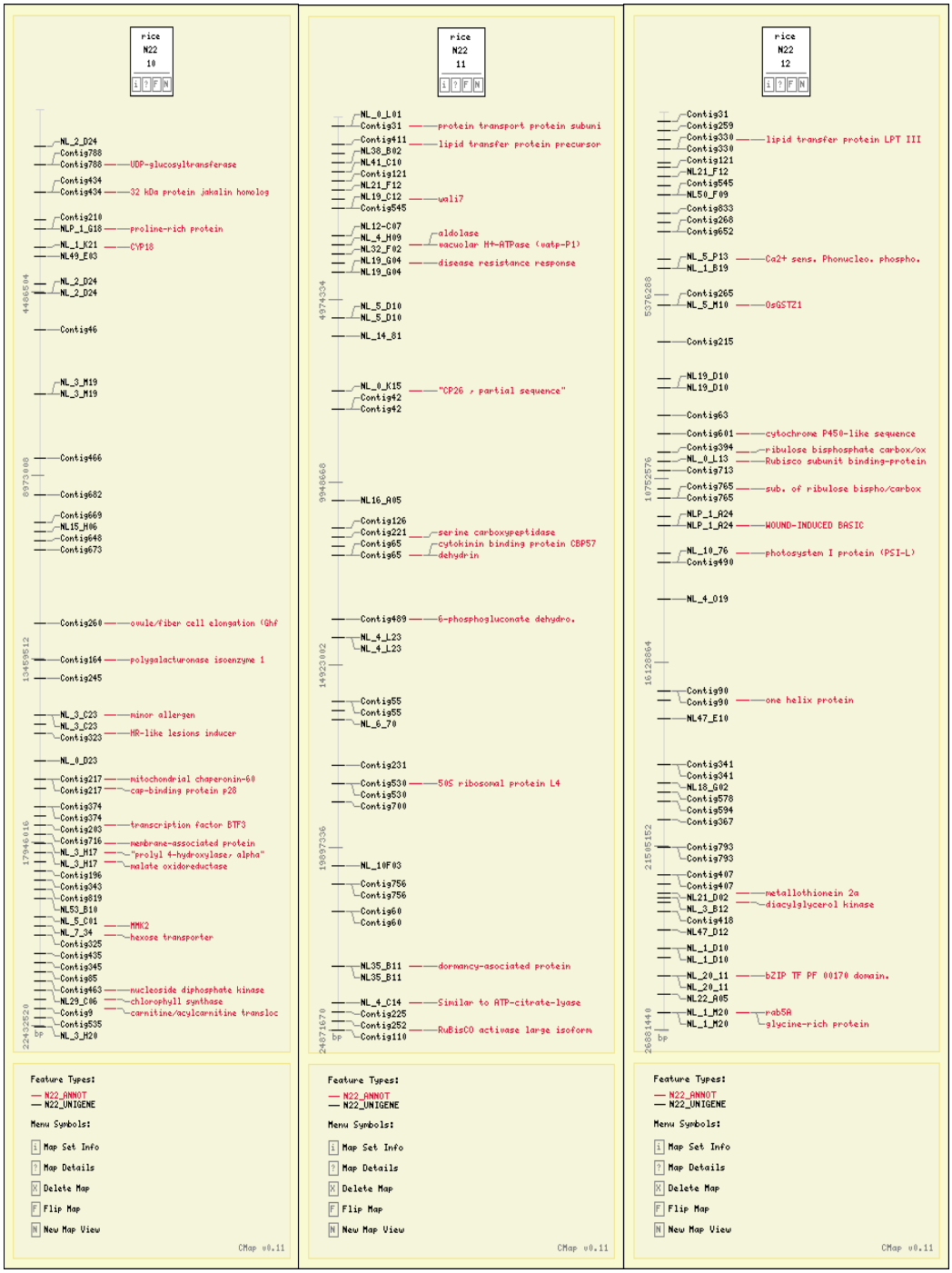

Fig 9: Physical maps of N22 unigene set with functional annotations of SRGs mapped on chromosome 10 and 11 & 12

Fig 10a: Rainout shelter (parking position) for field drought stress experiment showing 8 different plots used for inducing different levels of water stress

Fig 10b: Digital irrigation controls used for regulating water supply to different plots

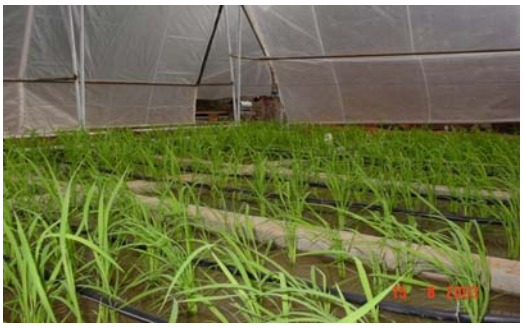

Fig 11: Biological replicates with in each stress regime which are pooled for sampling

Fig 12a: Complete experiment setup before drought stress initiation (Day 38)

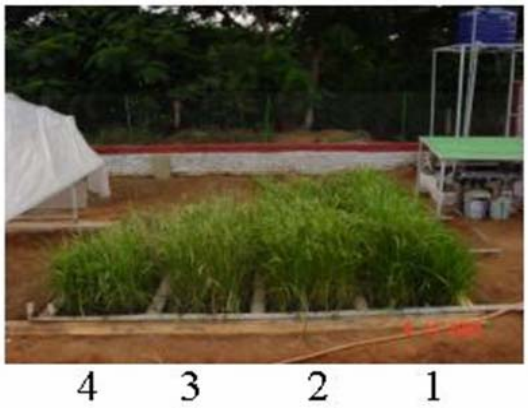

Fig 12b: Phenotypic difference reflected in plant height in different stress regimes

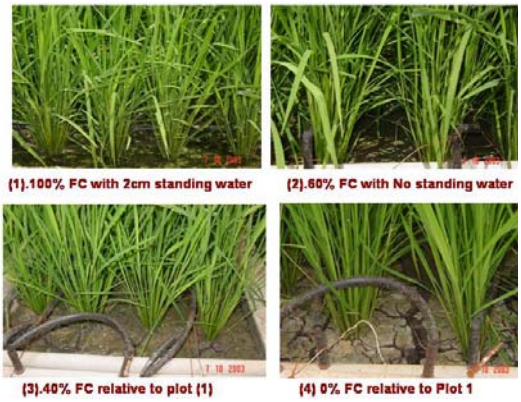

Fig 13: Drought stress induction in field in 4 different plot A1, A2 are shown on the top, A3 and A4 are shown in the bottom

Fig 14: Early booting in stressed samples compared to the control

Fig 15: Decreased plant height with the severity of drought stress; A1, A2, A3, A4 are shown from left to right

Fig 16: A view on day 71 where A4 dried up completely and A3 started drying up

Fig 17: A view on day 75 where A3 and A4 are dried up completely

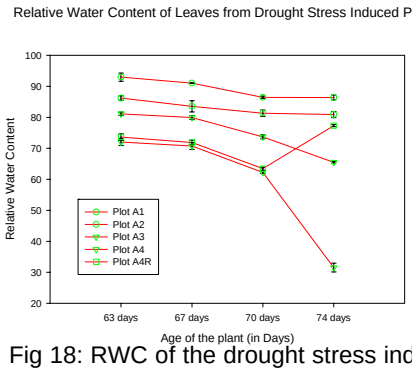

Fig 18: RWC of the drought stress induced rice seedlings shown from day 63 to day

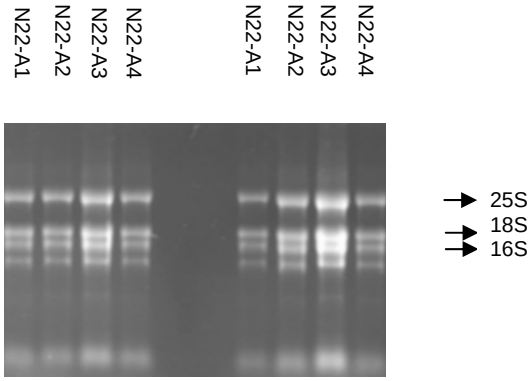

Fig 19: Denatured agarose gel photograph of total RNA isolated from drought stress induced plants.
(A) From 50 days old plants. (B) From 55 days old plants.

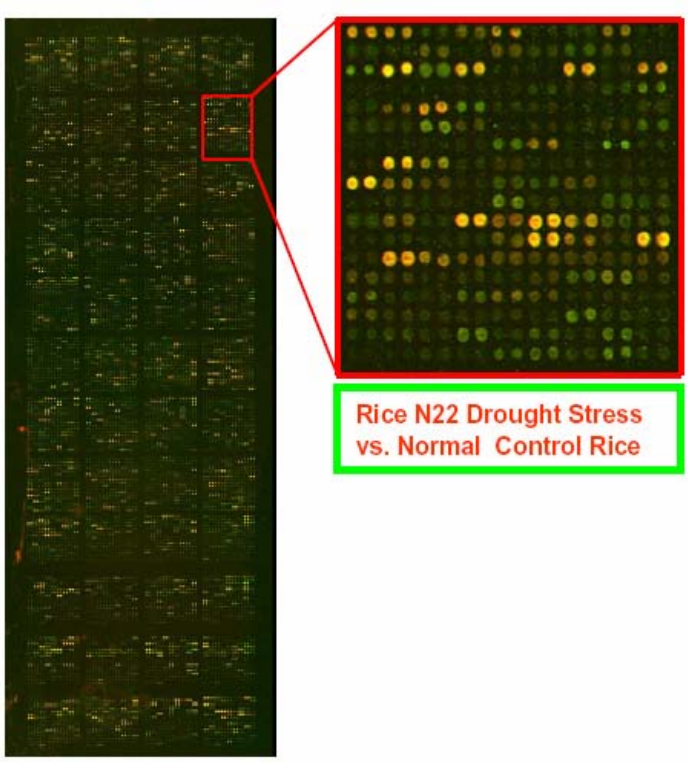

Fig 20: A 15.5 K chip printed on superamine substrate with low background; enlarged portion of a sub block showing uniformity of the spots.

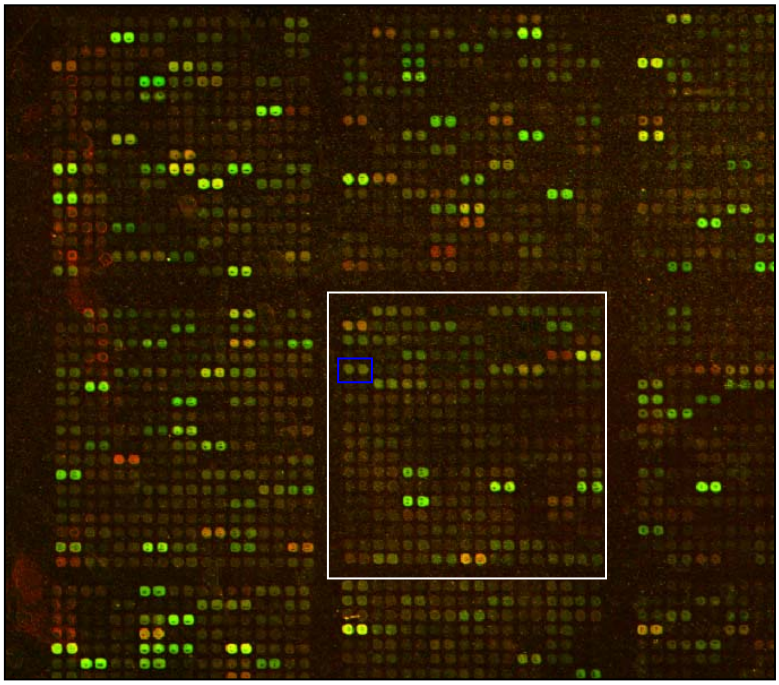

Fig 21: A block of 324 spots (shown in white outline) printed in duplicate (shown in blue outline) with 162 clones, with uniform distance between the blocks

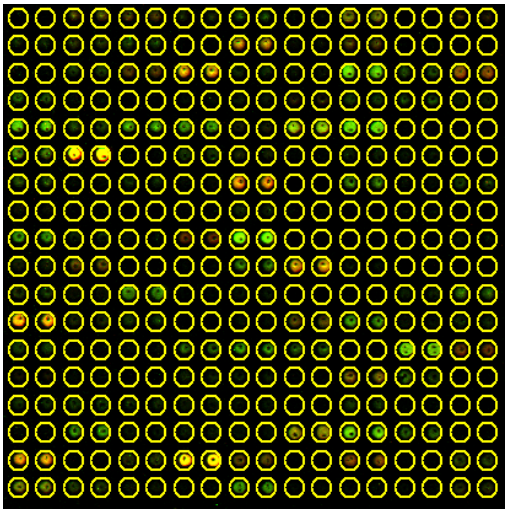

Fig 22a: A Block superimposed Cy3, Cy5 channels with 324 features after gridding to quantify the intensities

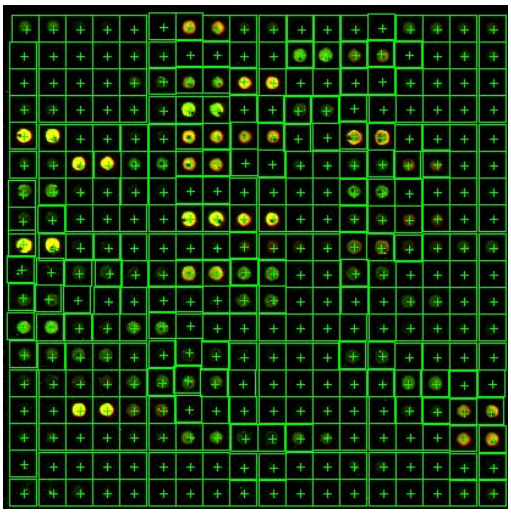

Fig 22b: A Block superimposed Cy3, Cy5 channels with 324 features with efficient gridding resulted in accurate spot detection indicated with symbol "+".

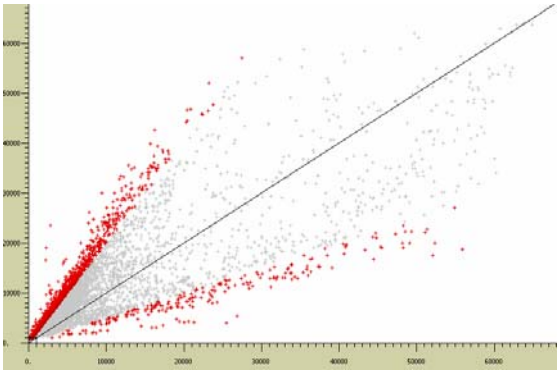

Fig 23: Scatter plot of two channel intensities before normalization

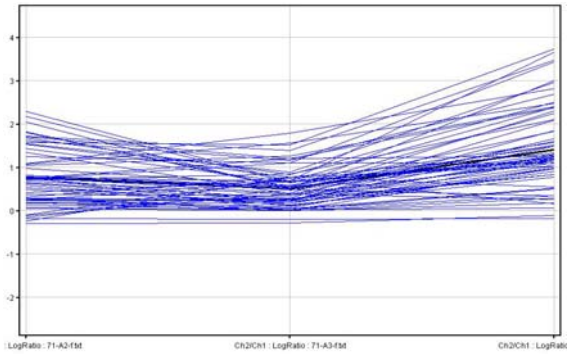

Fig 24a: Expression graph of spike in controls from Day 71 slides before normalization; non zero centered values are obvious from the expression plots

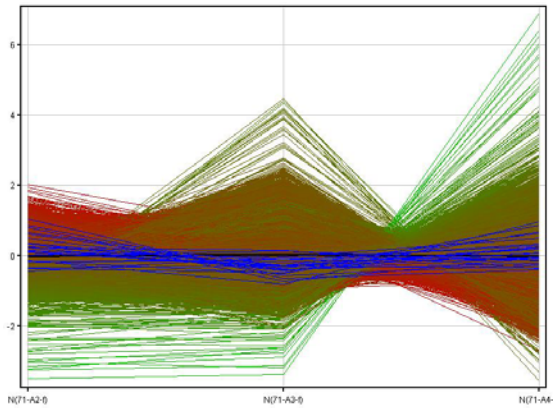

Fig 24b: Post normalized expression graph of the day 71 showing zero centered values; Spike in controls are highlighted with blue color

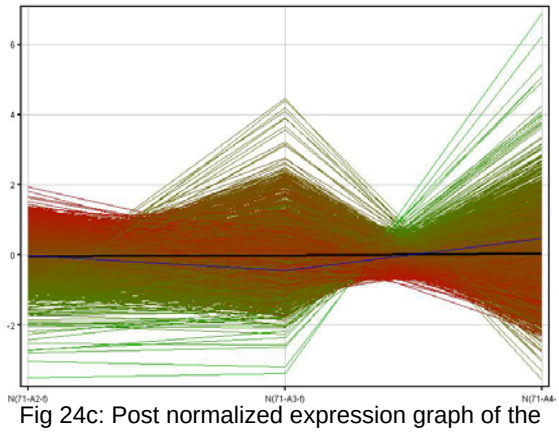

Fig 24c: Post normalized expression graph of the day 71 after folding the replicates with median; Expression graph with median of spike-in control shown in blue color

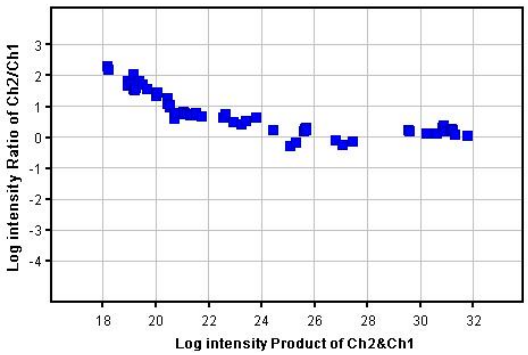

Fig 25a: Pre normalized MVA plot showing spike in controls from Day 71 A2 sample

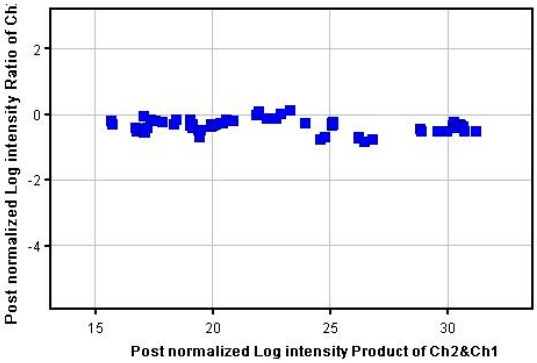

Fig 25b: Post normalized MVA plot showing spike in controls from Day 71 A2

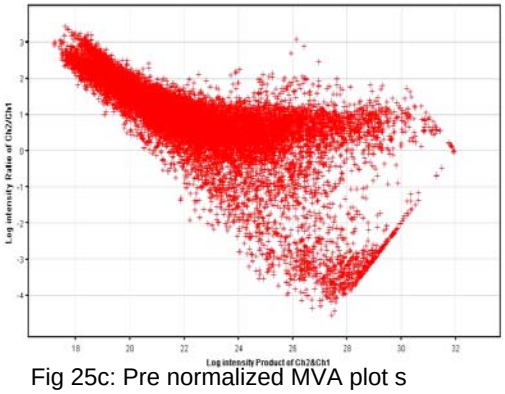

Fig 25c: Pre normalized MVA plot s showing all features of the array from Day 71 A2 sample

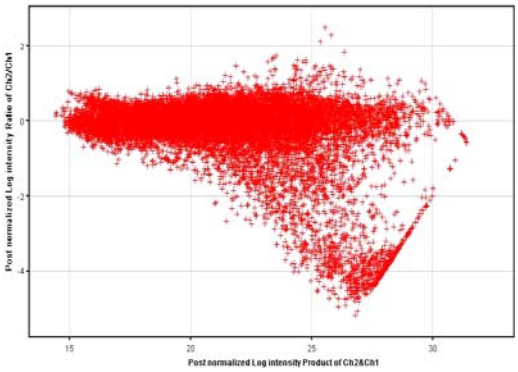

Fig 25d: Post normalized MVA plot s showing all features of the array from Day 71 A2 sample

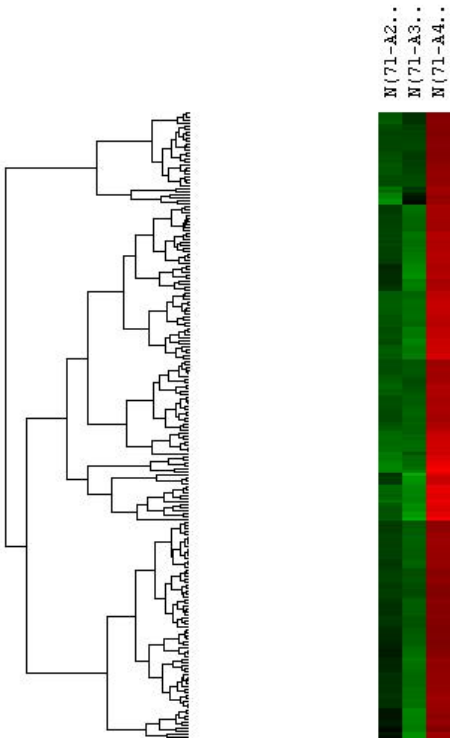

Fig 26a: Hierarchical clustering of genes showing different groups of genes classified based on the similarity in expression pattern from Day 71 A2, A3, and A4 samples (A partial view)

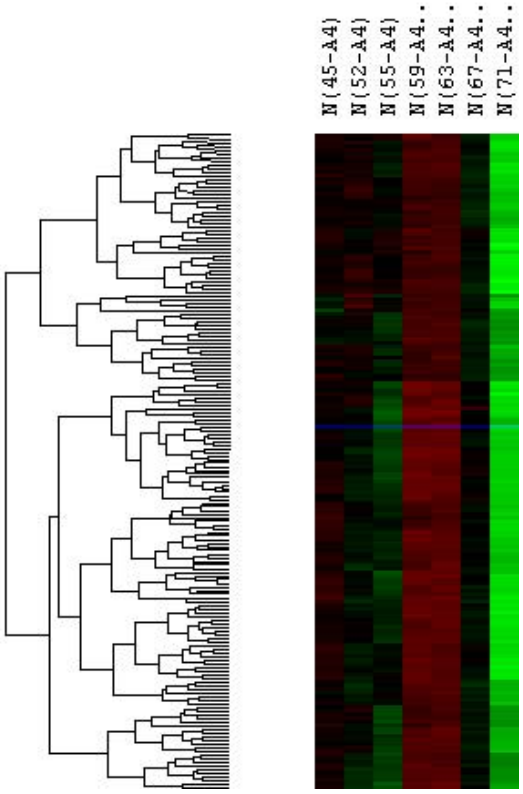

Fig 26b: Hierarchical clustering of genes showing different groups of genes classified based on the similarity in expression pattern from Day 45 to Day 71 from A3 samples (A partial view)

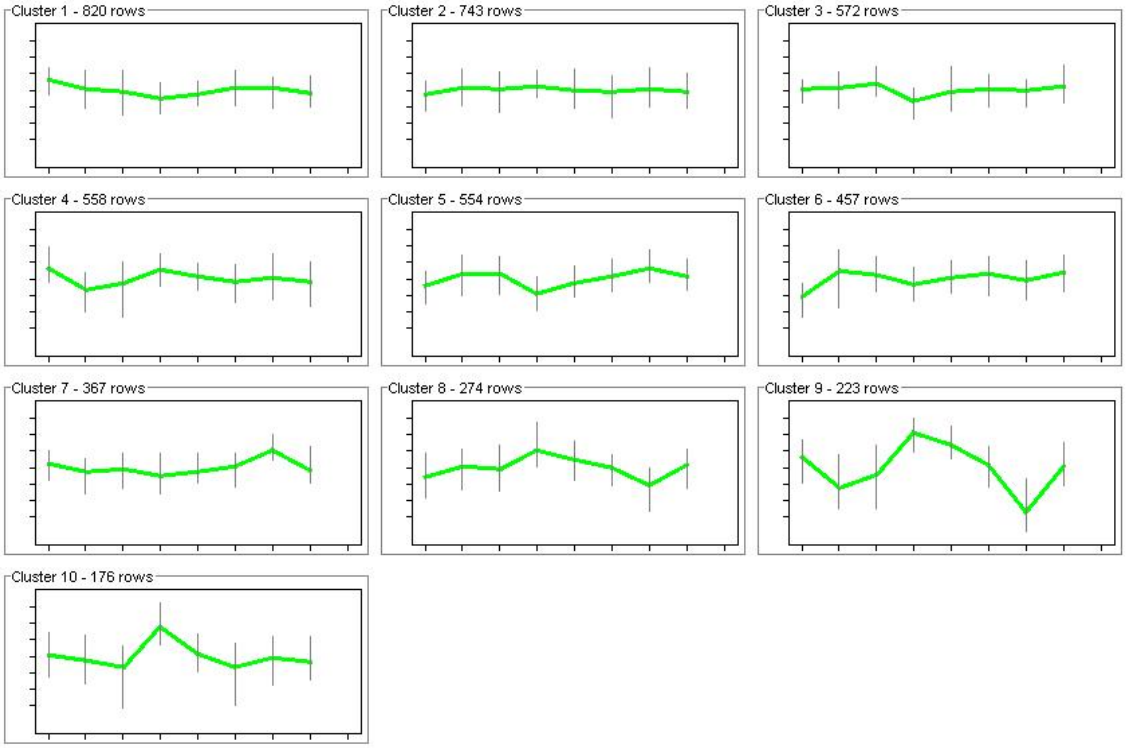

Fig 27a: K-means clustered expression graphs across the experiments from day 45 to day 75 in A2 (60%FC)

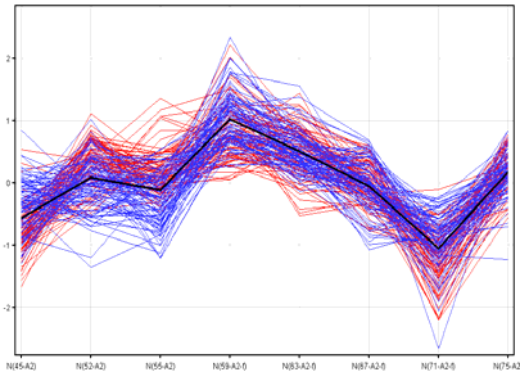

Fig 27b : Cluster 8 showing group of genes showing similar expression pattern from day

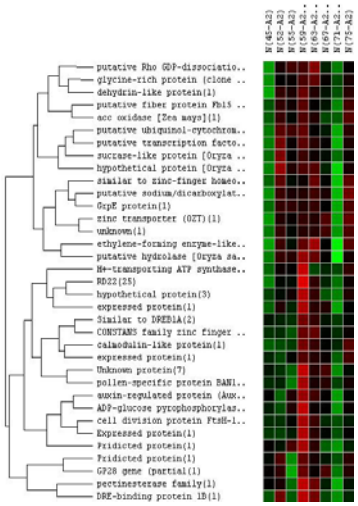

Fig 27c: Hierarchically clustered expression map of genes from cluster 8 from day 45 to day 75 in A2 stress regimes; a partial view

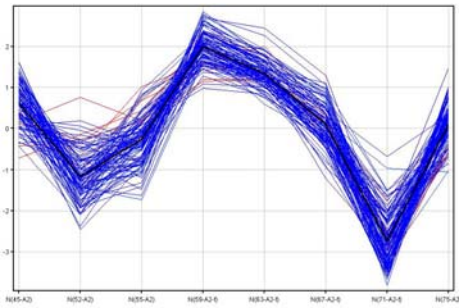

Fig 27d: Expression graph of cluster9 of A2 stress regime from day 45 to day 75; genes from rice are shown in blue and genes from pearl millet are shown in red color

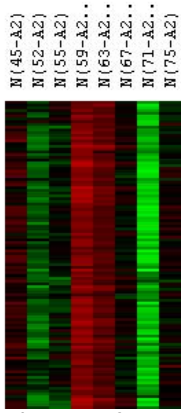

Fig 27e: Heat map of cluster9 of A2 stress regime (from day 45 to day 75) showing similar expression pattern

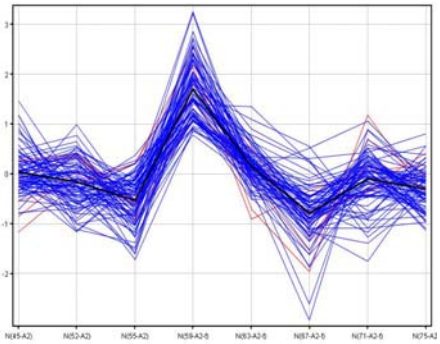

Fig 27f: Expression graph of cluster 10 of A2 stress regime from day 45 to day 75; Genes from rice are shown in blue and genes from pearl millet are shown in red color

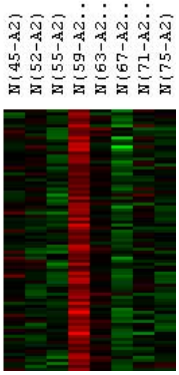

Fig 27g: Heat map of cluster10 of A2 stress regime (from day 45 to day 75) showing similar expression pattern

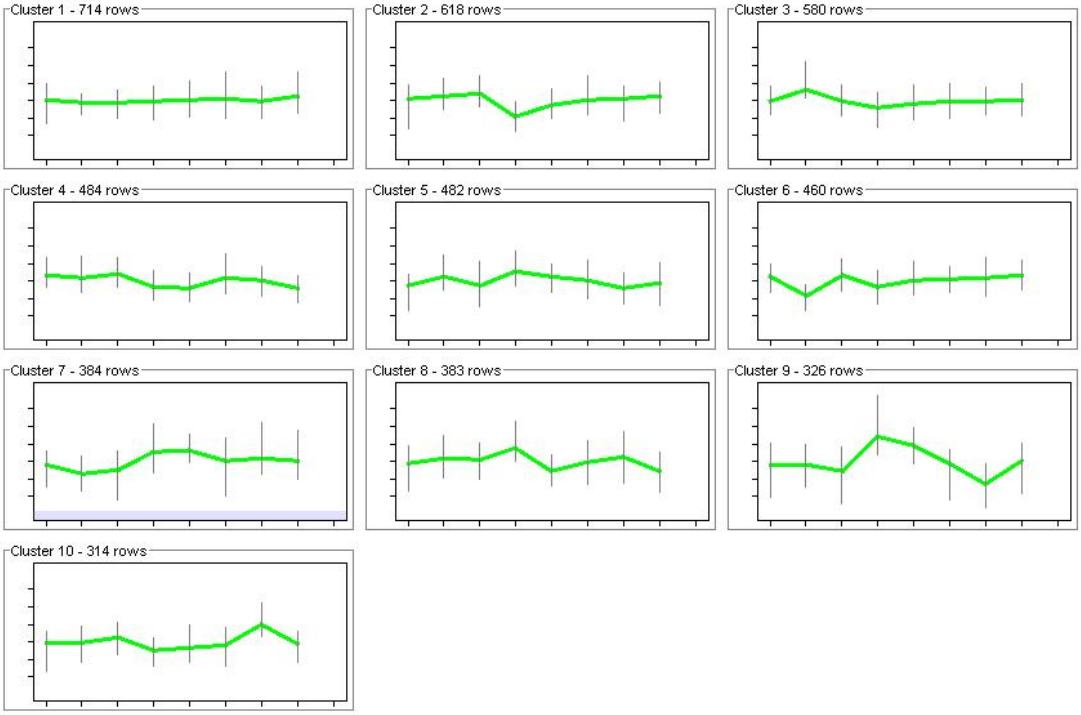

Fig 28a: K-means clustered expression graphs from day 45 to day 75 from A3 (40%FC)

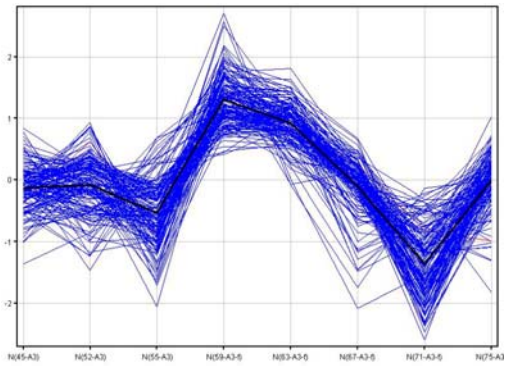

Fig 28b: Expression graph of cluster9 of A3 stress regime from day 45 to day 75; Genes from rice are shown in blue and genes from pearl millet are shown in red color

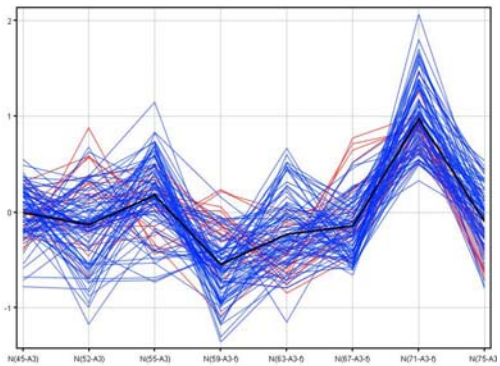

Fig 28c: Expression graph of cluster 10 of A3 stress regime from day 45 to day 75; Genes from rice are shown in blue and genes from pearl millet are shown in red color

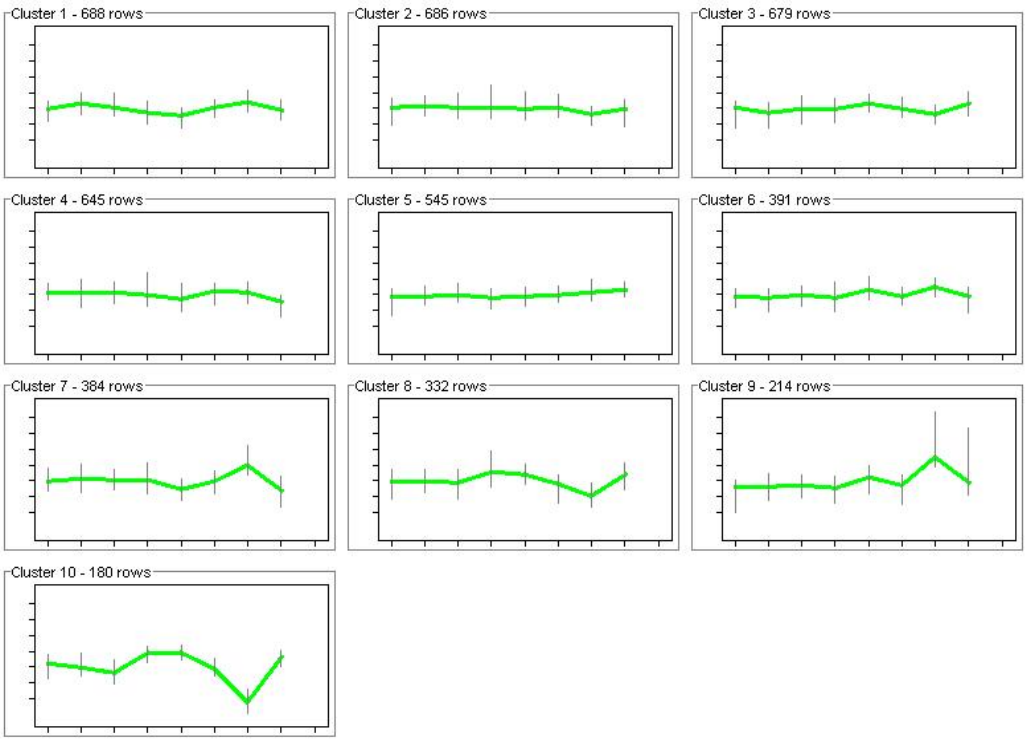

Fig 29a: K-means clustered expression graphs from day 45 to day 75 from A4 (15%FC)

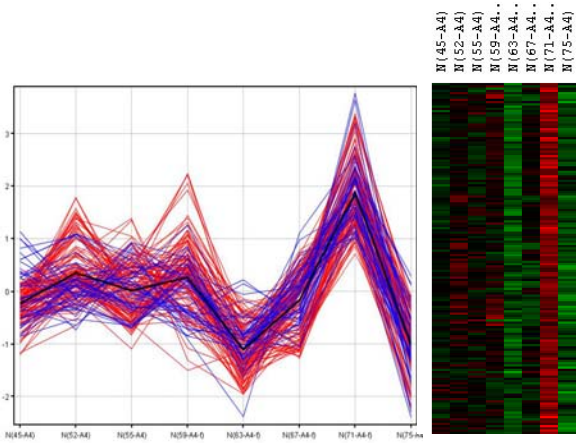

Fig 29b: Expression graph of cluster 7 of A4 stress regime from day 45 to day 75. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 29c: Expression map of cluster 7

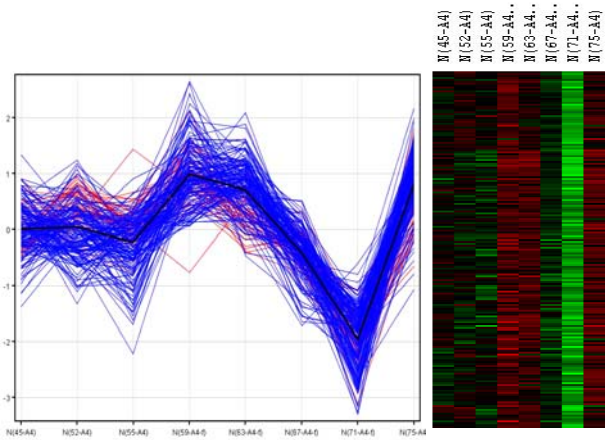

Fig 29d: Expression graph of cluster 8 of A4 stress regime from day 45 to day 75. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 29e: Expression map of cluster 8

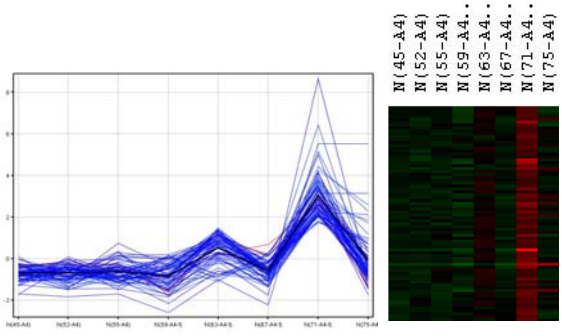

Fig 29f: Expression graph of cluster 9 of A4 stress regime from day 45 to day 75. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 29g: Expression map of cluster 9

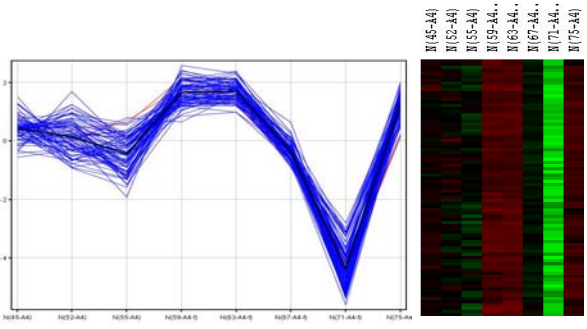

Fig 29h: Expression graph of cluster 10 of A4 stress regime from day 45 to day 75. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 29i: Expression map of cluster10

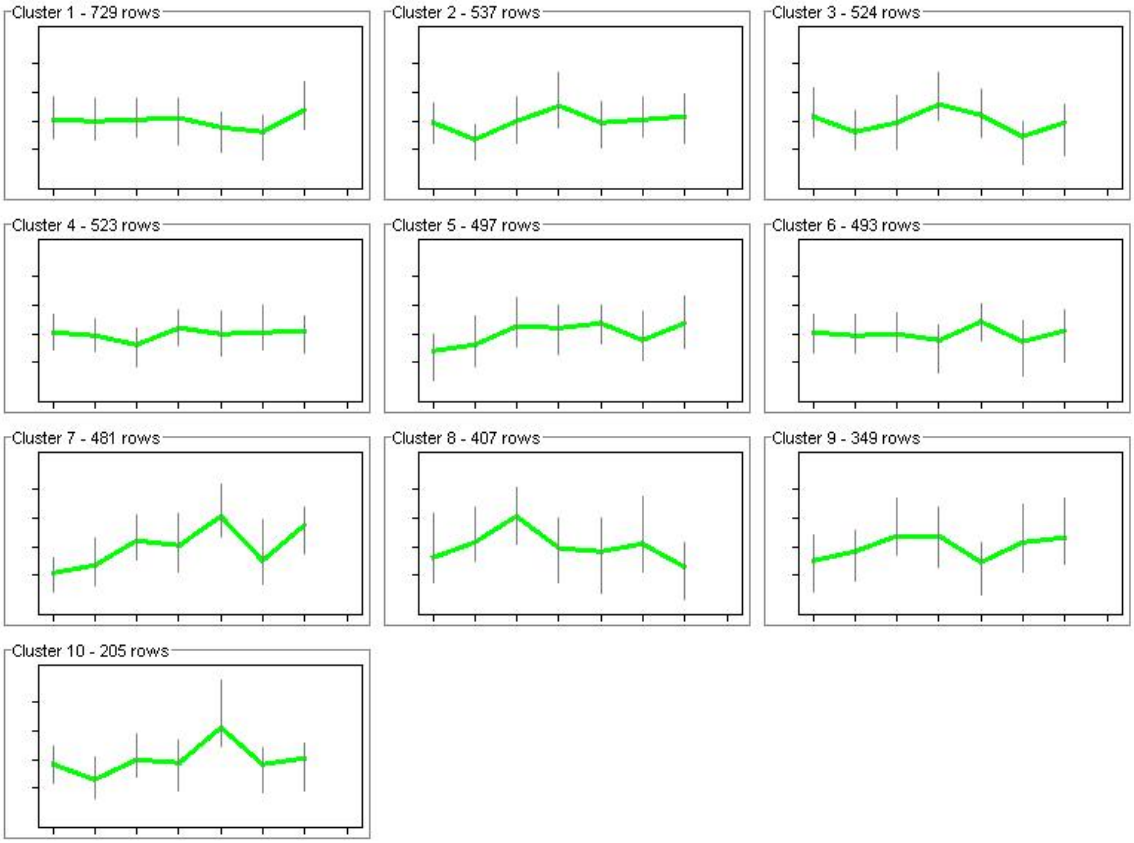

Fig 30a: K-means clustered expression graphs from panicle samples of day 59 to day 75 from A2 (60%FC), Expression graph also shows gene expression pattern at day 78 and day 81 recovery samples (last two data points on X axis) which are ignored in current analysis

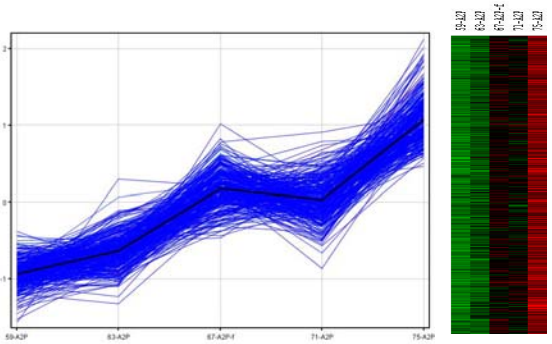

Fig 30b: Expression graph of cluster 7 of A2 stress regime from day 59 to day 75 from Panicle sample. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 30c: Expression map of cluster 7

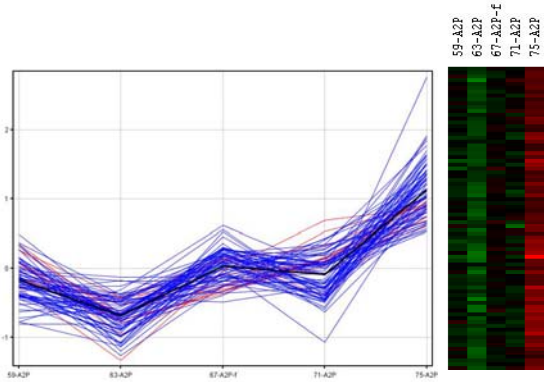

Fig 30d: Expression graph of cluster 10 of A2 stress regime from day 59 to day 75 from Panicle sample. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 30e: Expression map of cluster 10

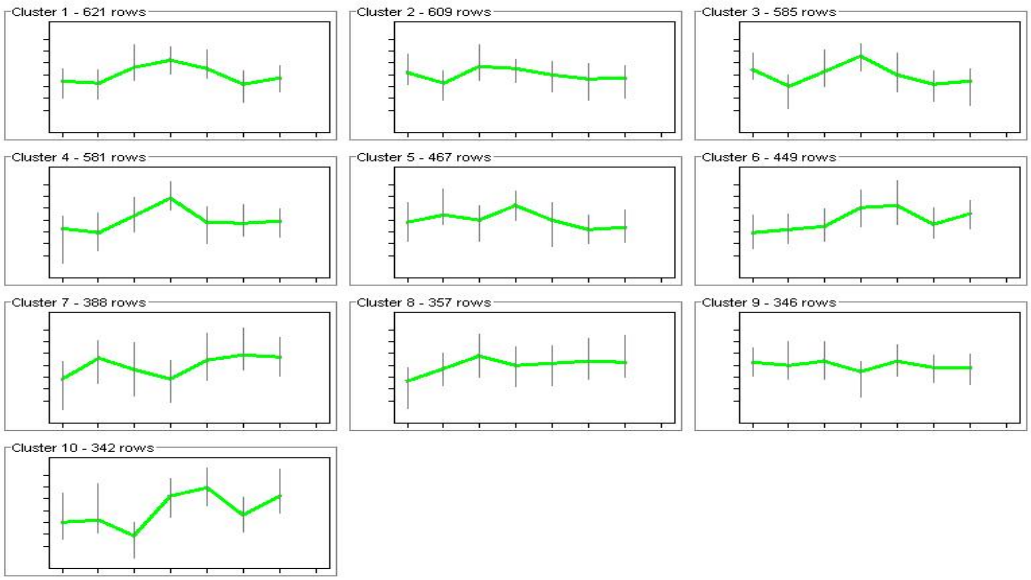

Fig 31a: K-means clustered expression graphs from panicle samples of day 59 to day 75 from A3 (40%FC), Expression graph also shows gene expression pattern at day 78 and day 81 recovery samples and are ignored in current analysis

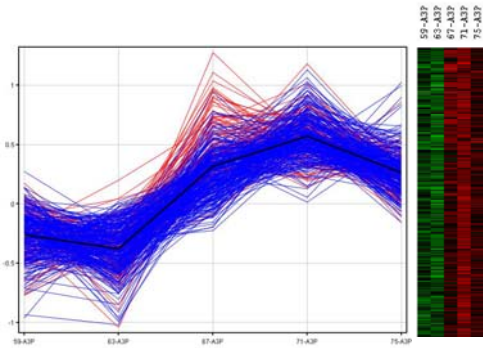

Fig 31b: Expression graph of cluster 1 of A3 stress regime from day 59 to day 75 from Panicle sample. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 31c: Expression map of cluster 1

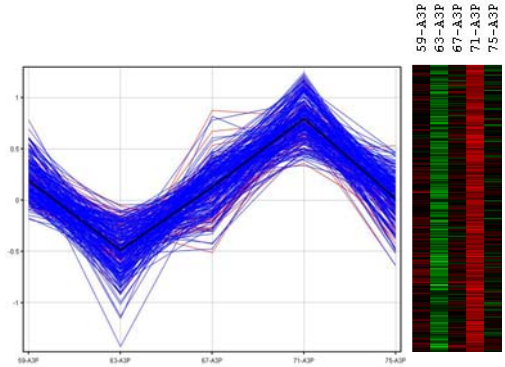

Fig 31d: Expression graph of cluster 3 of A3 stress regime from day 59 to day 75 from Panicle sample. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 31e: Expression map of cluster 3

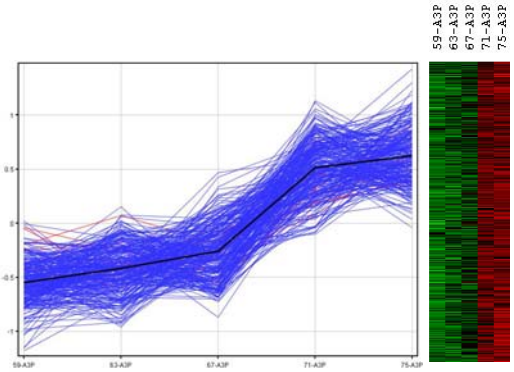

Fig 31f: Expression graph of cluster 6 of A3 stress regime from day 59 to day 75 from Panicle sample. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 31g: Expression map of cluster 6

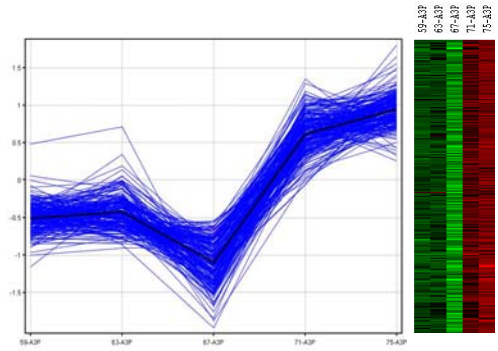

Fig 31h: Expression graph of cluster 10 of A3 stress regime from day 59 to day 75 from Panicle sample. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 31i: Expression map of cluster 10

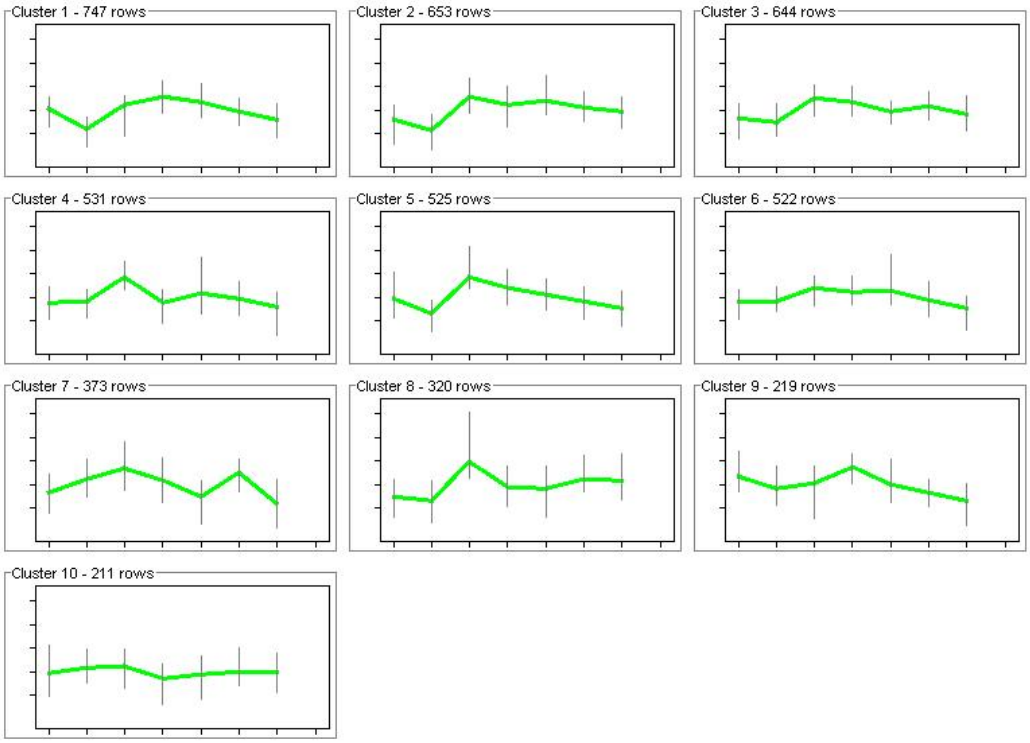

Fig 32a: K-means clustered expression graphs from panicle samples of day 59 to day 75 from A4 (15%FC), Expression graph also shows gene expression pattern at day 78 and day 81 recovery samples and are ignored in current analysis.

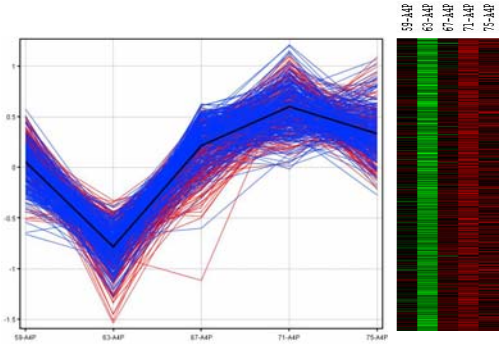

Fig 32b: Expression graph of cluster 1 of A4 stress regime from day 59 to day 75 from Panicle sample. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 32c: Expression map of cluster 1

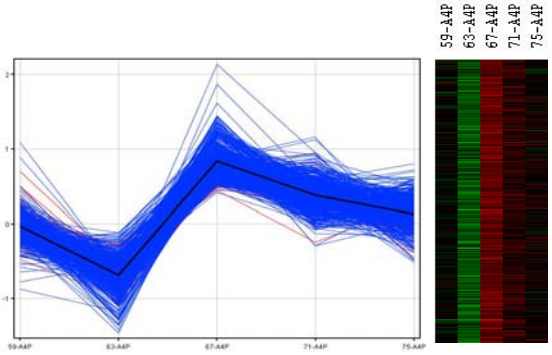

Fig 32d: Expression graph of cluster 5 of A4 stress regime from day 59 to day 75 from Panicle sample. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 32e: Expression map of cluster 5

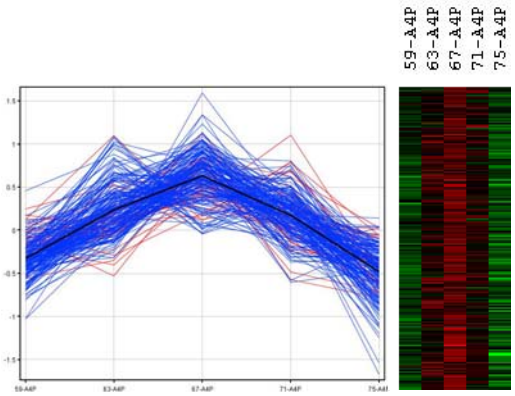

Fig 32f: Expression graph of cluster 7 of A4 stress regime from day 59 to day 75 from Panicle sample. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 32g: Expression map of cluster 7

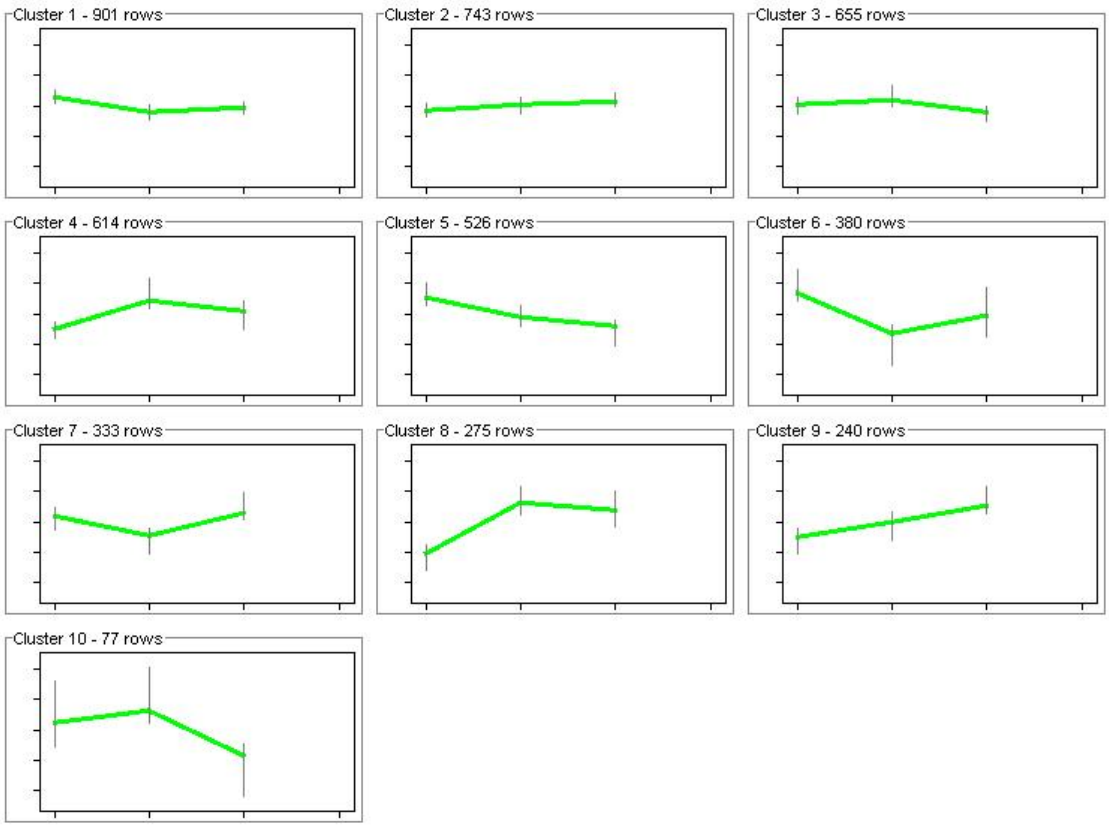

Fig 33a: K-means clustered expression graphs from leaf samples of day 45 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

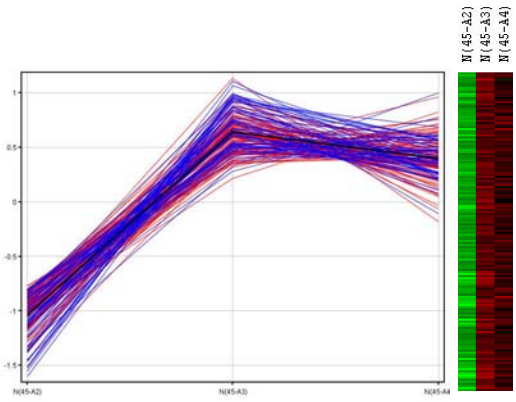

Fig 33b: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 45. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 33c: Expression map of Cluster 8

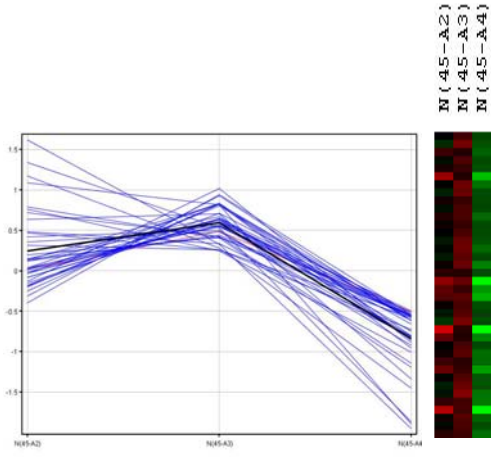

Fig 33d: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 45. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 33e: Expression map of Cluster 10

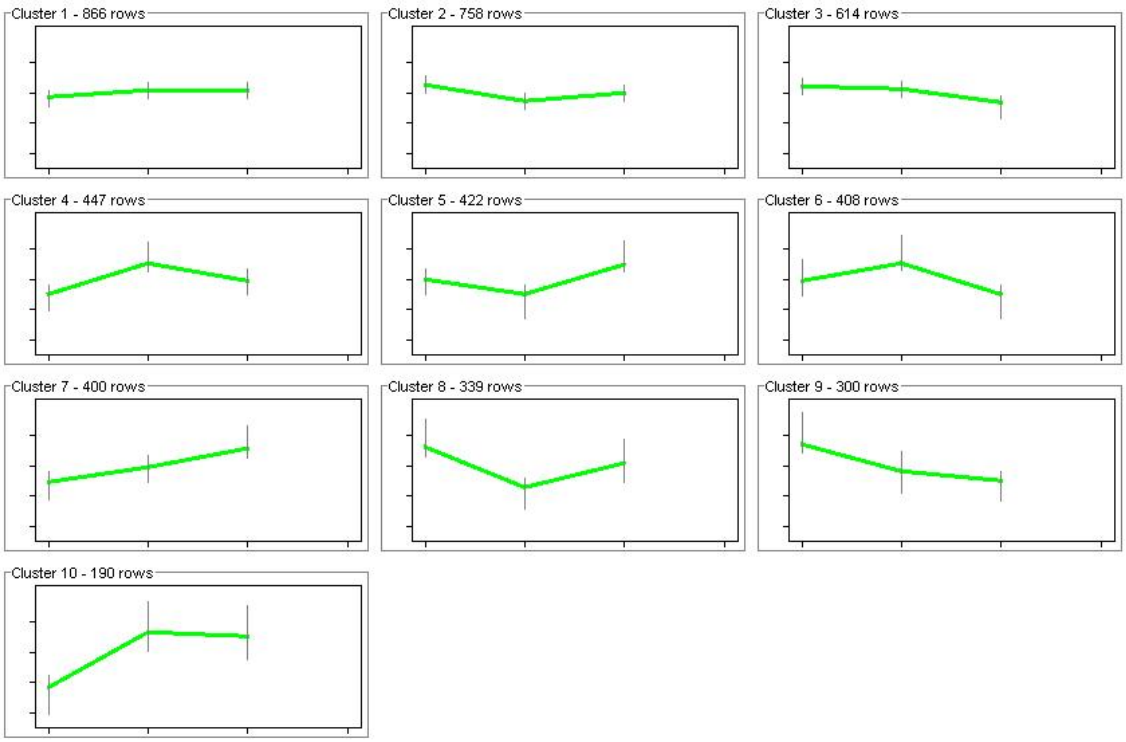

Fig 34a: K-means clustered expression graphs from leaf samples of day 52 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

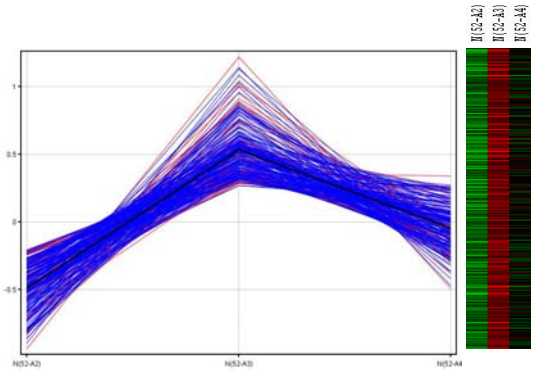

Fig 34b: Expression graph of Cluster 4 from samples A2, A3 and A4 of day 52. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 34c: Expression map of Cluster 4

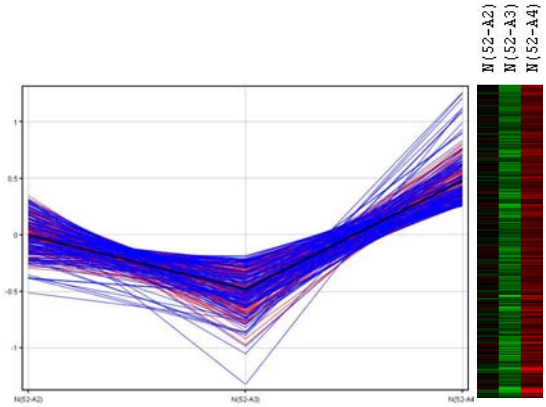

Fig 34d: Expression graph of Cluster 5 from samples A2, A3 and A4 of day 52. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 34e: Expression map of Cluster 5

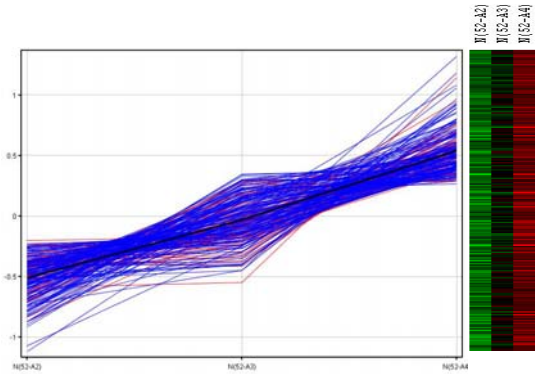

Fig 34f: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 52. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 34g: Expression map of Cluster 7

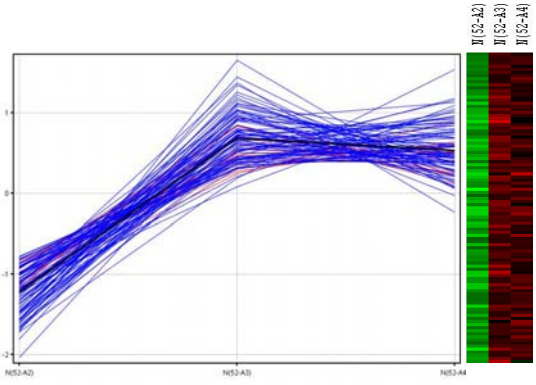

Fig 34h: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 52. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 34i: Expression map of Cluster 10

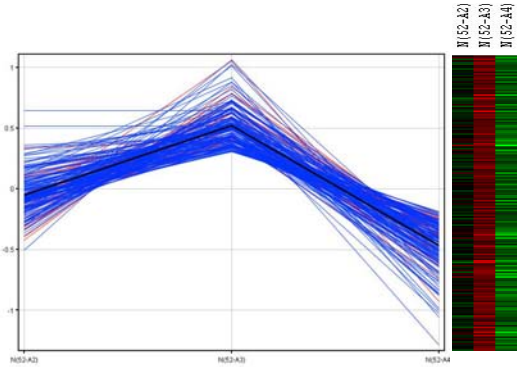

Fig 34j: Expression graph of Cluster 6 from samples A2, A3 and A4 of day 52. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 34k: Expression map of Cluster 6

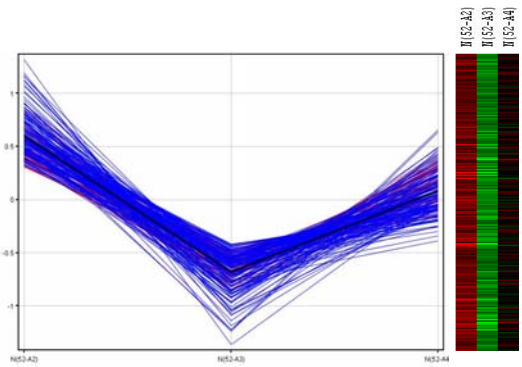

Fig 34l: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 52. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 34m: Expression map of Cluster 8

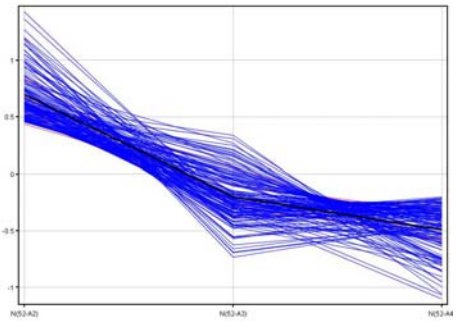

Fig 34n: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 52. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 34o: Expression map of Cluster 9

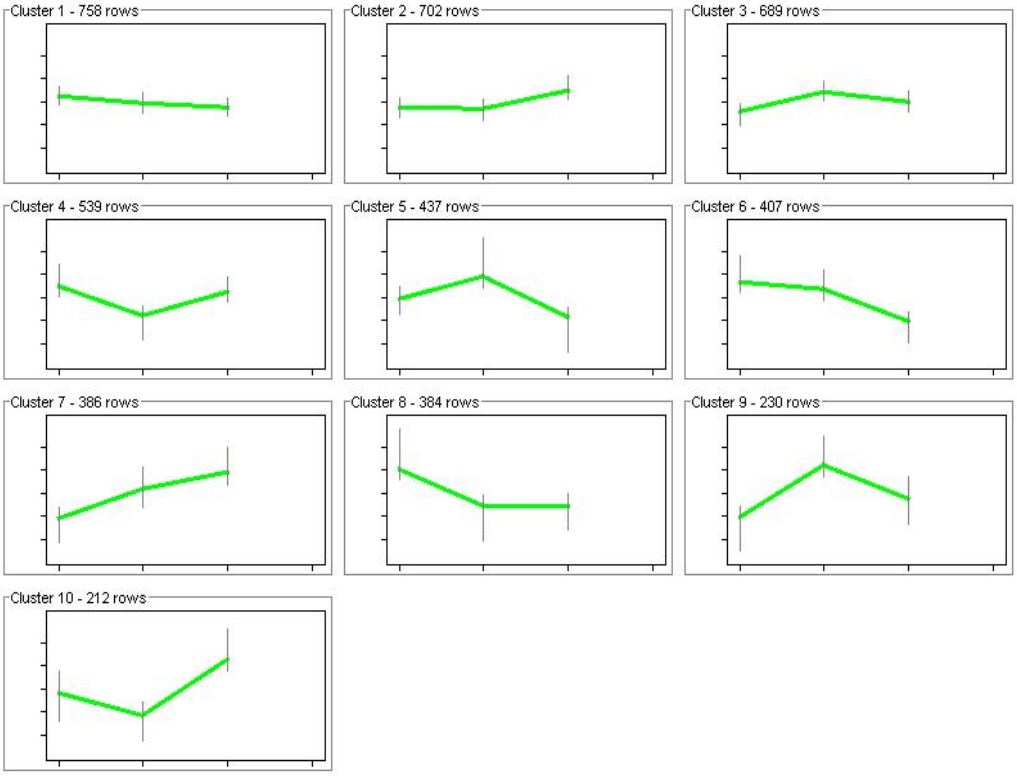

Fig 35a: K-means clustered expression graphs from leaf samples of day 55 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

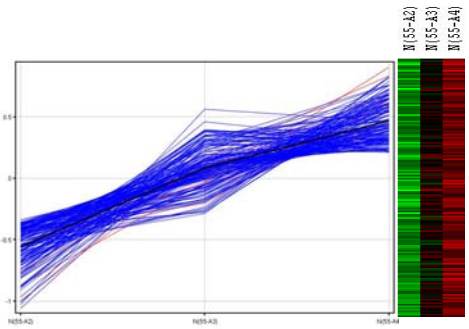

Fig 35b: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 55. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 35c: Expression map of Cluster 7

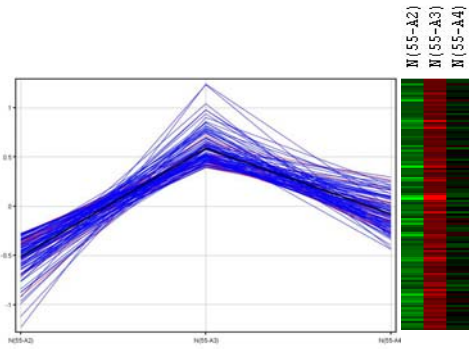

Fig 35d: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 55. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 35e: Expression map of Cluster 9

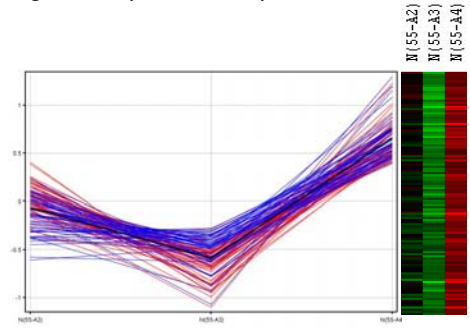

Fig 35f: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 55. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 35g: Expression map of Cluster 10

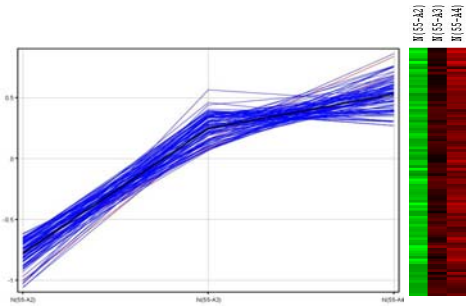

Fig 35h: Expression graph of Cluster 7_3 (sub cluster3 of cluster7) from samples A2, A3 and A4 of day 55. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 35i: Expression map of Cluster 7_3

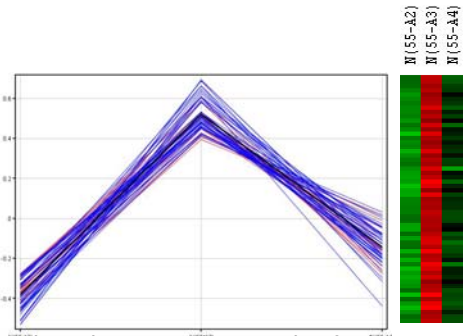

Fig 35j: Expression graph of Cluster 9_1 from samples A2, A3 and A4 of day 55. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 35k: Expression map of Cluster 9_1

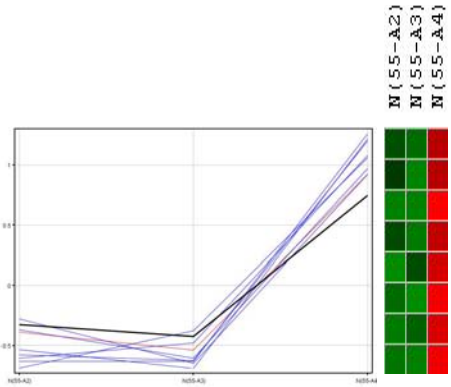

Fig 35l: Expression graph of Cluster 10_2 from samples A2, A3 and A4 of day 55. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 35m: Expression map of Cluster 10_2

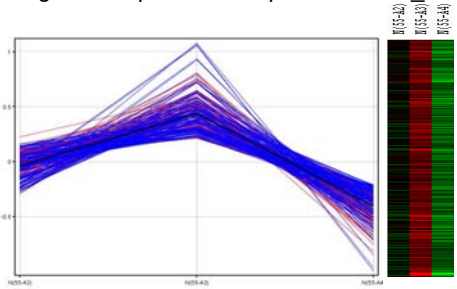

Fig 35n: Expression graph of Cluster 5 from samples A2, A3 and A4 of day 55. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 35o: Expression map of Cluster 5

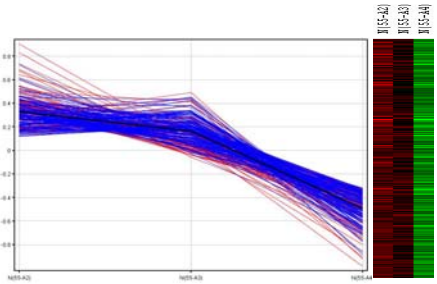

Fig 35p: Expression graph of Cluster 6 from samples A2, A3 and A4 of day 55. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 35q: Expression map of Cluster 6

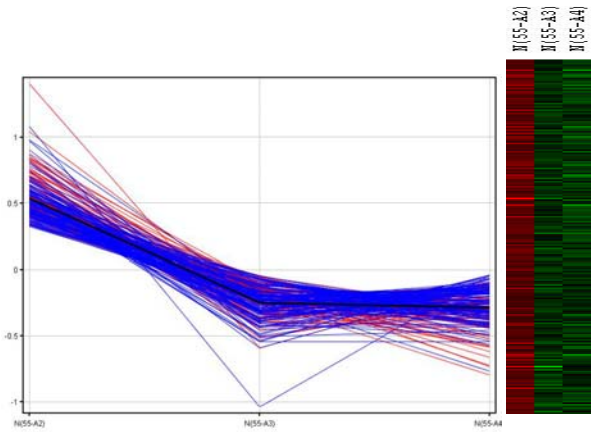

Fig 35r: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 55. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 35s: Expression map of Cluster 7

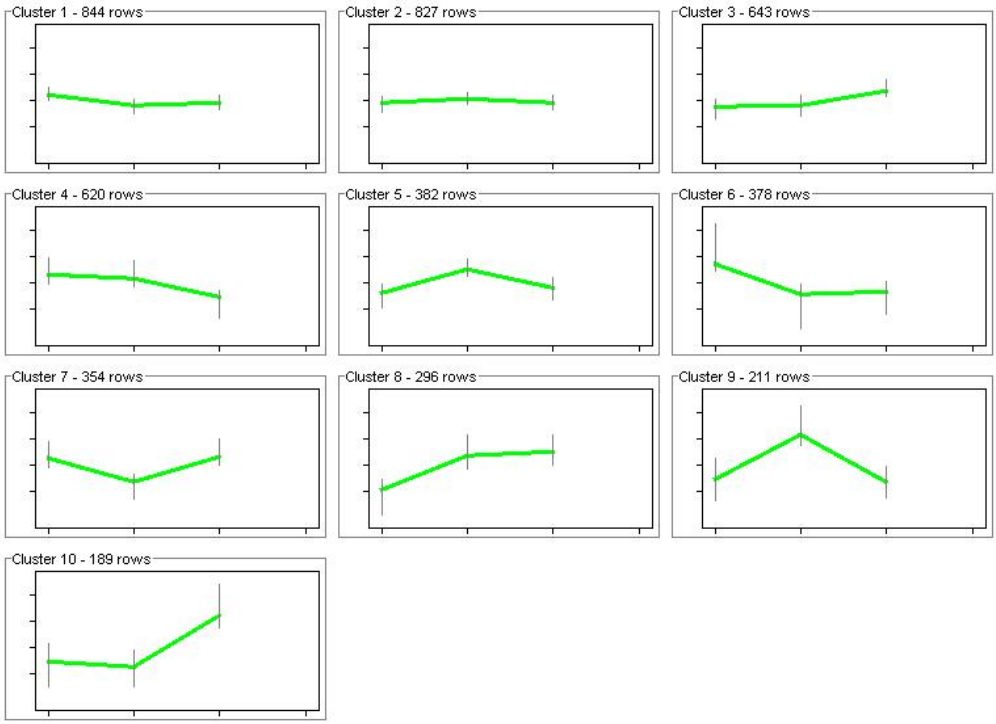

Fig 36a: K-means clustered expression graphs from leaf samples of day 59 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

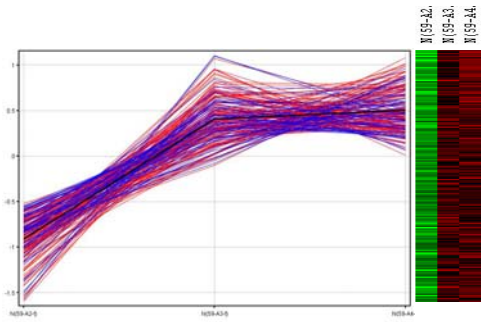

Fig 36b: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 59. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 36c: Expression map of Cluster 8

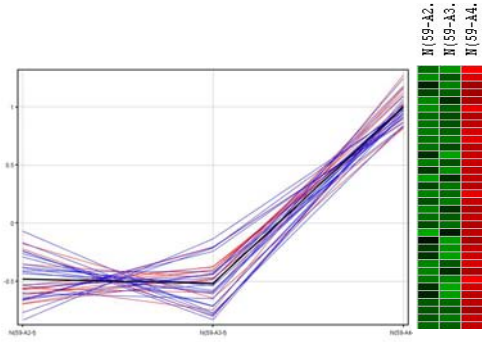

Fig 36d: Expression graph of Cluster 10_1 from samples A2, A3 and A4 of day 59. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 36e: Expression map of Cluster 10_1

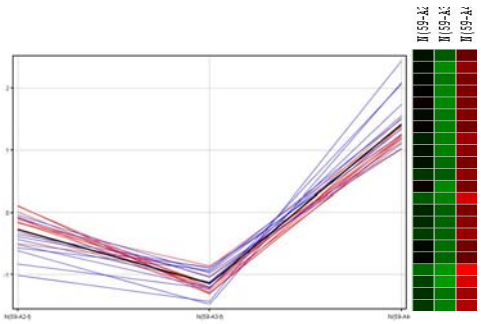

Fig 36f: Expression graph of Cluster 10_2 from samples A2, A3 and A4 of day 59. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 36g: Expression map of Cluster 10_2

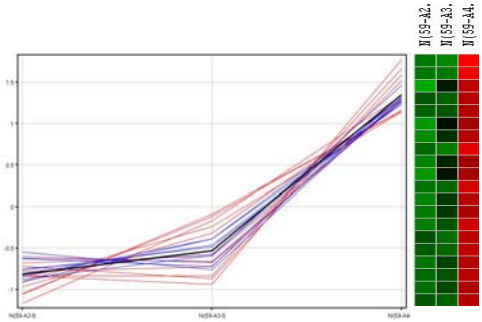

Fig 36h: Expression graph of Cluster 10_3 from samples A2, A3 and A4 of day 59. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 36i: Expression map of Cluster 10_3

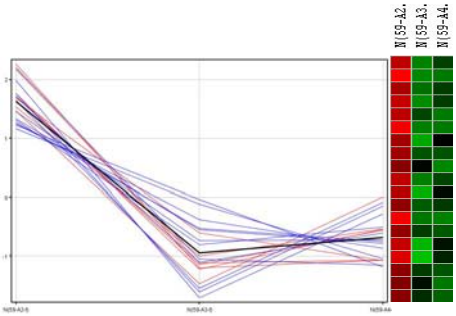

Fig 36j: Expression graph of Cluster 6_1 from samples A2, A3 and A4 of day 59. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 36k: Expression map of Cluster 6_1

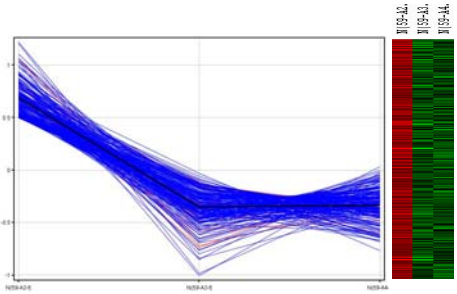

Fig 36l: Expression graph of Cluster 6_2 from samples A2, A3 and A4 of day 59. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 36m: Expression map of Cluster 6_2

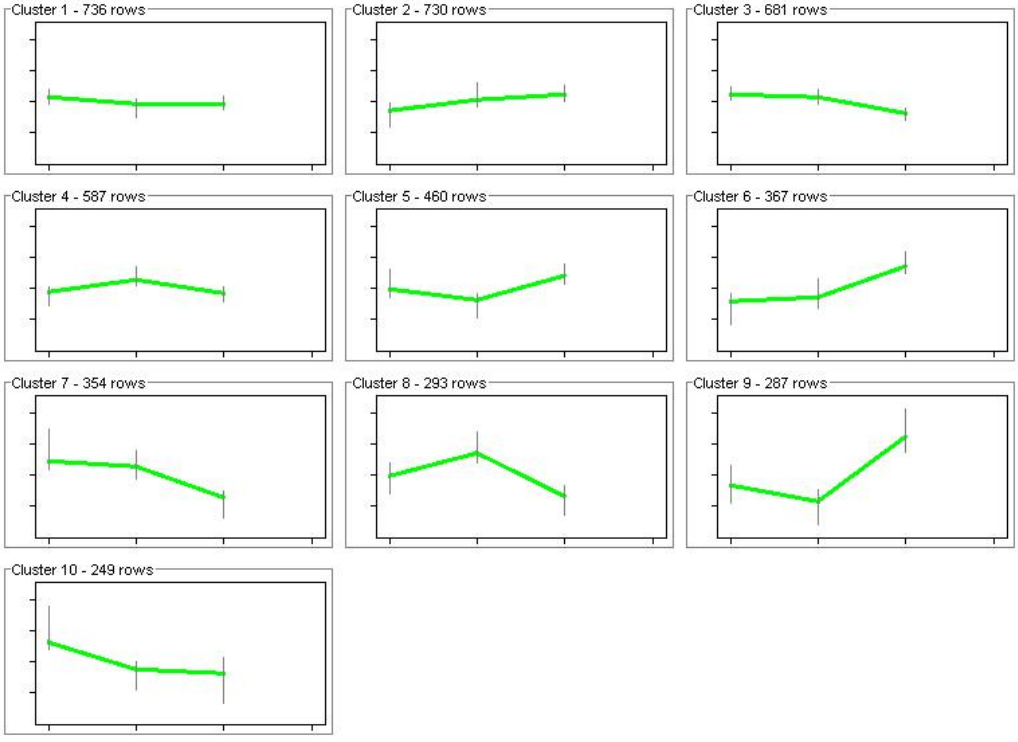

Fig 37a: K-means clustered expression graphs from leaf samples of day 63 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

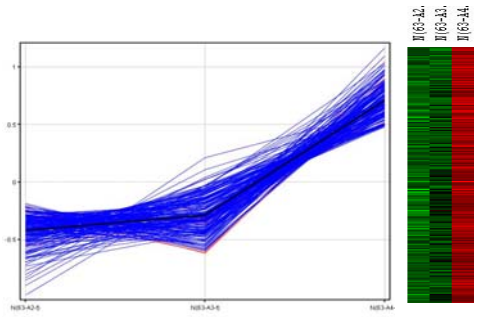

Fig 37b: Expression graph of Cluster 6 from samples A2, A3 and A4 of day 63. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 37c: Expression map of Cluster 6

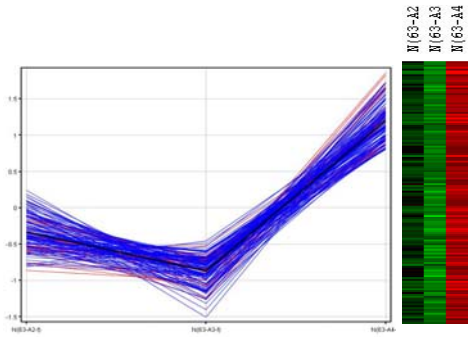

Fig 37d: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 63. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 37e: Expression map of Cluster 9

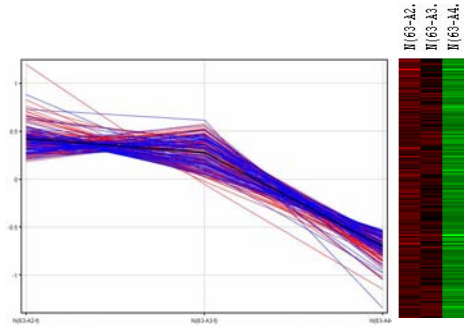

Fig 37f: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 63. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 37g: Expression map of Cluster 7

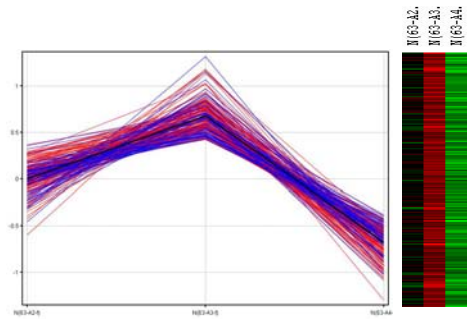

Fig 37h: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 63. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 37i: Expression map of Cluster 8

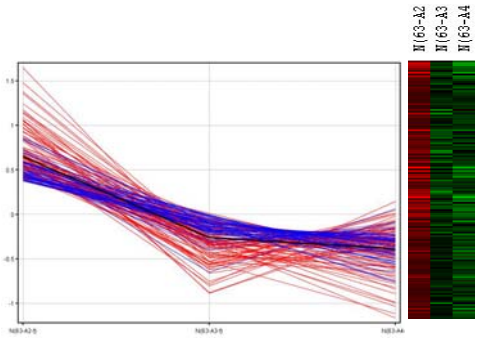

Fig 37j: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 63. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 37k: Expression map of Cluster 10

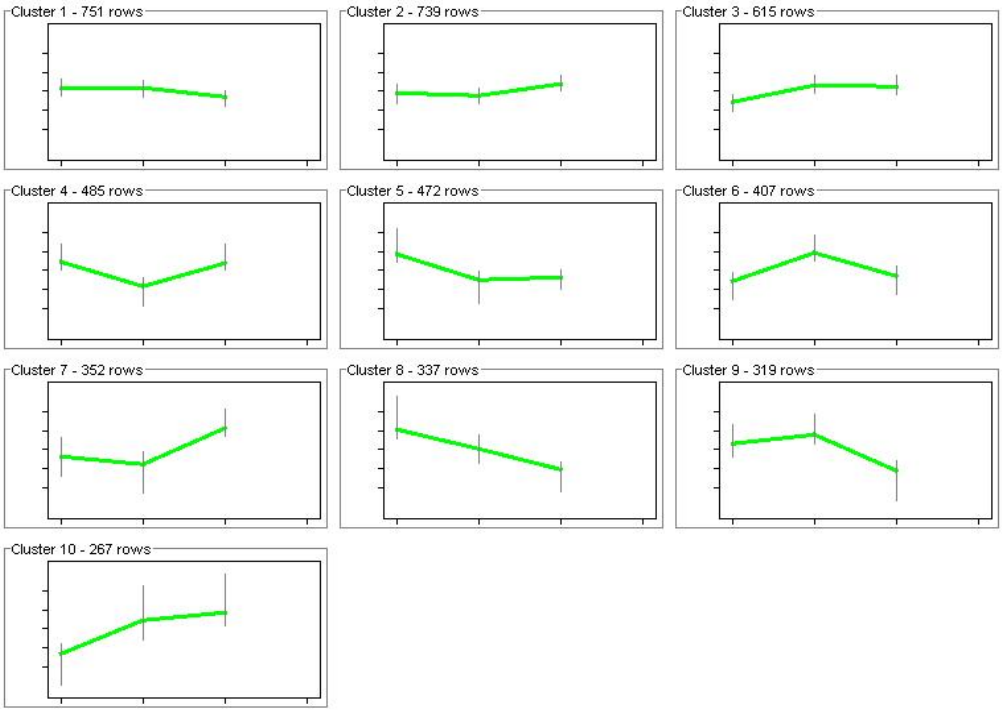

Fig 38a: K-means clustered expression graphs from leaf samples of day 67 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

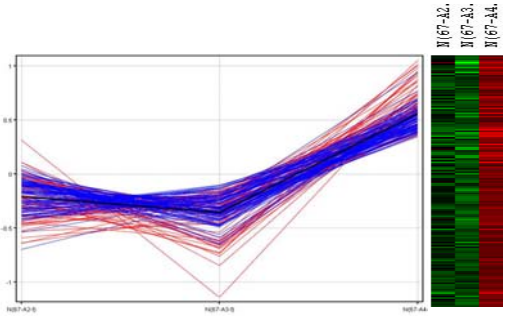

Fig 38b: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 67. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 38c: Expression map of Cluster 7

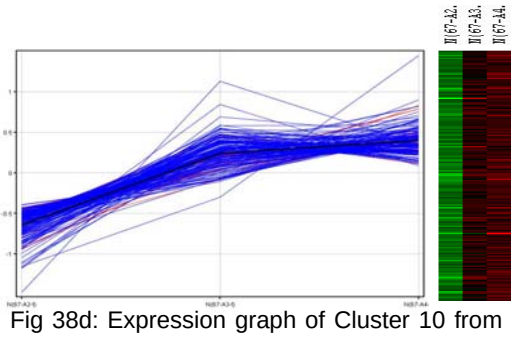

Fig 38d: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 67. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 38e: Expression map of Cluster 10

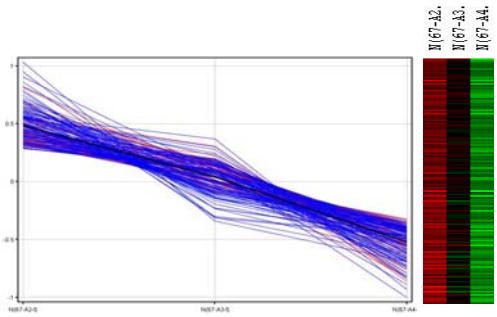

Fig 38f: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 67. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 38g: Expression map of Cluster 8

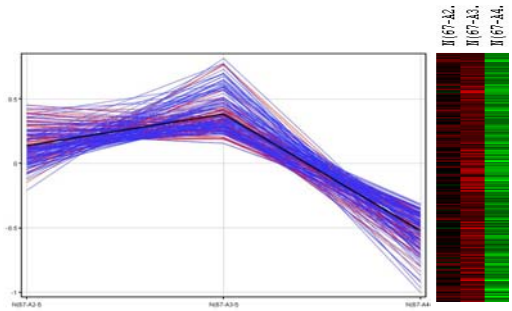

Fig 38h: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 67. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 38i: Expression map of Cluster 9

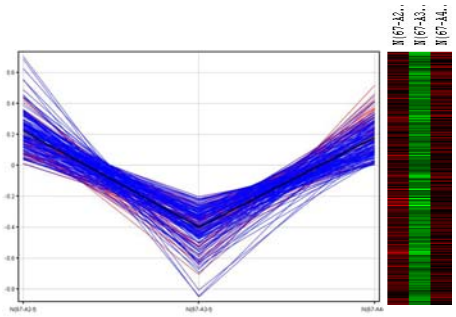

Fig 38j: Expression graph of Cluster 4 from samples A2, A3 and A4 of day 67. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 38k: Expression map of Cluster 4

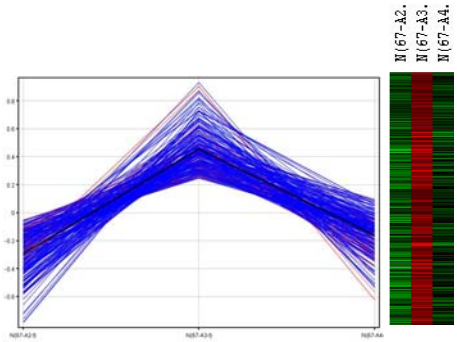

Fig 38l: Expression graph of Cluster 6 from samples A2, A3 and A4 of day 67. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 38m: Expression map of Cluster 6

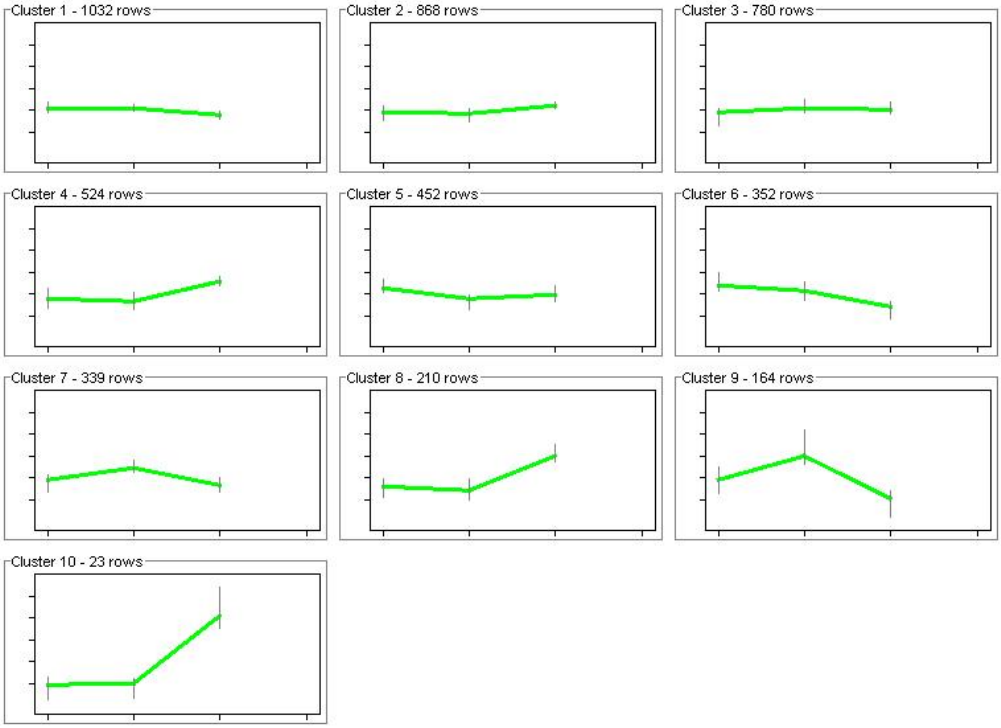

Fig 39a: K-means clustered expression graphs from leaf samples of day 71 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

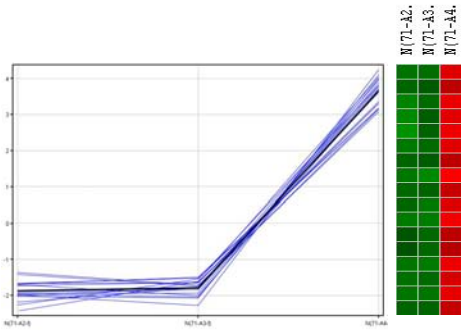

Fig 39b: Expression graph of Cluster 10_1 from samples A2, A3 and A4 of day 71. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 39c: Expression map of Cluster 10_1

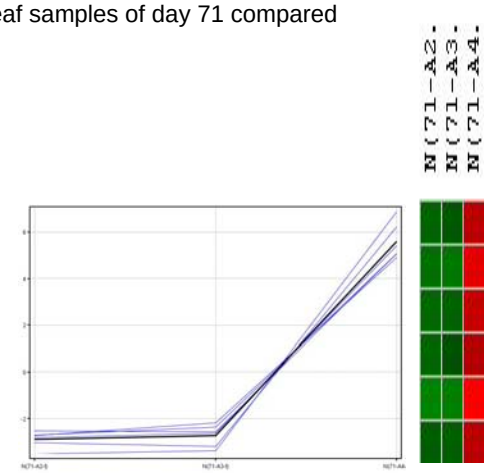

Fig 39d: Expression graph of Cluster 10_2 from samples A2, A3 and A4 of day 71. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 39e: Expression map of Cluster 10_2

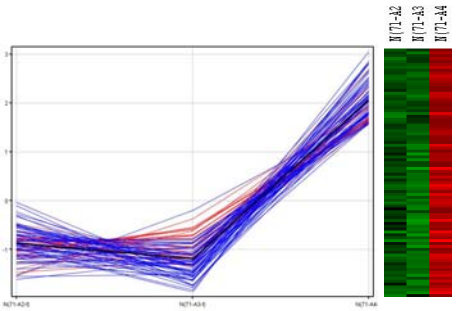

Fig 39f: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 71. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 39g: Expression map of Cluster 8

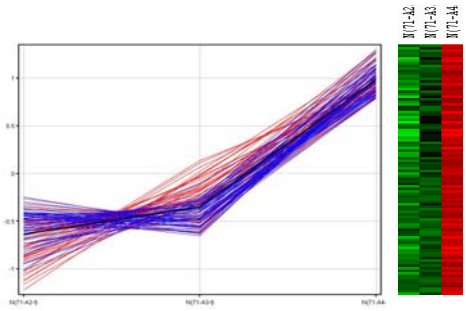

Fig 39h: Expression graph of Cluster 4_1 from samples A2, A3 and A4 of day 71. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 39i: Expression map of Cluster 4_1

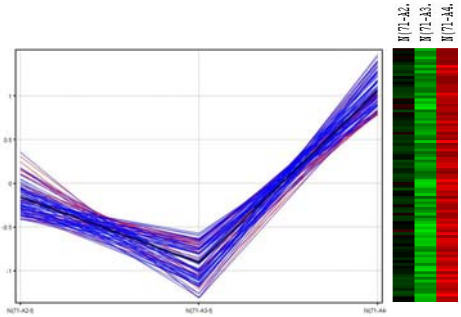

Fig 39j: Expression graph of Cluster 4_2 from samples A2, A3 and A4 of day 71. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 39k: Expression map of Cluster 4_2

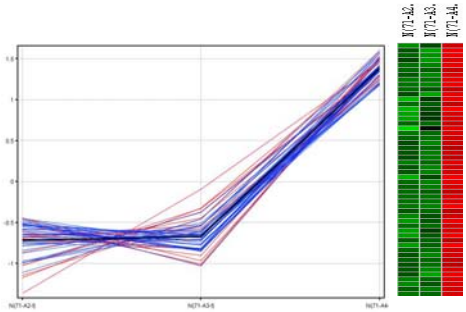

Fig 39l: Expression graph of Cluster 4_3 from samples A2, A3 and A4 of day 71. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 39m: Expression map of Cluster 4_3

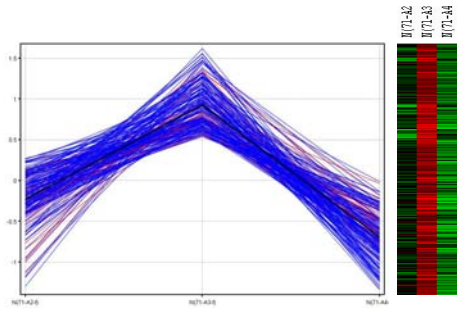

Fig 39n: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 71. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 39o: Expression map of Cluster 7

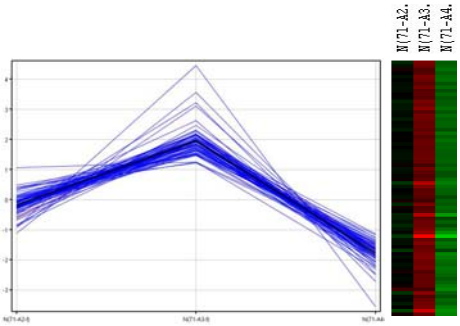

Fig 39p: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 71. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 39q: Expression map of Cluster 9

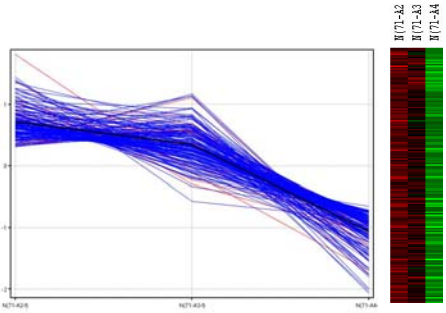

Fig 39r: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 71. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 39s: Expression map of Cluster 9

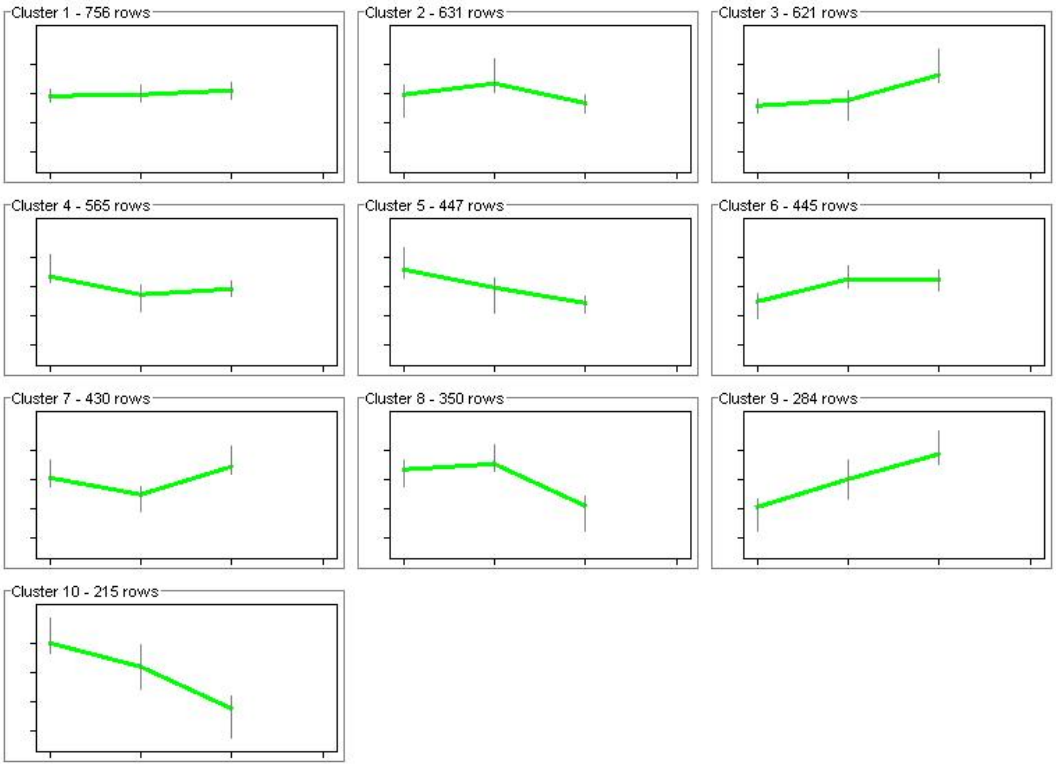

Fig 40a: K-means clustered expression graphs from leaf samples of day 75 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

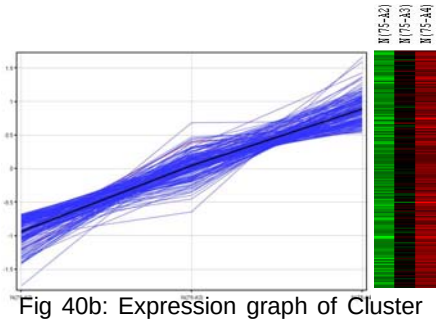

Fig 40b: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 75. Genes from rice are shown in blue and genes from pearl millet are shown in red color;
Fig 40c: Expression map of Cluster 9

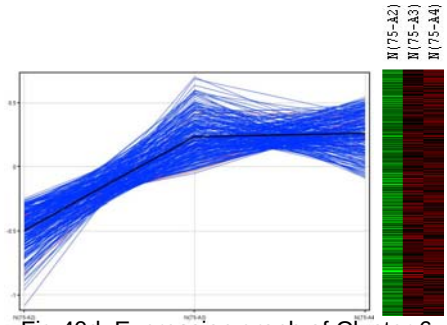

Fig 40d: Expression graph of Cluster 6 from samples A2, A3 and A4 of day 75. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 40e: Expression map of Cluster 6

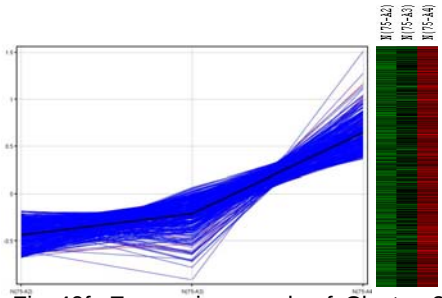

Fig 40f: Expression graph of Cluster 3 from samples A2, A3 and A4 of day 75. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 40g: Expression map of Cluster 3

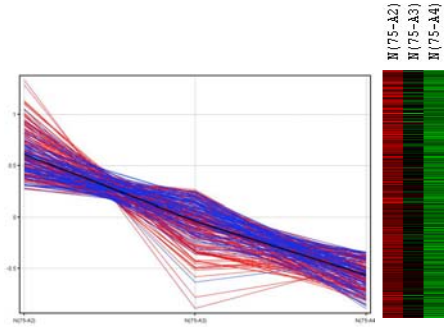

Fig 40h: Expression graph of Cluster 5 from samples A2, A3 and A4 of day 75. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 40i: Expression map of Cluster 5

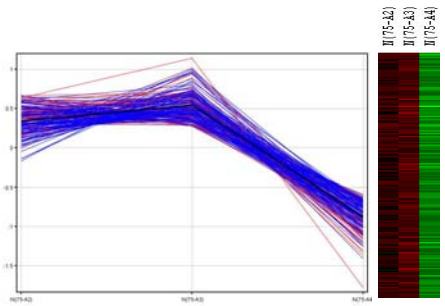

Fig 40j: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 75. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 40k: Expression map of Cluster 8

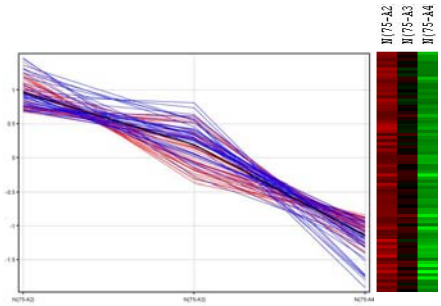

Fig 40l: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 75. Genes from rice are shown in blue and genes from pearl millet are shown in red color; Fig 40m: Expression map of Cluster 10

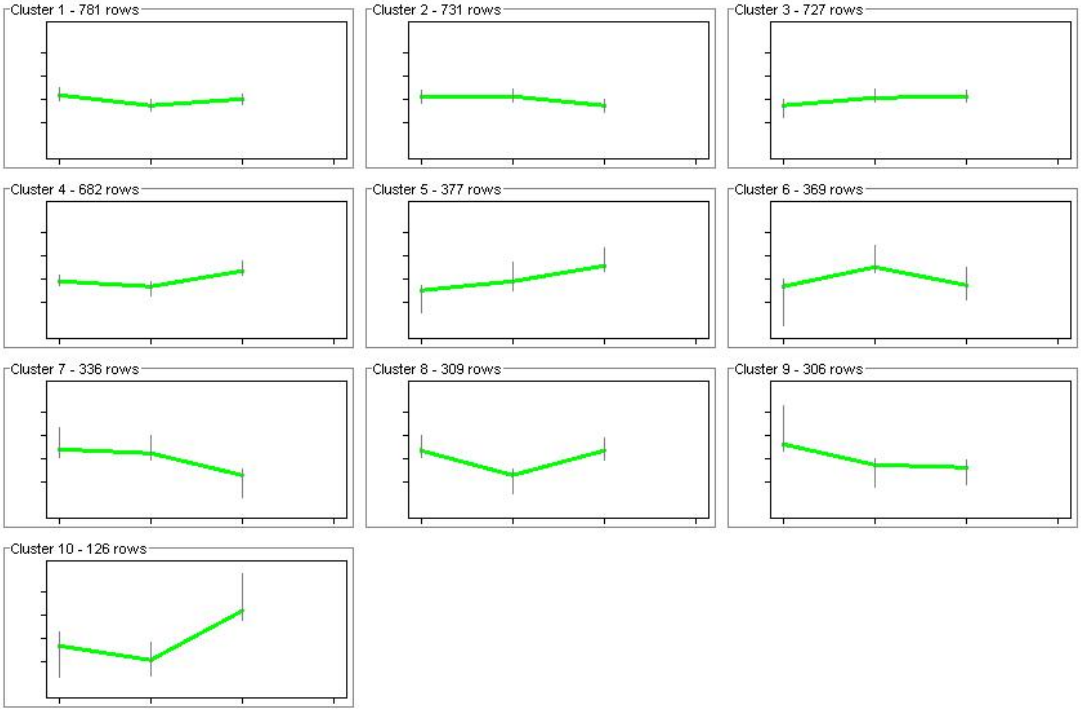

Fig 41a: K-means clustered expression graphs from panicle samples of day 59 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

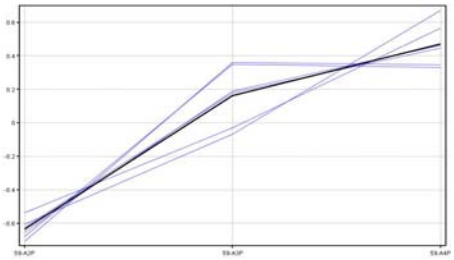

Fig 41b: Expression graph of Cluster 5 from samples A2, A3 and A4 of day 59 panicle. All the genes are from rice and shown in blue color.

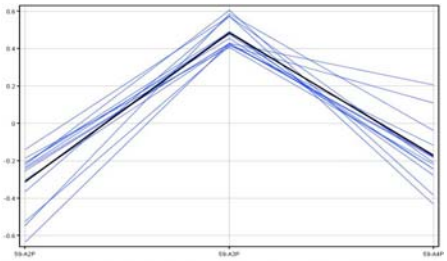

Fig 41c: Expression graph of Cluster 6 from samples A2, A3 and A4 of day 59 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

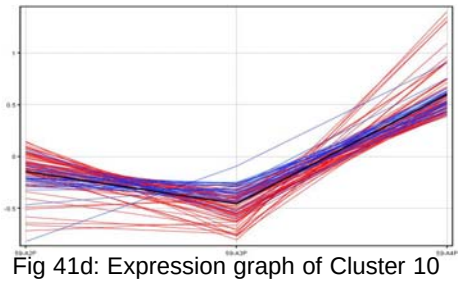

Fig 41d: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 59 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

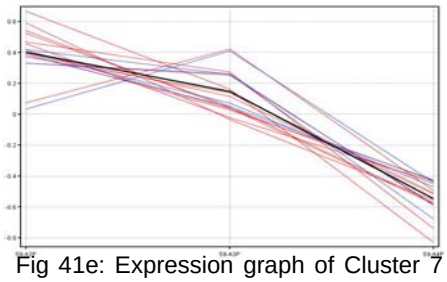

Fig 41e: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 59 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

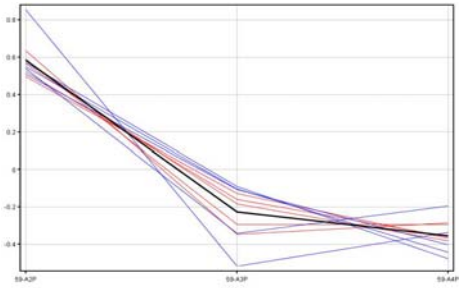

Fig 41e: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 59 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

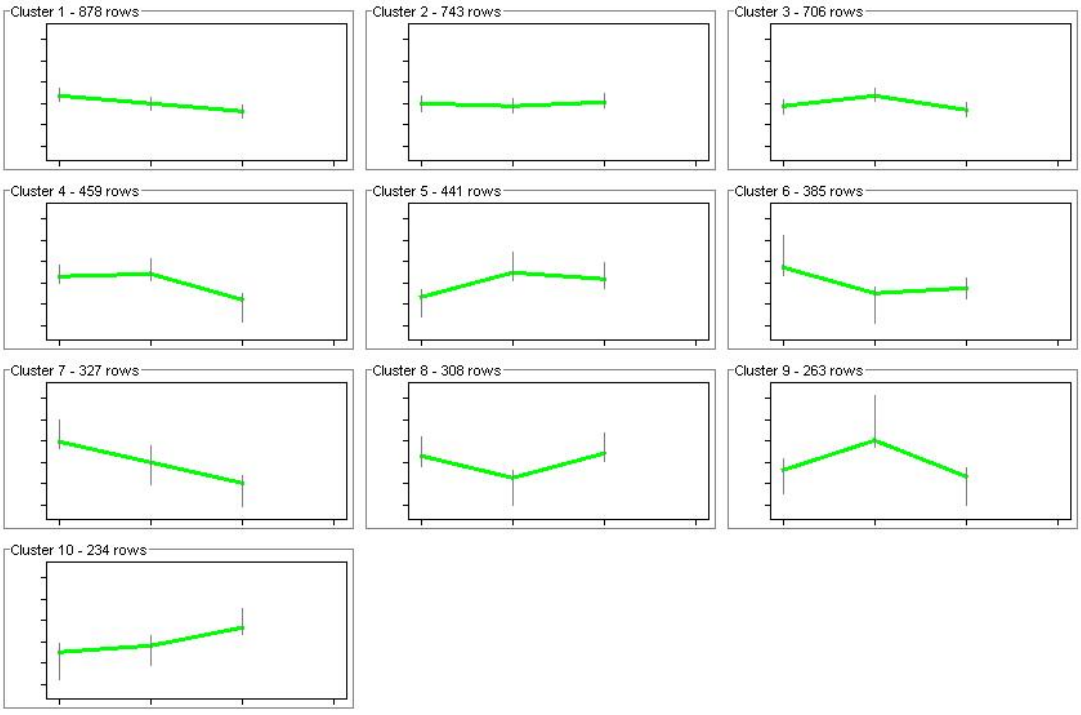

Fig 42a: K-means clustered expression graphs from panicle samples of day 63 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

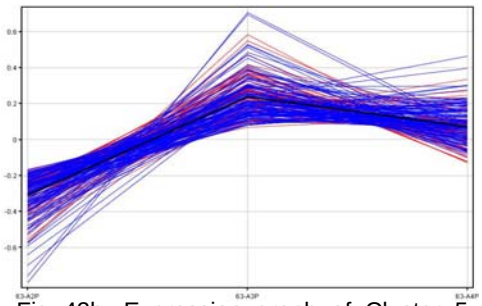

Fig 42b: Expression graph of Cluster 5 from samples A2, A3 and A4 of day 63 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

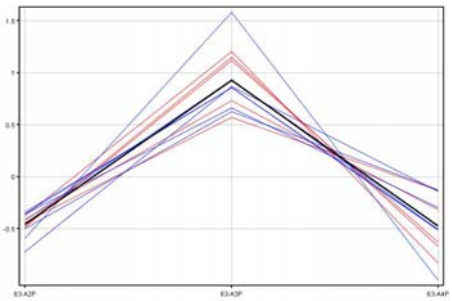

Fig 42c: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 63 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

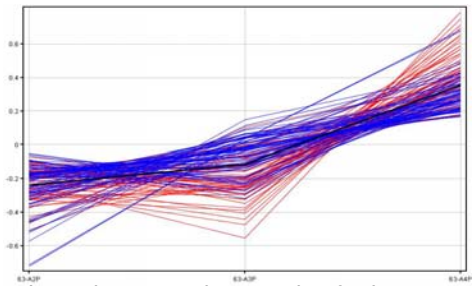

Fig 42d: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 63 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

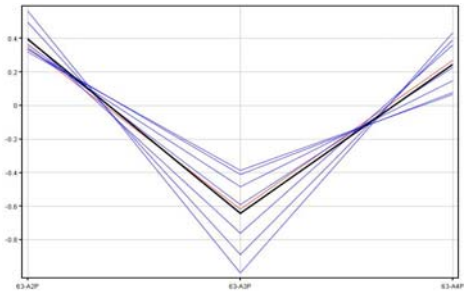

Fig 42e: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 63 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

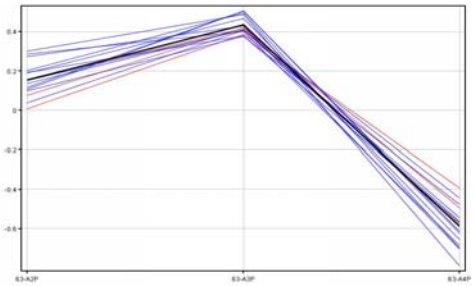

Fig 42f: Expression graph of Cluster 4 from samples A2, A3 and A4 of day 63 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

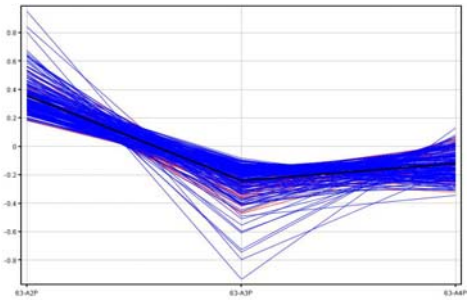

Fig 42g: Expression graph of Cluster 6 from samples A2, A3 and A4 of day 63 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

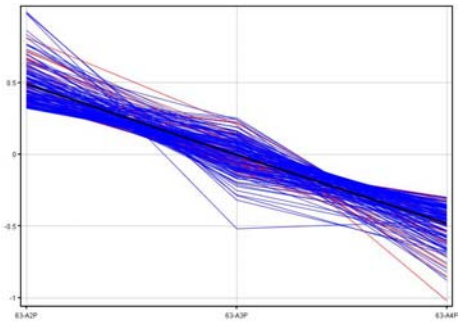

Fig 42h: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 63 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

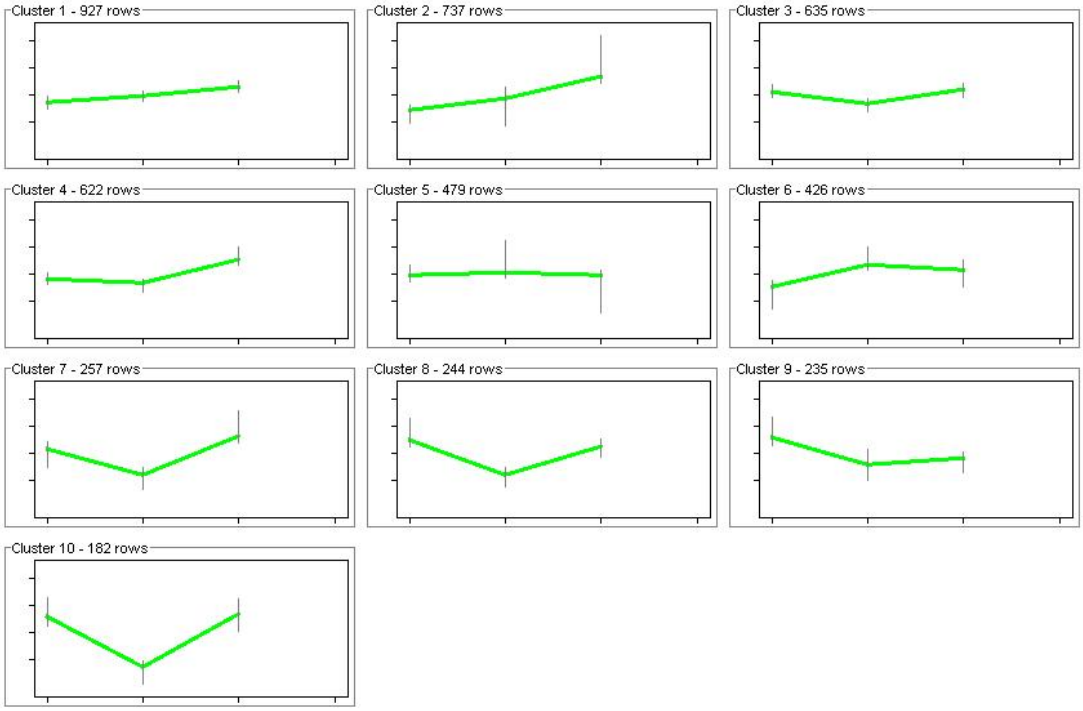

Fig 43a: K-means clustered expression graphs from panicle samples of day 67 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

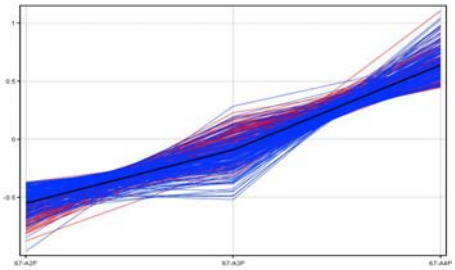

Fig 43b: Expression graph of Cluster 2 from samples A2, A3 and A4 of day 67 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

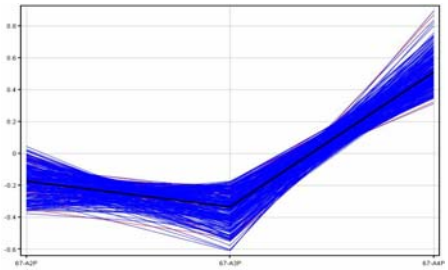

Fig 43c: Expression graph of Cluster 4 from samples A2, A3 and A4 of day 67 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

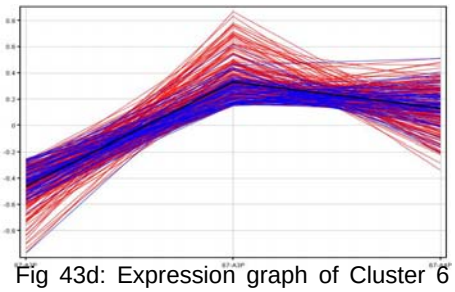

Fig 43d: Expression graph of Cluster 6 from samples A2, A3 and A4 of day 67 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

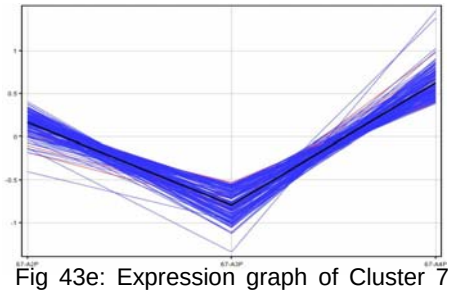

Fig 43e: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 67 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

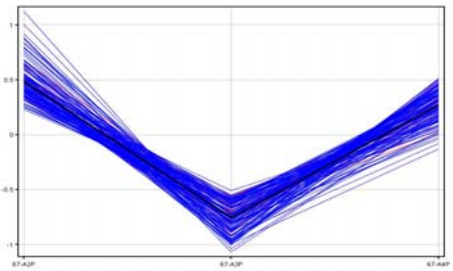

Fig 43f: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 67 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

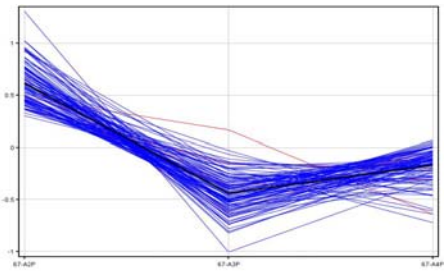

Fig 43g: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 67 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

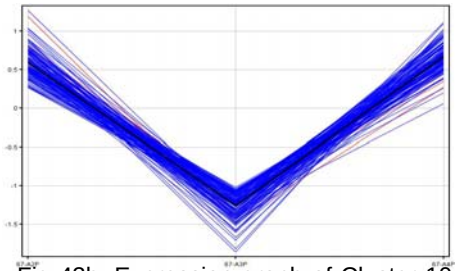

Fig 43h: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 67 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

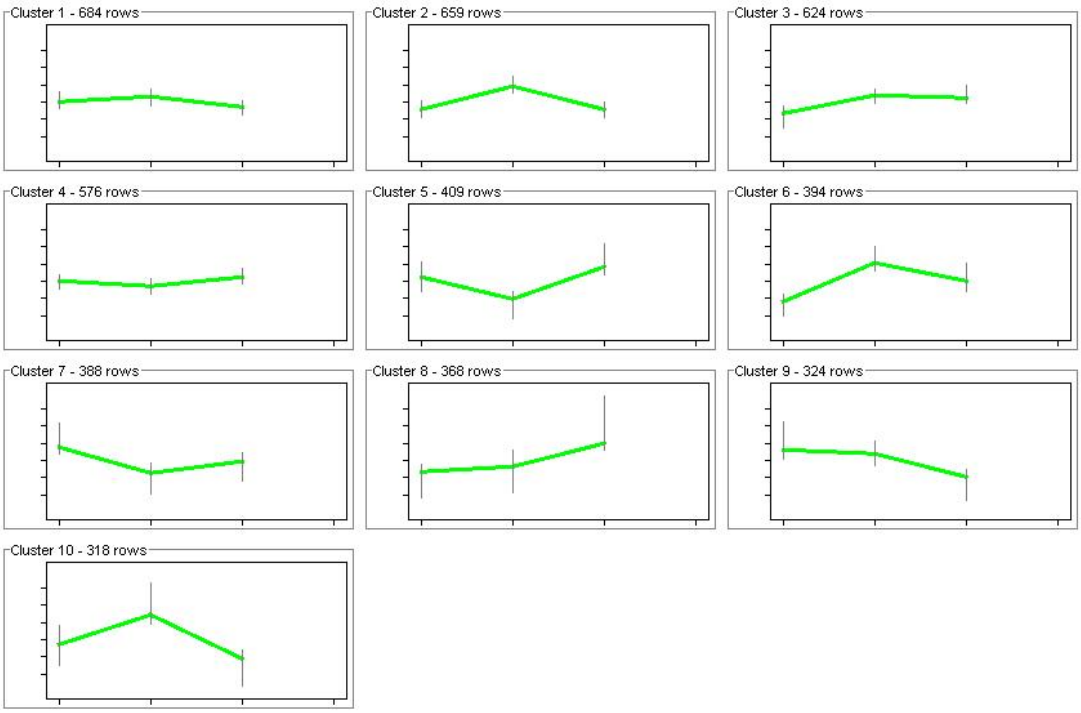

Fig 44a: K-means clustered expression graphs from panicle samples of day 71 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

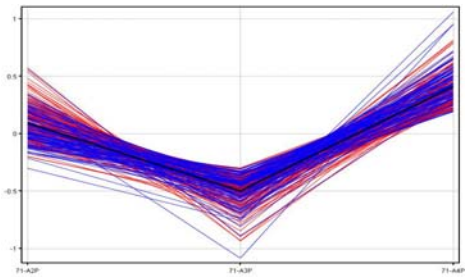

Fig 44b: Expression graph of Cluster 5 from samples A2, A3 and A4 of day 71 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

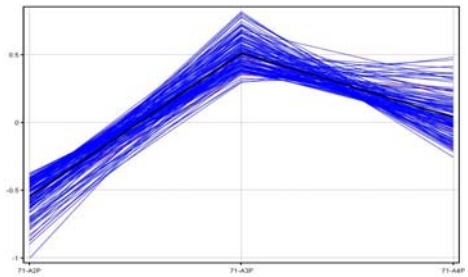

Fig 44c: Expression graph of Cluster 6 from samples A2, A3 and A4 of day 71 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

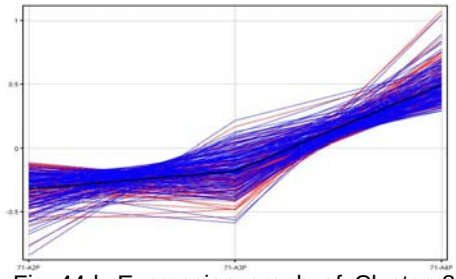

Fig 44d: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 71 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

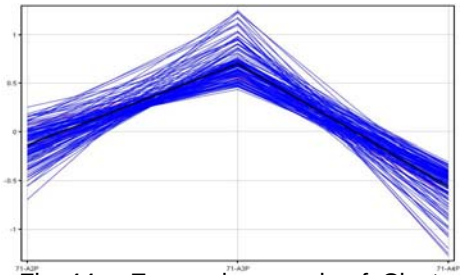

Fig 44e: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 71 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

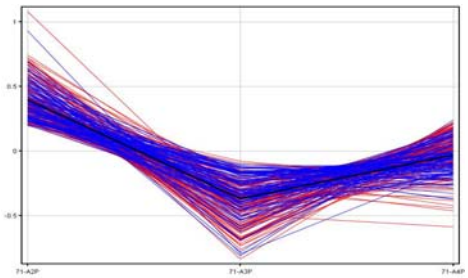

Fig 44f: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 71 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

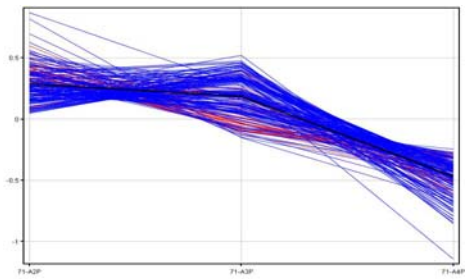

Fig 44g: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 71 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

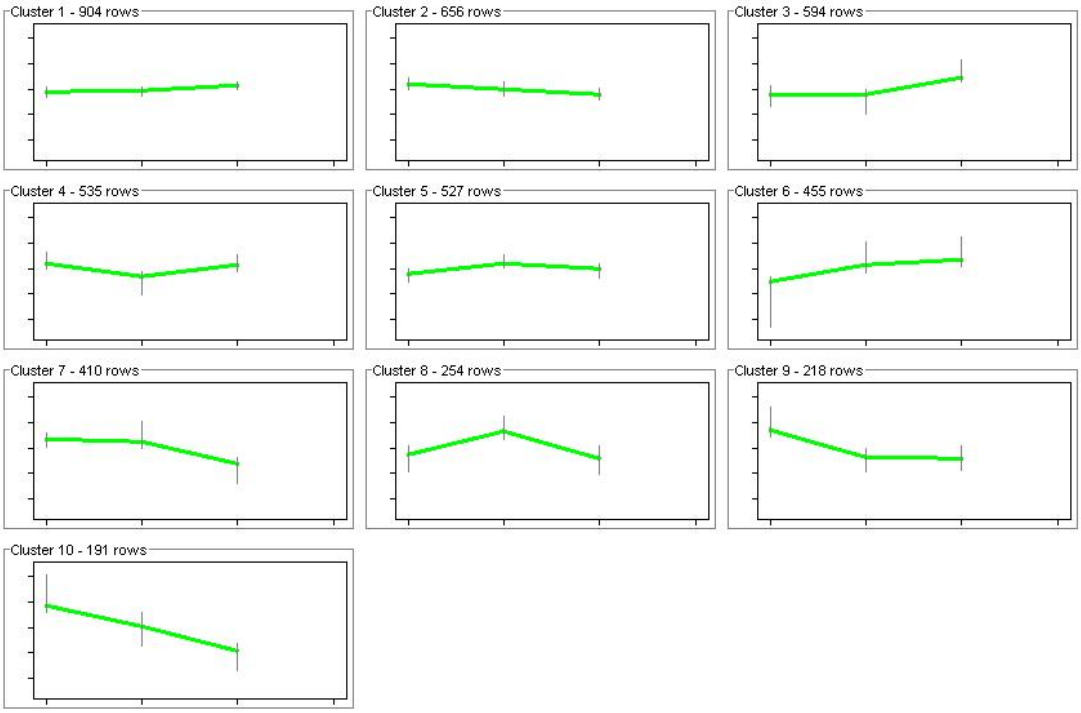

Fig 45a: K-means clustered expression graphs from panicle samples of day 75 compared among A2 (60%FC), A3 (40%FC), A4 (15%FC).

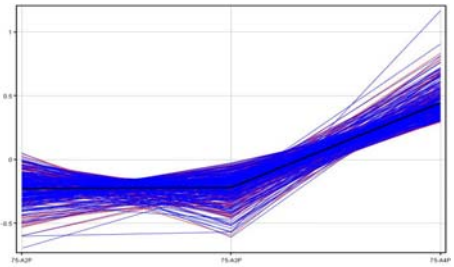

Fig 45b: Expression graph of Cluster 3 from samples A2, A3 and A4 of day 75 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

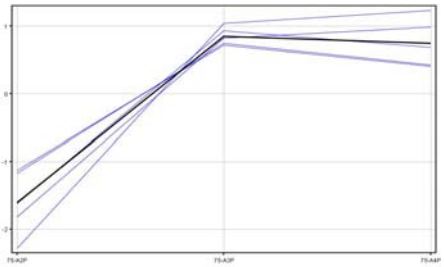

Fig 45c: Expression graph of Cluster 6_1 (sub cluster of 6) from samples A2, A3 and A4 of day 75 panicle.

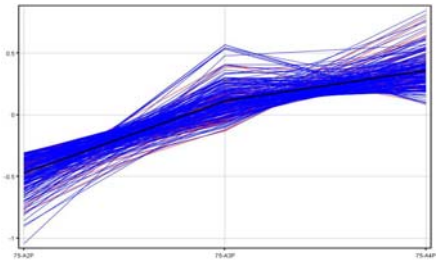

Fig 45d: Expression graph of Cluster 6_2 from samples A2, A3 and A4 of day 75 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

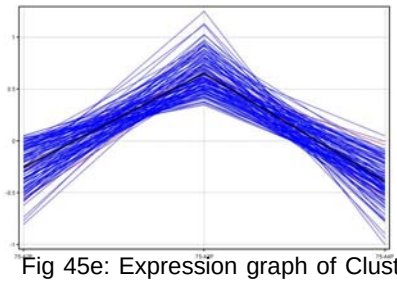

Fig 45e: Expression graph of Cluster 8 from samples A2, A3 and A4 of day 75 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

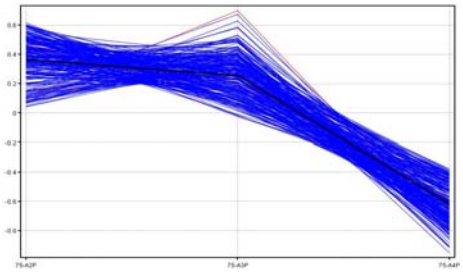

Fig 45f: Expression graph of Cluster 7 from samples A2, A3 and A4 of day 75 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

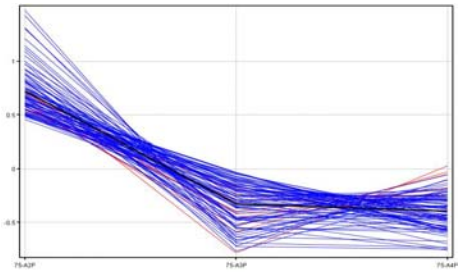

Fig 45g: Expression graph of Cluster 9 from samples A2, A3 and A4 of day 75 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

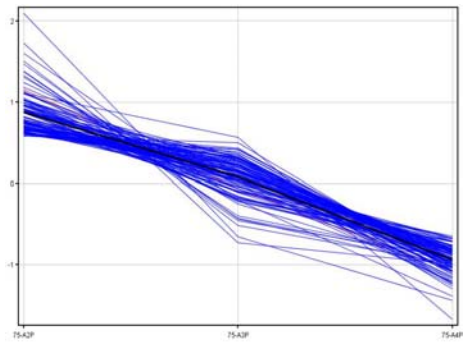

Fig 45h: Expression graph of Cluster 10 from samples A2, A3 and A4 of day 75 panicle. Genes from rice are shown in blue and genes from pearl millet are shown in red color

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

Table 3: Summary of EST generation and analysis

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

Table 4: Functional classification of N22 unigene sequences

Clone ID	Putative function	Chromosome	cM	Flanking marker	QTL acc. no.
NL_7_34	Mitogen-activated protein kinase	10	61.7–68.6	E10477S, R716	CQE77
NL_1_M20	Small GTP-binding protein (Rab5a)	12	108.2	R2292	CQE87, AQDZ005, AQCI012, CQE83
NL_14_33	14-3-3 protein homologue GF14-12	8	92.2–96.6	R2382	CQE69, CQE70, CQN37, CQN38
NL46_H02	Protein kinase	5	104.7	E31112S	AQDZ013, AQDZ002
NL_5_B21	1-Aminocyclopropane-1-carboxylate oxidase	7	55.6	C60626SB	CQG5, CQN36, CQN35, CQN47
NL_3_G15	EREBP-like protein	3	20.3	S2769	CQE15
NL_0_C09	Ethylene-responsive protein (ebp-89 gene)	3	20.3	S2769	CQE15
NRS2R_1_H13	Helicase-like transcription factor	4	109.9	R78	CQE34, CQN32
NL_3_C10	OSMYB1	5	85.7	C308	CQE41
NL_3_B07	RNAse S-like protein	9	82.4, 83.2	S10578, S955	CQE74, CQE76, CQE75
NL_1_A17	Vesicle-soluble NSF attachment protein receptor	3	65.4	C12845S	CQE18
NL_1_F05	Mitochondrial F0 ATP synthase D chain	8	92.2	S4036S	CQE69, CQE70, CQN37, CQN38
NL_1_N15	Photosystem I chain IV precursor	7	50.0	R658	AQM001, AQDN004, CQN36, CQN35

Table 5: Known stress responsive genes at QTL locations in rice

S.No.	Controls	Channel 1	Channel 2	Expected Ratio
1	cYIR01	30000	30000	1
2	cYIR02	10000	10000	1
3	cYIR03	3000	3000	1
4	cYIR04	1000	1000	1
5	cYIR05	300	300	1
6	cYIR06	100	100	1
7	cYIR07	30	30	1
8	cYIR08	10	10	1
9	cYIR09	3	3	1
10	cYIR10	1	1	1
11	rYIR1	100	300	0.333333333
12	rYIR2	300	100	3
13	rYIR3	1000	3000	0.333333333
14	rYIR4	3000	1000	3
15	rYIR5	30	300	0.1
16	rYIR6	300	30	10
17	rYIR7	1000	10000	0.1
18	rYIR8	10000	1000	10
19	uYIR1	300	300	1
20	uYIR2	300	300	1
21	uYIR3	300	300	1
22	nYIR1	0	0	0
23	nYIR2	0	0	0
24	Lambda	1000	1000	1

Table 6: Various external spike-in controls used in microarray analysis