Database Mining to Identify Salinity Tolerance Genes in Rice (*Oryza sativa,* L.).

El-Iskandarani[1], M. A., A. S. Soliman[2], H. M. El-Wakil[2], Amira M. Elmaghawry[1].

[1]Institute of Graduate Studies & Research, Alexandria University, Egypt.
[2]Faculty of Agriculture, Saba Basha, Alexandria University, Egypt.

ADDITIONAL INDEX WORDS: Data mining, Rice, salinity, Expressed sequence Tags (ESTs).

ABSTRACT

Data mining is the process of identifying patterns and relationships in data that often are not obvious in large, complex data sets. As such, DNA microarrays generate huge amounts of data of the expression profiles of thousands of genes. A variety of visualization and data mining techniques have been developed in the recent years to help the biologists find interesting relationships in microarray data. The increasing role of public databases of model organisms and bioinformatics in data mining presents a new opportunity as well as a challenge to researchers to develop more focused molecular tools for gene discovery and deployment. Abiotic stress is a major limiting factor in agricultural crop production in many countries. The major abiotic stresses of economic importance include drought, cold, heat, salinity, soil mineral deficiency, and toxicity. Salinity is usually exacerbated by intensive irrigation. World wide, crop production is affected by salinity on nearly 50% of all irrigated land. Approximately 10 million hectares are not being used for agricultural production due to salinity problems.

To limit the effect of salt in plant productivity, amelioration and utilization of salt affected soils are needed. Two approaches are used to solve the problem: by using technical approaches (water and soil management) and biological approaches. Among species, rice (*Oryza sativa*) may play a major role because of its role as the second most consumed cereal not only in the world but also in Egypt. Since the genome of rice became completely sequenced, rice is increasingly becoming the model plant for cereals. As a step towards the characterization of genes that contribute to combating salinity stress, ESTs libraries of rice (*Oryza sativa*) roots treated with 150 mM NaCl for different time series and data mining techniques have been utilized to discover genes and proteins relationship in salt tolerant public dataset and gene expression microarray datasets. This analysis has identified novel genes specifically effective against salinity; cDNA libraries that have been established for *Oryza sativa* L. 1,418 ESTs generated with RNA from stressed rice, In addition 1,106 ESTs from unstressed rice. ESTs libraries of Rice (*Oryza sativa*) roots treated with 150 mM Nacl for different time series that ranged from 1 hour, 3 hours, 6 hours, 24 hours and 1 week after salt shock.

INTRODUCTION

Computational methods with the current deluge of data have become indispensable to biological investigations. Originally developed for the analysis of biological sequences, bioinformatics now encompasses a wide range of subject areas including structural biology, genomics and

gene expression studies. Expression analysis using large numbers of genes (i.e. microarray analysis) is rapidly providing a vast amount of information. The rate-limiting step in comprehensive expression analysis is not the data-generation experiments but the processing and mining of the expression data. Furthermore, biologists use various types of data in many different types of microarray systems, making the data difficult to combine. Plant growth is greatly affected by environmental abiotic stresses such as drought, high salinity and low temperature. These abiotic stresses are severe limiting factors of plant growth and crop production. These abiotic stresses induce various biochemical and physiological responses in plants to acquire stress tolerance. The mechanism of the molecular response of higher plants against water stress has been analyzed by studying a number of genes responding to drought, high-salinity and cold stress at the transcriptional level (Bray, 1997; Hasegawa *et al*., 2000; Ingram and Bartels, 1996; Thomashow, 1999). Salinity is a major environmental threat for agricultural production that affects ionic and osmotic as well as nutritional relation of plants. It is a factor that greatly affects plant growth and development and is a major constraint for crop production. This stress is complex and causes a number of determinant effects. Among them ionic and water constraints constitute the most important. The water constraint even called osmotic pressure is characterized by difficulties to absorb water. The ionic constraint interferes with the uptake of nutrients, and causes direct toxicity due to the ions Na^+ and Cl^-. Salinity also interferes with the structure of the soil, causes an indirect stress and increases the sensitivity to diverse biotic stresses (Araya *et al*., 1991). Efficient experimental techniques, primarily DNA sequencing and microarrays, have led to an influx of basic biological data. Genome Net records show that the size of six major public databases has doubled every two years since 1982. The area of database searching has seen much growth recently, mostly focusing on providing more sensitive searches. Biological databanks have proven useful to bioscience researchers, especially in the analysis of raw data. Computational tools for sequence identification, structural analysis, and visualization have been built to access these databanks. Progress in DNA sequencing technology has produced a tremendous increase in the number of nucleotide sequences from diverse organisms in a relatively short period of time. Recently, the collections of DNA and RNA sequences submitted to public databases such as GenBank, DDBJ, and EMBL had reached 100 gigabases from 165,000 organisms. As sequencing advances, it is important to evaluate the accuracy and quality of the sequence data. Positional accuracy indicates that the sequence is mapped onto the correct position on the genome. Sequence accuracy means that the nucleotide

evaluation is performed correctly (National Library of Medicine (NLM), 2005). The accumulation of sequence data allows detailed genome analysis by using-friendly database access and information retrieval. Genetic and molecular genome colinearity allow efficient transfer of data revealing extensive conservation of genome organisation between species. The genome research goal is the identification of the sequenced genes and the deduction of their functions by metabolic analysis and reverses genetic screens of gene knockouts. This information enables new strategies to study gene expression patterns in plants. Available information from news technologies, as the database stored DNA microarray expression data, will help plant biology functional genomics. Expressed sequence tags (ESTs) also give the opportunity to perform digital northern comparison of gene expression levels providing initial clues toward unknown regulatory phenomena. Crop plant networks collections of databases and bioinformatics resources for crop plants genomics have been built to harness the extensive work in genome mapping. This resource facilitates the identification of agronomically important genes, by comparative analysis between crop plants and model species, allowing the genetic engineering of crop plants selected by the quality of the resulting products. Bioinformatics resources have evolved beyond expectation, developing new nutritional genomics biotechnology tools to genetically modify and improve food supply, for an ever-increasing world population, and to solve unanswered research questions on the mechanisms of plant development (Rudd *et al*., 2003). Plant adaptation to salinity requires alterations of various cellular physiological and metabolic mechanisms that are controlled by specific gene expression. These specific genes could encode for proteins implicated in Na^+ sequestration, (H^+-ATPases, NHX-type transporters), in synthesis of compatible osmolytes (proline, glycinebetaine, polyols), in the detoxification of toxic compounds (ROS scavenging enzymes), in signal perception and regulating factors and other unknown functions. However, fundamental importance to understand the mechanism by which rice plants take up Na^+ in cells and how they deal with excess Na^+ in the cytosol. On the other hand, in terms of adaptive responses, it is of prime importance to understand how rice cells perceive Na^+ stress and how the signal is transmitted to the downstream cellular machinery (Chinnusamy *et al*., 2005). Recent improvements in automated DNA sequencing have made whole-genome shotgun sequencing an attractive approach for gene discovery in both small and large genomes Goff *et al*., 2002 was released the draft rice sequence with efforts from Syngenta was sequenced and assembled by whole-genome shotgun sequencing. They described the random-fragment shotgun sequencing of *Oryza sativa* L. ssp.

Japonica cv. Nipponbare to discover rice genes. The assembled sequence covers 93% of the 420 megabase genome. Gu *et al*., 2005 isolated four full-length cDNAs containing complete ORF, which encode putative Ser/Thr/Tyr kinases, from rice by RT-PCR and named as OsDPK1 (*O. sativa* Dual-specificity Protein Kinase), OsDPK2, OsDPK3 and OsDPK4, respectively. These OsDPKs share high nucleotide similarity (about 70%) in coding region and similar genomic organization, each with two exons and one intron. Analysis of the deduced amino acid sequences showed that OsDPKs were significantly homologous to the protein kinase TaPK1 (about 85% identity) from *Triticum aestivum* L. and STY (83%) from *Arachis hypogaea*, and also homologous to the GmPK6 (49%) from *Glycine max* L. and ATN1 (46%) from *Arabidopsis thaliana*. RT-PCR assay revealed that the OsDPKs were tissue-specific expression at the adult stage. The OsDPK1 was only highly expressed in leaves and low in immature spikes, the OsDPK2 moderately in leaves and very low in the stems, and the OsDPK3 moderately in roots and leaves. The OsDPK4 was little detected in all detected rice tissues. The transcript levels of OsDPK1, OsDPK2, and OsDPK3 in the seedlings were significantly induced by ABA, high salinity, drought, and blast (*Magnaporthe grisea* Barr). The results suggest that these genes may belong to a novel rice gene family encoding dual-specificity protein kinases involved in the plant responses to abiotic and biotic stresses. In rice cDNA arrays comparison between the salt tolerant line Nona and the salt sensitive line IR28, Chao *et al*., 2005 showed earlier and strong expression of genes related to transcription factors and signal transduction under salt treatment. Differences could be seen in the regulation of gene correlating with salt tolerance in different species. However, many authors seem to agree to the importance of earlier perception and transmission of signals in the tolerance to stress namely salt stress. Zhou *et al*., 2005 reported that data mining the complete rice genome sequences revealed a genomic fragment encoding a characteristic metallothionein (MT) protein, and its full-length cDNA was isolated from rice developing seeds by RT-PCR. This cDNA, designated *OsMT-II-1a*, contains an open reading frame of 264 bp encoding a protein of 87 amino acid residues. The predicted amino acid sequence was shown to have structural features characteristic of plant class II MT proteins. By sequence analysis of its 50-flanking region, one putative TATA box, four putative CAAT boxes, and several short sequences homologous to previously reported regulatory cis-elements were identified. Northern blot analysis showed that accumulation of *OsMT-II-1a* mRNA is specifically abundant in developing seeds and 2-day glumes after pollination, and *OsMT-II-1a* transcription can markedly be induced by H_2O_2, paraquat, SNP, ethephon,

ABA and SA, but barely by metal ions or other exogenous abiotic factors such as low temperature and PEG. These results coincide with the prediction of existing regulatory cis-elements in its 50- flanking region. Taken together, these results suggest that the processes of pollination and seed development might be mediated, at least in part, by expression of the OsMT-II-1a gene that is regulated by several abiotic factors.

MATERIALS AND METHODS

1. Libraries from stressed plants and EST sequencing

For the time being, approximately fifty cDNA libraries have been generated and listed at www.stress-genomics.org. As a step towards the characterization of genes that contribute to combating salinity stress. Signals from triplicate spots were averaged. The Cy3/Cy5 signal intensities were adjusted with the help of exogenously added control genes, which had been placed in different sections of the microarray slides to compensate for variable background levels. All ESTs derived from root tissue. Included are the **LR** (**L**$\log_{10}$ **R**atio [Cy3/Cy5]) values, for the Cy5 and Cy3 signals of all clones on the arrays. Microarray data should always be log transformed because this transformation makes variation in signal ratios more independent of signal magnitude, reduces distribution skew, and provides a more realistic sense of data variability. Figure (1) illustrates the study strategy to analyze microarray data and identify genes response to salinity in *O. sativa*.

2. Data mining and domain analysis

To identify genes preferentially expressed in *O. sativa* at different stress times, a microarray containing a 1728 clone unique gene set was constructed. A set of samples was spotted in triplicate, as it enabled good spatial distribution without causing interference between adjacent spots.

2. 1. Experimental data

The rice genome has now been sequenced and there is a vast amount of experimental data available for analyzes. The data used in the study was obtained from www.stress-genomics.org. The gene expression data set used was an aggregation of data from several experiments on rice as a reduced dataset (Kawasaki *et al*, 2001).

On the basis of root cDNA libraries and expressed sequence tags (ESTs) from the salt-tolerant rice microarray elements were assembled to monitor changes in transcripts a period that ranged from 1 hour to 1 week after salt shock to compare the behavior of salt-tolerant rice variety with that of the salt-sensitive variety to detect, describe, and understand plant responses to high salinity. The present study discloses on the response to salt shock in

hybridization to 110 transcripts derived from the roots of stressed plants under stress (salt shock).

2.2. Similarities in gene sequence between and other plants

A BLASTX search (http://www.ncbi.nlm.nih.gov/BLAST/Genome/PlantBlast.shtml) of 110 *O. sativa* ESTs, with an expected cutoff value of ≤0.0, was conducted against the GenBank non-redundant database available through the National Center for Biotechnology

2.3. Data analysis

To determine the threshold for changes in gene expression that could be attributable to salt treatment, $\log_{10}$ of the expression ratios was used. $\log_{10}$ expression ratios (i.e. ratios bettween the normalized signal intensity of each time point for treated and control samples of each EST) which were normally distributed and centered on a ratio of 0 (i.e. the average of the expression ratios was 1). The average and the standard deviation of the expression ratios ($\log_{10}$ transformed) for each salt-treatment time point were estimated from the raw data. The software Microsoft office Excel 2003 was used to analyze the raw data.

In order to identify clusters of genes, the data set is filtered to focus on genes that are differentially expressed in different times. The Filter Data need to remove genes that do not have certain desired properties from the dataset. Data filtration can be accomplished in a small variety of ways, but as a minimum should include the omission of data from entire arrays that clearly did not yield reliable fluorescence signals and the omission of fluorescence signals that are 'flagged' by image analysis software. After determination of differentially expressed gene set, K- means and Hierarchical methods were applied on the data to produce clusters of genes with similar expression patterns.

2.4. Cluster analysis

Clustering means the partitioning of data into subsets (clusters), so that each element of the subset shares a common feature. Clustering methods allow visual insight into the data and can be used for class discovery.

Unsupervised classification of stress- response genes was based on log transformed ratios of all up or down regulated genes detected in salt treatment. Hierarchical clustering analysis performed to display the expression pattern and tree diagram at different stress times by employing Cluster and Tree View. However, the object of clustering is to identify genes that respond the same way to the environmental treatment. Each gene is compared to every other gene and a gene similarity score (metric) is produced. If X_i is the log odds value for gene X at time I, then for two genes X and Y and N observations, a similarity score is calculated (Eisen *et*

al., 1998). The computer program that performs the cluster analysis is Cluster 3.0 (De Hoon *et al*., 2002).

RESULTS AND DISCUSSION

Modern data gathering techniques such as those used for genomic mapping of biological sequences are producing vast amounts of data. However, data can be useless in the absence of understanding. The work described here provides a generalized methodology that can address problems in bioinformatics. Preliminary results are promising; Digital expression analysis combined with gene annotation helped in the identification of several pathways associated with salt stress. The genomic resources and knowledge developed by this study will contribute to a better understanding of the different mechanisms that govern stress tolerance in rice and other cereals.

1. Database mining of *Oryza sativa* L. ESTs subjected to salt stress

The information provided by ESTs of randomly isolated gene transcripts generated under specific abiotic stress conditions provides an opportunity for gene discovery in addition to identifying the biochemical pathways involved in plant physiological responses. The study present an initial set of data from database mining that is focusing on defining the complete gene set associated with salinity stress.

1.1. Database searching for ESTs

Expressed Sequence Tags (ESTs) are partial, single- pass sequence from randomly selected cDNA clones. ESTs in public databases are a useful tool for identifying specific and novel genes. Among the cDNA libraries that have been established for *Oryza sativa* L. 1,418 ESTs generated with RNA from salt- stressed rice at different times during stress. In addition 1,106 ESTs from unstressed rice were found and analyzed into functional categories.

1.2. Comparison of *Oryza sativa* ESTs with those in Genbank

BLASTX research was utilized to investigate the similarity in the EST sequences of *Oryza* to GenBank non-redundant database.

NCBI produces a portable database containing assembled nucleotide sequences, the BLASTX results of the closest homologies and provides an assessment of portable function based on automated functional assignment algorithms. This functional assignment algorithm with NCBI utilizing the information generated by BLASTX homology analysis is for estimation of function using Geramene database, providing a preliminary metabolic

overview for raw ESTs sequences data found. The criteria used to select the assembled sequences were the E value. Assembled sequences with an E value ≤ 0.0 were considered to represent putative salt- inducible gene homologs.

1.3. Functional categorization of rice ESTs and putative stress-regulated genes

The ESTs were searched against the non-redundant database using BLASTX, and the database matched ESTs were further grouped according to their putative function. The expression of genes involved in protein synthesis, protein destination and cell defense are strongly affected by salt stress. In contrast, the cell growth genes and metabolism genes respond weakly to salt stress at certain time points. Gene functional groups showed different responses to salt stress. These genes cover a broad range of the functional categories. However, due to the lack of gene products information, many transcripts cannot be functionally categorized.

To compare between transcripts from unstressed and salt- stressed rice roots under salt stress treatment conditions with sequences in public databases were done. Data from libraries of salt-stressed roots demonstrated that gene expression profiles of *O. sativa* changed significantly at different stress time points. The results indicate a progression of regulated functions such that different categories of transcripts show regulation at different time points.

Table (1) and figures (2& 3) provide a transcript abundance profile for roots from salt- stressed and unstressed rice libraries annotated in general functional categories.

The annotation process assembled approximately 40.5% of the sequences were categorized as unknown proteins, including those in the "no hits" category from library of unstressed rice roots. This number was increased to be 47.25% in libraries from stressed plants. In the category no hits, 274 ESTs were found in the non-redundant database, 180 ESTs from stressed plants and only 94 from unstressed library did not show significant homology with ESTs in the databases.

These unknown and no hits genes are likely the source of candidate salt-tolerant genes and further functional analysis will help elucidate their specific roles in salt tolerance. A large proportion of ESTs was found to participate in the biological process as a function of stress, a precipitous decrease in the protein synthesis (22.9%) in unstressed plants to (10.7%) in stressed plants. A large proportion of genes was found to participate in the biological process of metabolism (17%) in stressed plants.

In contrast, other major distinctions in transcript composition between the unstressed and stressed library ESTs included, transcripts for the transport

facilitation and cell defense categories. The accumulation of osmoprotectants by either altering metabolism or increasing metal ion homeostasis is an important process of plants for the adaptation to environmental stress. The increase in cell rescue, defense and aging proteins could be essential as a strategy the plants adopted to cope with the changed environment. Results showed that 6.35 % of the categories respond to environmental stimuli. This category from the basis for mining the stress regulated genes. Late embryogenesis abundant (LEA) OE06D06 protein and other induced proteins were found in this category. There were stress induced genes (stress- induced protein oz11 precursor) OE06B01 and (Asr1) OC03F06. These were also other genes function as scavenges of reactive oxygen species, such as glutalthione s- transferase (OC01B11), these gene products are needed to maintain the redox homeostasis under abiotic stress. A plethora of physiological and metabolic adjustments occur during salinity acclimation and in response to other stresses. The regulation of genes involved in temperature, drought and salt stresses is known to reflect the cross-talk between different signaling pathways (Seki *et al*.,2002). However, few studies have identified multiple genes that are stress-regulated and that belong to a same metabolic pathway. Our analyses enabled us to position several genes in their respective metabolic pathway, suggesting that these pathways are involved in stress responses.

It was reported that overexpression of H_2O_2 scavenging enzymes increased the tolerance of plants to abiotic stress (Yuasa *et al*., 2001), there are also found in the EST sequenced (OC103G02, OC04D05, OC102E03, and OC02F04). Metallothioneins (MT) are a group of low- molecular weight (LMW) metal- binding proteins with a high cysteine content that are thought to be involved in metal ion metabolism and detoxification; which protect cells from superoxide radicals by catalyzing the dismutation of the superoxide radical to molecular O_2 and H_2O_2 (Wu *et al*., 1997).

In plant cells, calcium functions as a second messenger coupling a wide range of extra cellular stimuli to intra cellular responses (Saijo *et al*., 2000). Calmodulin, one major class of Ca^{2+} sensor characterized in plant, which was present in the *oryza sativa* ESTs (OC103B03), is involved in stress signal transduction suggested by several lines of evidence (Chinnusamy and Zhu, 2003). Genes for transcription factors that contain typical DNA binding motifs, such as GTP; have been demonstrated to be stress inducible. Transcription factors containing similar domains are present in the Rice EST (OC101F12) and may be important in regulating the response to salt stress.

1.4. Microarray analysis for salt- tolerant and salt- sensitive rice

Unless all genes from an organism are available, microarray analyses will remain incomplete, but even partial sets of arrayed ESTs, transcripts or genes can provide useful information. EST databases have proven to be a tremendous resource for finding genes and for interspecies sequence comparison, and have provided markers for genetic and physical mapping and clones for expression analyses. The relative abundance of ESTs in libraries prepared from different organs and plants in different physiological conditions also provides preliminary information on expression patterns for the more abundant transcripts (Bouchez and HÖfte 1998).

The type of information gained by microarray analysis of rice in which ESTs from the collection of salt-tolerant rice were spotted together with ESTs from salt-sensitive plants. The resulting microarrays comprised 1728 transcripts spotted in triplicate. The changes in salt- tolerant and salt-sensitive rice were described at various time points during salt stress: 1 hr, 3 hr, 6 hr, 24 hr and 7 days. Changes in signal intensity were converted into $\log_{10}$ expression ratios, **LR** (**L**$\log_{10}$ **R**atio [Cy3/Cy5]), for the Cy5 and Cy3 signals of all clones on the arrays. From repeat experiments, log-10 +/-0.2 (1.6-fold upregulation or downregulation) considered as the threshold that to signified a significant change. In this data 48 ESTs from the salt- tolerant rice libraries were upregulated (48/1728= 2.8%), while 24 ESTs from the salt-sensitive plants were upregulated (24/ 1728= 1.4%). A similar result characterized repressed cDNAs: 49 ESTs from the salt- tolerant rice cDNAs libraries were downregulated (49/1728= 2.8%) and 24 ESTs from salt- sensitive rice were strongly downregulated (24/ 1728= 1.4%). Results focused on the response to salt shock in rice in hybridization to 110 transcripts derived from the roots of salt stressed plants (Table. 2)

2. Generation of ESTs from rice subjected to salt stress

In order to identify clusters of genes, the data set is filtered to focus on genes that are differentially expressed in different times resulting 110 ESTs exhibited.

Selected ESTs were sequenced and generated for further compare gene expression under salt stress, all regulated genes were selected for cluster analysis.

2.1. Nucleotide sequence statistics

The information provided by ESTs of randomly isolated gene transcripts generated under specific abiotic stress conditions provides an opportunity for gene discovery in addition to identifying the biochemical pathways involved in plant physiological responses. This study describes ESTs obtained from salinity- induced cDNA libraries prepared from the

roots of the *Oryza sativa* exposed to stress from 1 hour to 7 days. The average read- length of cleaned ESTs was found to be 530.70 bp. The average G+C content of *Oryza* ESTs was proven to be 48.94%.

3. Clustering gene expression data:

Cluster analysis is the art of finding groups in a given data set such that objects in the same group are similar to each other while objects in different groups are as dissimilar as possible. Because of the large number of genes and the complexity of biological networks, clustering is a useful exploratory technique for analysis of gene expression data. By clustering gene expression profiles or whole cDNA library profiles, Ewing *et al*., 1999 showed that genes with similar functions, or cDNA libraries expected to share patterns of gene expression, are grouped together.

3.1. K-means clustering

The 110 ESTs were assembled into 10 clusters using Cluster 3.0 program. The frequency of EST distribution after clustering is shown in figure (4).

Analysis of the obtained data revealed that five clusters had 11 or more ESTs, with the largest one containing 20 ESTs. Most clusters contained six to eight ESTs.

Cluster I contained six transcripts. Among these genes encoding several glycine/ serine rich proteins were the most highly upregulated transcripts. Two several glycine/ serine rich proteins isoforms were found in this cluster (OC101E09 and OC103E04).

The second cluster consisted of twelve ESTs. Among them were an EST encoding polyubiquitin 6 (OE201B05) involved in protein destination and an EST encoding receptor protein-like (OC103D06) functioning as transcription signaling.

The third cluster which include seven transcripts; the annotation for some of those ESTs suggested that they include various stress (drought, high-salinity, cold, and ABA) responsive genes (cytochrome b5 (OE202C08), heat shock protein 70 (OC103F07), Nramp1 protein (OCP06F06), nucleoside diphosphate kinase (OC05H08), putative glycosylation enzyme (OE05E02), gibberellin-induced receptor-like kinase (OE04A04), and water channel protein (WCP-IV) (OE202F11).

Finally the last cluster contained twelve ESTs in which they represent Osr40g2 (OE201G12) protein induced by salt stress and abscisic acid and H^+-transporting ATPase (OD102G11) involved in transport facilitation.

Genes of unknown function that belong to a cluster of genes displaying similar expression over time, or tissue types, may be assigned the function of the most prevalent gene function in that cluster (provided the cluster has genes with known function as members). In this way functional information

may be transferred between genes with little or no sequence similarity (Baldi & Brunak, 2001).

3.2 Hierarchical clustering

Hierarchical clustering has been used since the very beginning of the microarray field. Hierarchical clustering aims at the more ambitious task of providing the definitive clustering that characterizes a set of patterns in the context of given distance metric. The result of K- means clustering is a set of K clusters, as well as all elements of a given cluster are on the same level, no particular inferences can be made about the relationship between members of a given cluster or between clusters. In contrast, the results of a hierarchical clustering are a complete tree with individual patterns (genes or experiments) as leaves and root as convergence point of all branches. The diagrams produced by the hierarchical clustering are also known as dendrograms. A dendrogram is a branching diagram representing a hierarchy of categories based degree of similarity. Different genes and/ or experiments are grouped together in clusters using the distance chosen. Different clusters are also linked together based on a cluster distance such as the average distance between all pairs of objects in the clusters. A combined dendrogram with gene clustering plotted vertically and experiment clustering plotted horizontally is presented in figure (5).

Thus Hierarchical clustering is widely accepted for the analysis of microarray data. Genes and experiments are often displayed as a color-coded table of hierarchically clustered rows and columns (Spellman *et al*., 1998).

The genomic-scale EST; genome sequencing; and cDNA microarray analyses that are now under way promise to rapidly isolate and identify all candidate genes of the gene complement essential for tolerance of salt stress. The large datasets generated by these efforts will be integrated and comparisons made between different cellular plant models to identify the cellular tolerance mechanisms that are evolutionarily conserved. Mining of these data will supply a systematic agenda for functional analysis with the use of tagged mutant collections, complementation and overexpression tests accompanied by microarray analyses to reveal hierarchical relationships between specific signaling components and downstream effector genes.

The sensing of Na^+ and the subsequent signal transduction to switch on adaptive responses are critical steps for plants exposed to NaCl dominated salt environments. Under high salinity both osmotic stress and ionic stress are perceived at the cellular level. Until now, however, very little information is known about the sensors in plants for ionic toxicity and osmotic stress under high salt, although several osmosensors are reported to be involved

in plant signal perception at osmotic stress (Urao *et al*., 1999; Reiser *et al*., 2003; Tamura *et al*., 2003). The information provided by data mining searching databases and analysis were used to attempt to draw a model for rice responses to salt (Figure 6).

Molecular genetics; and cell biological approaches to identify signaling components and biochemical characterization of signaling complexes will be required to further understand salt-stress signaling pathways and their use in crop improvement. All of these data shown in this study would provide a platform to delineate a conceptual framework for cytosolic Na^+ homeostasis and salt tolerance in rice. This would eventually be an important ground for genetic engineering to develop salt-tolerant high yielding rice cultivars.

Understanding of the molecular genetics and physiology of the traits conferring salinity tolerance will form the basis for further improvements in the salinity tolerance of agricultural species. Knowledge gained from rice data will not only be used in applied rice but might also serve for basic discovery that can be transferred to other plants. Armed with such information from established models, it will be possible to rationally manipulate and optimize tolerance traits for improved crop productivity well into the twenty-first century.

REFERENCES

Araya, F., O. Abarca, G. Zúñiga and J. Corcuera. 1991. Effects of NaCl on glycine betaine and on aphids in cereal seedlings. Phytochemistry. 30: 1793–1795.

Baldi, P. and S. Brunak. 2001. Bioinformatics: the machine learning approach, 2nd ed. The MIT Press, Massachusetts London, England.

Bouchez, D. and H. HÖfte. 1998. Functional Genomics in Plants. Plant Physiol. 118: 725-732.

Bray, E. A. 1997. Plant responses to water deceit. Trends Plant Sci. 2: 48-54.

Chao, D., H. Luo, M. Shi, D. Luo and L. Xuan. 2005. Salt-responsive genes in rice revealed by cDNA microarray analysis. Cell Research. 10: 796-810.

Chinnusamy, V. and J. Zhu. 2003. Plant salt tolerance. Topics Curr. Genet. 4: 241-270.

Chinnusamy, V., A. Jagendorf and J. Zhu. 2005. Understanding and improving salt tolerance in plants. Crop Sci. 45:437– 448.

De Hoon, M., S. Imoto and S. Miyano. 2002. Statistical analysis of a small set of time-ordered gene expression data using linear splines. Bioinformatics.18: 1477- 1485.

Eisen, M., P. Spellman, P. Brown and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. 95(25): 14863-14868.

Ewing, R., O. Poirot and J. Claverie. 1999. Comparative analysis of the Arabidopsis and rice expressed sequence tag (EST) sets. In Silico Biol. 1: 18. [http://www.bioinfo.de/isb/1999/o1/0018].

Goff, S., D. Ricke, T. Lan, G. Presting, R. Wang and M. Dunn. 2002. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). Science. 296: 92- 100.

Gu, Z., J. Wang, J. Huang and H. Zhang. 2005. Cloning and characterization of a novel rice gene family encoding putative dual-specificity protein kinases involved in plant responses to abiotic and biotic stresses. Plant Science. 169: 470- 477.

Hasegawa, M., R. Bressan, J. Zhu and H. Bohnert. 2000. Plant cellular and molecular responses to high salinity. Annu. Rev. Plant Physiol. Plant Mol. Biol. 51: 463- 499.

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

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

National Library of Medicine (NLM). 2005. Public Collections of DNA and RNA Sequence Reach 100 Gigabases. Press Release, [http://www.nlm.nih.gov/news/press_releases/dna_rna_100_gig.html].

Reiser, V., D. Raitt and H. Saito. 2003. Yeast osmosensor Slnl and plant cytokinin receptor Crel respond to changes in turgor pressure. J. Cell. Biol. 161: 1035-1040.

Rudd, S., H. Mewes and K. Mayer. 2003. Sputnik: a database platform for comparative plant genomics. Nucleic Acids Research. 31(1): 128- 132.

Saijo, Y., S. Hata, J. Kyozuka, K. Shimamoto and K. Izui. 2000. Overexpression of a single Ca^{2+} dependent protein kinase confers both cold and salt/drought tolerance on rice plants. Plant J. 23: 319- 327.

Seki, M., M. Narusaka, J. Ishida, T. Nanjo, M. Fujita, Y. Oono, A. Kamiya, M. Nakajima, A. Enju, T. Sakurai, M. Satou, K. Akiyama, T. Taji, K. Yamaguchi-Shinozaki, P. Carninci, J. Kawai, Y. Hayashizaki and K. Shinozaki. 2002. Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J. 31: 279-292.

Spellman, T., G. Sherlock, M. Zhang, V. Iyer, K. Anders, M. Eisen, P. Brown, D. Botstein and B. Futcher. 1998. Comprehensive

identification of cell cycle regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol. Biol. Cell. 9: 3273- 3297.

Tamura, T., K. Hra, Y. Yamaguchi, N. Koizumi and H. Sano. 2003. Osmotic stress tolerance of transgenic tobacco expressing a gene encoding a membrane-located receptor-like protein from tobacco plants. Plant Physiot. 131: 454- 62.

Thomashow, F. 1999. Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annu. Rev. Plant Physiol. Plant Mol. Biol. 50: 571- 599.

Urao, T., B. Yakubov, R. Satoh, K. Yamaguchi-Shinozaki, M. Seki, T. Hirayama and K. Shinozaki. 1999. A transmembrane hybridstresses type histidine kinase in *Arabidopsis* functions as an osmosensor. Plant Cell. 11:1743-1754.

Wu, Y. J. Kuzma, E. Marechal, R. Graeff, C. Lee, R. Foster and N. Chua. 1997. Abscisic acid signalling through cyclic ADPribose in plants. Science. 278: 2126- 2130.

Yuasa, T., K. Ichimura, T. Mizoguchi and K. Shinozaki. 2001. Oxidative stress activates ATMPK6, an *Arabidopsis* homologue of MAP kinase. Plant Cell Physiol. 42: 1012-1016.

Zhou, G., Y. Xu and J. Liu. 2005. Characterization of a rice class II metallothionein gene: Tissue expression patterns and induction in response to abiotic factors. Journal of Plant Physiology. 162: 686- 696.

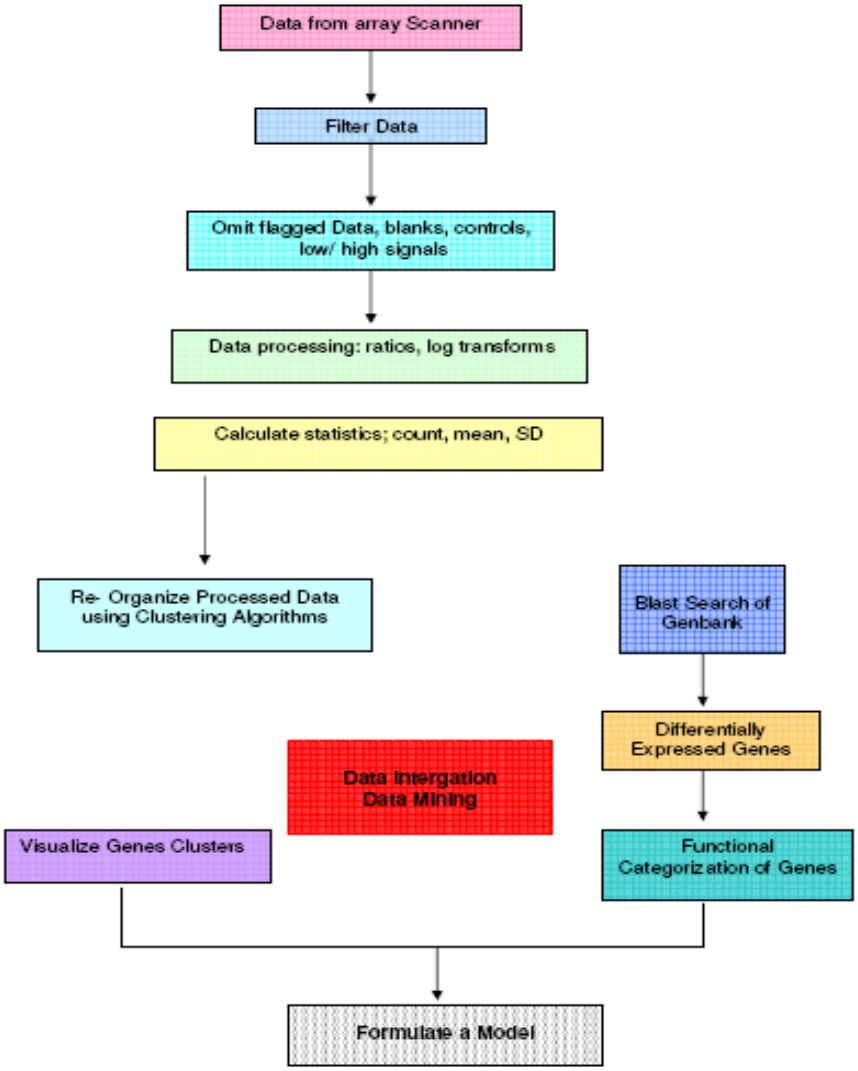

Figure 1. Strategy to analyze microarray data and identify genes response to salinity in *O. sativa*.

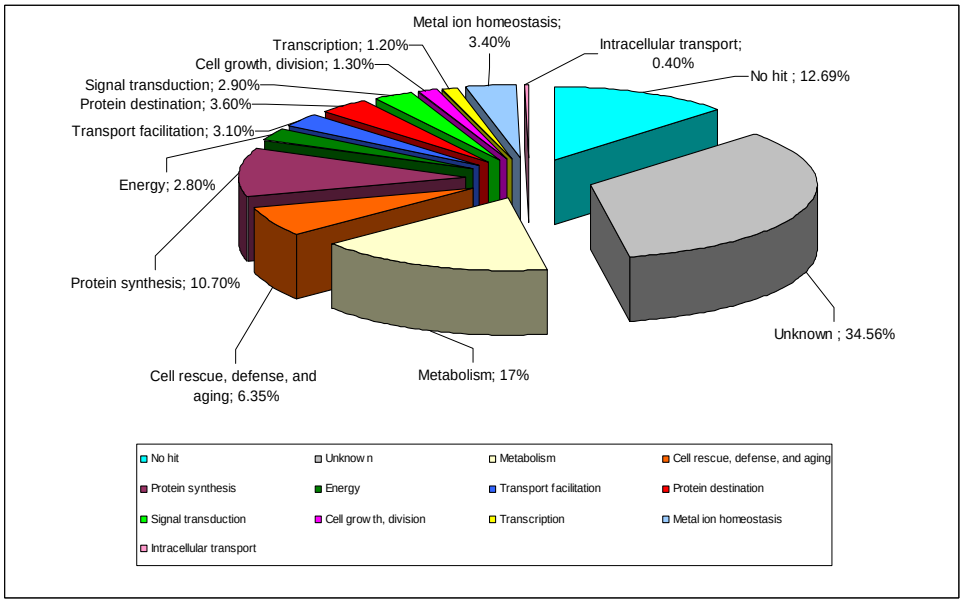

Figure 2. Chart of ESTs categorized by function. A total of 1418 ESTs from salt- stressed rice roots cDNA libraries.

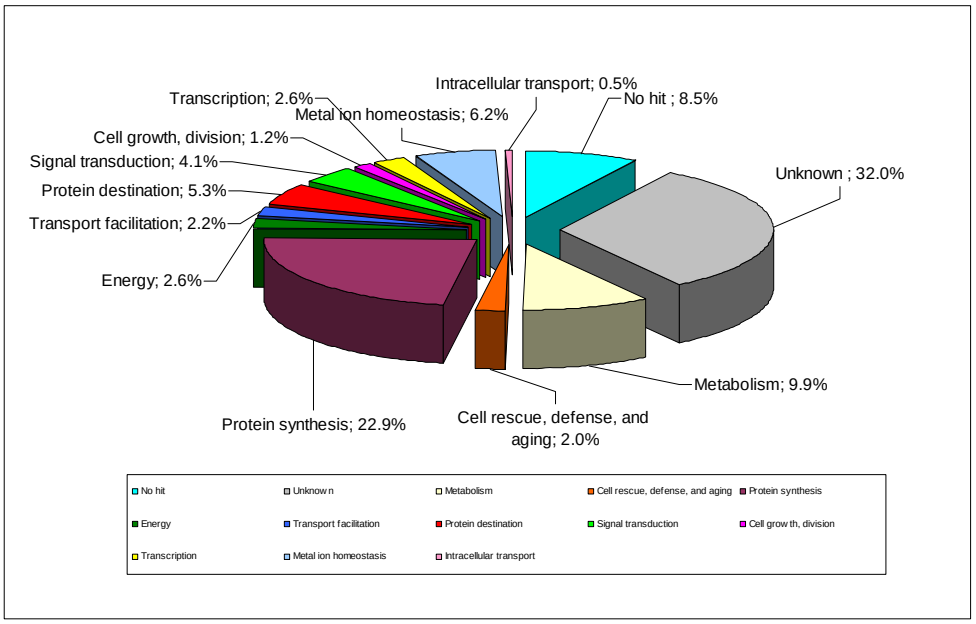

Figure 3. Chart of ESTs categorized by function. A total of 1106 ESTs from unstressed rice roots cDNA library.

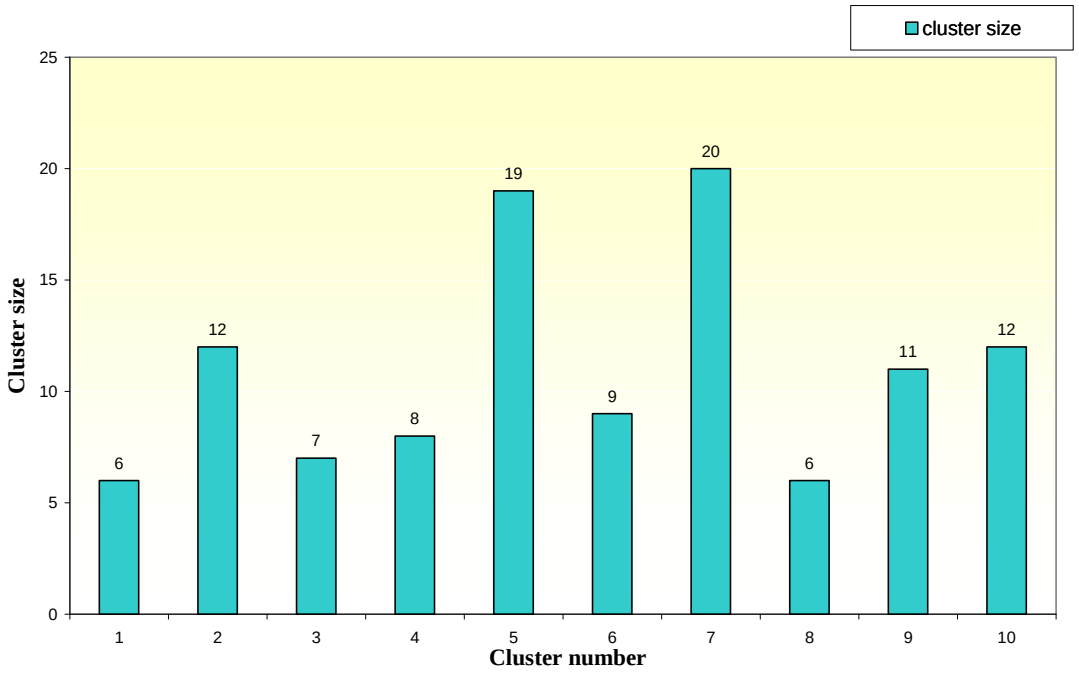

Figure 4: Distribution and number of clustered sequences

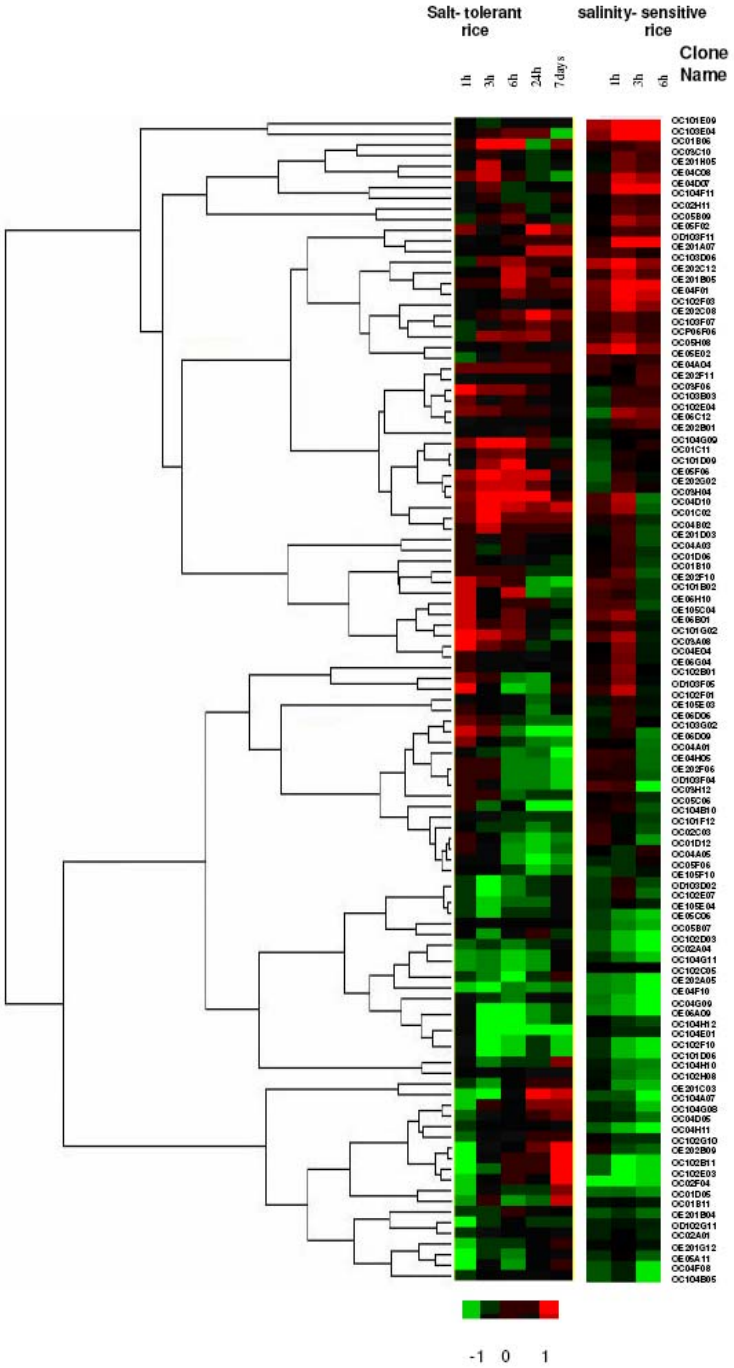

Figure 5. Cluster analysis of all responsive genes in each of the salt-tolerant and salinity- sensitive rice

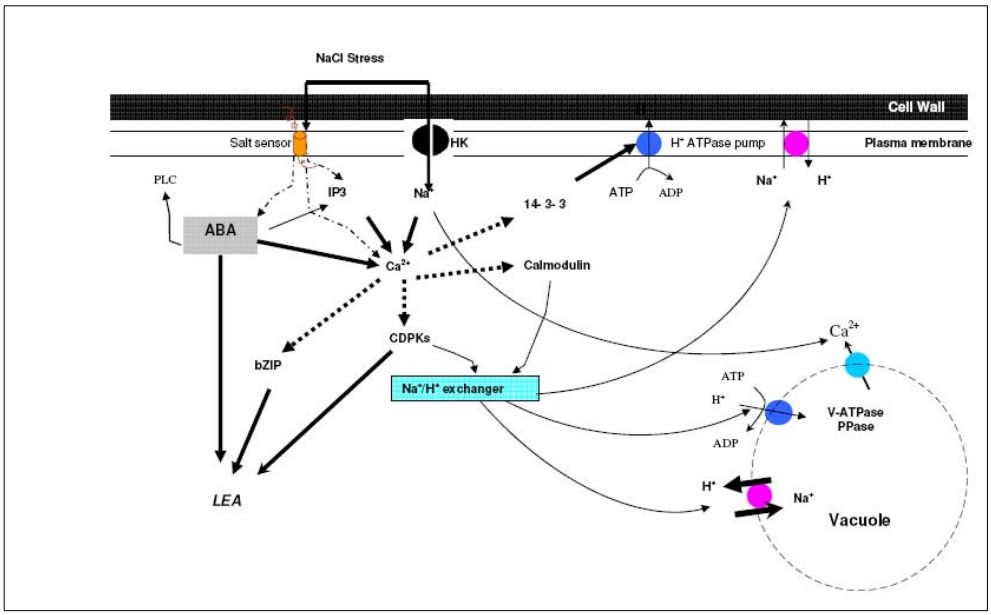

Figure 6. Hypothetical model of the rice responses to salinity stress.

Table 1. Abundance profile of ESTs in root cDNA libraries of salt- stressed and unstressed rice.

Gene category	salt- stressed rice		unstressed rice		Category description
	No.	**%**	**No.**	**%**	
No hit	180	12.69	94	8.5	Blast search in non-redundant GenBank finds no homologs.
Unknown	490	34.56	354	32	Blast search GenBank identifies unknown, hypothetical, and putative proteins, or proteins with unknown functions.
Metabolism	242	17.00	109	9.9	Amino acid, nucleotide, C-compound and carbohydrate, lipid, fatty acid and isoprenoid, nitrogen and sulfur, and secondary metabolisms.
Cell rescue, defense, and aging	90	6.35	22	2	Environmental stimuli responses (stress, wounding, phytohormone regulation, and ion deficiency), radical scavenging, detoxification, DNA repair, and cell death
Protein synthesis	152	10.7	253	22.9	Ribosomal proteins, translation (initiation, elongation and termination), translational control, and tRNA synthetases
Energy	39	2.8	29	2.6	Glycolysis and gluconeogenesis, pentose-phosphate pathway, TCA pathway, respiration, metabolism of energy reserves (e.g. glycogen, trehalose), and fatty acid oxidation
Transport facilitation	44	3.1	24	2.2	Ion channels, water channels, ion, sugar and carbohydrate, amino acid, lipid, purine and pyrimidine transporters, transport ATPases
Protein destination	51	3.6	59	5.3	Protein folding and stabilization, protein targeting, sorting and translocation, protein modification (e.g. myristylation, farnesylation, palmityolation, glycosylation) and proteolysis
Signal transduction	41	2.9	45	4.1	Receptor proteins, second messenger such as calmodulins, key kinases, key phosphatases, and G-proteins
Cell growth, division	18	1.3	13	1.2	Cell growth, development, cell cycle control and mitosis, cytokinesis, and DNA synthesis and replication
Transcription	17	1.2	29	2.6	mRNA synthesis (including general transcription activities, transcriptional control such as transcriptional factors, and chromatin modification), mRNA splicing, and mRNA stabilization and degradation
Metal ion homeostasis	48	3.4	69	6.2	Homeostasis of metal ions and other ions
	6	0.4	6	0.5	Vesicular transport (Golgi network), vacuolar transport, cytoskeleton dependent transport
	1,418	**100**	**1,106**	**100**	

Table 2. ESTs expression profiles at different time treatment in salt tolerance and sensitive rice.

EST Name	Annotation[a]	salt tolerance rice*					salt sensitive rice*		
		1 h	3 h	6 h	24 h	7 days	1 h	3 h	6 h
OC101E09	Glycine-serine rich protein	0.24[b]	0.13	-0.17	-0.33	-0.3	-0.15	-0.24	-0.33
OC103E04	Glycine-serine rich protein-(2)	0.17	0	-0.1	-0.15	-0.21	-0.14	-0.13	-0.19
OC01B06	Photosystem II	0.18	0.09	-0.15	-0.22	0	-0.08	-0.19	-0.26
OC03C10	High mobility group protein HMGd1	0.09	0.05	-0.17	-0.15	-0.23	-0.11	-0.24	-0.39
OE201H05	Nicotianamine synthase 1	0.3	0	-0.23	-0.2	0.1	-0.17	-0.24	-0.33
OE04C08	Similar to *Oryza sativa* bZIP transcriptional activator RF2a	0.09	0.05	-0.17	-0.15	-0.23	-0.11	-0.24	-0.39
OE04D07	Branched-chain amino acid aminotransferase	0.02	-0.37	-0.51	-0.3	-0.34	-0.33	-0.41	-0.28
OC104F11	Hypothetical protein	-0.24	-0.3	-0.17	-0.19	-0.14	-0.1	-0.11	-0.48
OC02H11	OSK4	-0.07	-0.41	-0.17	-0.09	0.02	-0.12	-0.42	-0.28
OC05B09	Phosphoribosyl pyrophosphate synthase isozyme 4	-0.19	-0.15	-0.23	-0.21	-0.01	-0.06	-0.24	-0.3
OE05F02	Putative glycosylation enzyme	-0.19	-0.15	-0.23	-0.21	-0.01	-0.06	-0.24	-0.3
OD103F11	Putative seed imbibition protein	-0.08	-0.17	-0.31	-0.1	0.06	-0.12	-0.24	-0.68
OE201A07	Pkutative translation initiation factor eIF-2B alpha subunit	-0.07	-0.41	-0.17	-0.09	0.02	-0.12	-0.42	-0.28
OC103D06	Receptor protein-like	0.02	-0.3	-0.35	-0.21	-0.07	-0.17	-0.2	-0.37
OE202C12	Similar to amino peptidase	0.02	-0.3	-0.35	-0.21	-0.07	-0.17	-0.2	-0.37
OE201B05	Polyubiquitin 6	-0.04	-0.32	-0.28	-0.08	-0.23	-0.11	-0.04	-0.3
OE04F01	Unknown	-0.04	-0.32	-0.28	-0.08	-0.23	-0.11	-0.04	-0.3
OC102F03	Unknown protein	-0.07	-0.29	-0.11	-0.11	0.03	-0.03	-0.07	-0.25
OE202C08	Cytochrome b5	-0.41	0.03	-0.19	-0.01	0.05	-0.01	-0.19	-0.24
OC103F07	Heat shock protein 70	-0.27	0.02	-0.07	-0.1	0.15	0.06	0.04	-0.14
OCP06F06	Nramp1 protein - rice	-0.24	-0.1	-0.09	-0.01	0.08	-0.03	-0.05	-0.15
OC05H08	Nucleoside diphosphate kinase (NDPK-2)	-0.12	-0.05	-0.17	-0.06	-0.03	-0.08	-0.17	-0.23
OE05E02	Putative glycosylation enzyme	-0.54	-0.08	-0.15	-0.01	-0.02	-0.05	-0.15	-0.26
OE04A04	TMK (gibberellin-induced receptor-like kinase)	-0.39	-0.1	0.01	-0.06	-0.07	0.02	0.05	-0.11
OE202F11	Water channel protein (WCP-IV)	-0.22	0.07	-0.21	-0.14	0.24	0.11	0.08	-0.36
OC03F06	ABA and stress induced protein (Asr1)	0.06	0.43	0.54	-0.21	0.17	0.16	0.48	0.32
OC103B03	Calmodulin (CaM1)	0.13	0.26	0.26	0.14	0	0.23	0.34	0.18
OC102E04	Gda-1	0.12	0.3	0.29	0.16	0.25	0.13	0.39	0.26
OE06C12	Gigantea protein	0.04	0.21	0.34	0.03	0.06	0.06	0.44	0.35
OE202B01	Osr40c1 ABA and salt responsive	0.15	0.47	0.47	0.35	0.05	0.05	0.51	0.56
OC104G09	S-adenosylmethionine decarboxylase 2	0.13	0.35	0.14	0.12	0.11	0.23	0.45	0.16
OC01C11	Subtilisin-chymotrypsin inhibitor 2	0.18	0.3	0.44	0.11	-0.1	0.22	0.35	0.41
OC101D09	Trypsin inhibitor-(1)	0.19	0.32	0.26	0.25	0.04	0.17	0.45	0.42
OE05F06	5-methyl tetrahydropteroyltriglutamate-homocysteine S-methyltransferase	0.05	0.09	-0.11	-0.07	-0.09	-0.12	0.08	-0.02
OE202G02	ADP-ribosylation factor 1	0.09	0.03	0.03	-0.04	0	-0.02	0.08	0
OC03H04	Cleft lip and palate transmembrane protein 1	-0.02	0.08	0.03	-0.08	-0.01	-0.02	0.01	0.01

Table. 2 continues

EST Name	Annotation[a]	salt tolerance rice*					salt sensitive rice*		
		1 h	3 h	6 h	24 h	7 days	1 h	3 h	6 h
OC04D10	Cyclin	0.11	0.09	-0.09	-0.12	-0.12	0	0.09	-0.1
OC01C02	Elongation factor 1-beta	0.09	-0.08	0.06	0	0.01	-0.02	0.06	-0.02
OC04B02	Glutathione S-transferase II	-0.04	0.12	-0.05	0.01	0	-0.08	0.07	0.06
OE201D03	H^+-transporting ATP synthase (EC 3.6.1.34) beta chain, mitochondrial	0.06	0	0.02	-0.15	-0.01	-0.12	0.03	0
OC04A03	Hypothetical protein	0.07	-0.01	0.04	-0.03	-0.1	0.03	0.04	-0.12
OC01D06	Low temperature and salt responsive protein LTI6B	0.02	0.27	0	-0.06	-0.04	0.12	0.09	-0.07
OE202F10	Plasma membrane H^+-ATPase	0.07	0.24	0.05	0.05	0.06	0.03	0.13	-0.12
OC101B02	S-adenosyl-L-methionine synthetase	-0.03	0.1	-0.05	-0.07	0.05	0	0.16	-0.03
OE06H10	Similar to H^+-ATPase	0.12	0.04	0.08	0.03	0.01	0.03	0.15	-0.01
OE105C04	Similar to zinc finger transcription factor Kaiso	0.02	0.02	0.04	0.05	0	0.01	0.14	-0.03
OE06B01	Stress-induced protein OZI1 precursor -Arabidopsis thaliana	0.04	0.03	0.03	0.03	0	0	0.09	0.08
OC101G02	Cysteine proteinase (EC 3.4.22.-) precursor	-0.04	0	0.08	-0.05	0.01	0.02	0.07	-0.08
OC03A08	Unknown	0.04	0.03	0.03	0.02	0.01	0.01	0.08	-0.06
OC04E04	Unknown	0.08	0.07	0.07	-0.09	-0.1	-0.06	0.07	0.08
OE06G04	Vacuolar H^+-ATPase c subunit	-0.03	0	0.06	0.03	0.04	0.03	0.17	0
OC102B01	Alpha2-tubulin	0.02	-0.05	-0.18	-0.14	-0.3	-0.09	0.04	0.01
OD103F05	Alpha-tubulin	-0.14	-0.02	-0.04	0.05	0.08	0.01	0.03	-0.06
OC102F01	Alpha-tubulin	0.08	0.02	-0.17	-0.15	-0.27	-0.08	0	0.05
OE105E03	No hit	0.03	-0.03	-0.22	-0.33	-0.22	-0.04	-0.02	-0.04
OE06D06	Late embryogenesis abundant protein	-0.06	0.02	0	0.06	0.07	0.09	0.06	-0.1
OC103G02	Peroxidase ATP19a	0.07	-0.14	0	-0.47	-0.33	0.04	-0.01	-0.08
OE06D09	S-adenosyl-L-methionine synthetase, pOSSAMS2	0.06	-0.03	-0.19	-0.29	-0.14	-0.06	0.01	-0.12
OC04A01	S-adenosylmethionine decarboxylase	-0.01	-0.06	-0.08	-0.17	-0.09	-0.02	-0.01	-0.03
OE04H05	Unknown	0.06	-0.01	-0.16	-0.29	-0.16	-0.01	0	-0.02
OE202F06	Alpha 7 subunit of 20S proteasome	-0.06	0	-0.02	0.03	-0.01	-0.03	-0.15	-0.17
OD103F04	Asparagine synthetase	-0.01	-0.03	-0.01	0	-0.01	-0.04	0	-0.05
OC03H12	ATP/ADP translocator	0.04	-0.06	-0.16	-0.04	-0.02	-0.07	0.06	-0.14
OC05C06	Calmodulin (CaM2)	-0.02	0.01	0.04	0.04	0	-0.01	-0.05	-0.05
OC104B10	DnaK-type molecular chaperone hsp70	0.03	-0.16	-0.07	-0.1	0.16	0.09	0.05	-0.07
OC101F12	GTP-binding regulatory protein beta chain	0.06	0.02	-0.02	0.01	0.02	0.01	-0.06	-0.06
OC02C03	H^+-transporting ATP synthase	0.02	-0.04	-0.07	-0.08	-0.07	-0.04	-0.06	0.02
OC01D12	High mobility group protein (HMGP-1)	0.02	0	0.03	-0.12	0.02	-0.03	0.05	-0.17
OC102C05	Hydroxyproline-rich glycoprotein 1 - 60S ribosomal protein L14	0.02	-0.04	0.03	0	-0.06	-0.01	-0.12	-0.02
OC04A05	Hypothetical protein	-0.01	-0.05	0.03	0.02	-0.04	0	-0.06	0.05
OC05F06	Hypothetical protein	-0.14	0.04	0.01	0.06	0.12	-0.0002	-0.0003	-0.00054
OE105F10	Hypothetical protein	0	-0.03	0	-0.04	0.04	0.04	-0.09	-0.13
OD103D02	Hypothetical protein	-0.08	0.01	0	0.01	0.03	0.07	-0.01	-0.1

Table. 2 continues

EST Name	Annotation[a]	salt tolerance rice*					salt sensitive rice*		
		1 h	3 h	6 h	24 h	7 days	1 h	3 h	6 h
OE105E04	Similar to small zinc finger-like protein	-0.07	-0.04	0.01	-0.01	-0.02	-0.02	-0.14	-0.19
OE05C06	Sucrose synthase-1	0	-0.2	-0.02	0.06	-0.07	-0.11	-0.06	-0.14
OC05B07	Transport inhibitor response 1	0.02	0	0.05	0.06	0.08	0	-0.06	-0.1
OC102D03	Unknown	-0.06	-0.19	0.01	0.06	0.08	0.04	0.07	-0.06
OC02A04	Histidine kinase	-0.09	-0.05	0.05	0.03	0	-0.07	0.05	-0.09
OC104G11	Unknown	-0.05	-0.26	-0.06	-0.1	0.03	0.01	0.01	-0.15
OE202A05	Phospholipase C	-0.24	-0.01	0.05	0.01	0.35	0.07	0	0.1
OE04F10	Metallothionein-like protein, OsMT-1	-0.34	-0.14	0.06	0.08	0.45	0.08	0.01	-0.19
OC04G09	Tonoplast intrinsic protein, gamma-Tip	-0.27	0.06	0.01	0.18	0.18	0.07	0.06	-0.14
OC104E01	Water channel protein (WCP-I, isoform)	-0.36	0.04	-0.01	0.13	0.4	0.1	0.22	-0.04
OC102F10	40S ribosomal protein S4	0.26	0.06	0.11	-0.02	-0.1	0.05	0.12	0.06
OC101D06	40S ribosomal protein S7	0.26	0	0.06	0.05	-0.09	0.06	0.12	0.06
OC104H10	40S ribosomal protein S9	0.28	0	0.242	-0.19	-0.14	-0.04	0.09	0.12
OC102H08	Beta-glucosidase, chloroplast	0.13	0.29	0.08	-0.05	-0.2	-0.05	0.08	-0.03
OE201C03	Calcium dependent protein kinase	0.48	0.26	0.18	0.04	-0.13	0.12	0.31	0.18
OC104A07	Elongation factor-1 alpha; EF-1 alpha	0.25	-0.01	0.12	0.03	-0.01	0.04	0.07	0.11
OC104G08	Nucleoside diphosphate kinase (NDPK-1)	0.26	0.12	0.05	-0.22	-0.24	0	0.11	0.08
OC04D05	Peroxidase-1 (EC 1.11.1.7)	0.32	0.16	0.18	0.08	-0.01	0.09	0.21	-0.12
OC04H11	Receptor-like protein kinase	0.16	0.11	0.09	0.09	-0.03	-0.02	0.15	0.09
OC102G10	Trehalose-6P phosphatase	0.01	0.1	0.12	0.12	-0.27	0.13	0.2	-0.04
OE202B09	Unknown	0.32	0.13	0.05	-0.09	0	0.07	0.25	-0.03
OC102B11	14-3-3 protein	0.03	0.11	0.16	0.03	0.03	0.06	0.26	0.1
OC102E03	Ascorbate peroxidase, cytosolic	-0.02	0.1	0.14	0.31	0.16	0.07	0.1	0.03
OC02F04	Cationic peroxidase	0.14	0.11	0.11	0.11	0.09	0.1	0.24	-0.15
OC01D05	Cyclophilin 2	0.07	0.09	0.26	0.1	0.11	-0.01	0.15	0.08
OC01B11	Glutathione S-transferase (auxin-induced)	0.16	0.04	0	0.39	0.15	0.06	0.28	0.2
OE201B04	Glyceraldehyde-3-phosphate dehydrogenase	-0.11	-0.04	0.1	0.1	0.08	-0.08	0.09	0.02
OD102G11	H^+-transporting ATPase (EC 3.6.1.35)	-0.06	0.04	0.08	0.13	0.09	0.08	0.22	0.09
OC02A01	Homolog of the Contains similarity to protein phosphatase 2C	0	0	0.24	0.13	0	-0.02	0.09	-0.1
OE201G12	Osr40g2	-0.05	0.2	0.17	0.22	0.05	-0.143	0.21	0.15
OE05A11	Putative methyltransferase	0.03	0.01	0.15	0.05	0.05	0	0.02	0.11
OC04F08	Thioredoxin h	-0.07	0.12	0.2	0.07	0.09	0.08	0.04	0.06
OC104B05	Unknown	0	0	0	0.24	0.23	0.04	0.08	-0.02

[a]Indicates the putative functions of the gene products that are expected on the basis of sequence similarity

[b]Log_{10} transformed ratios of all genes at series times of salt stress were listed

الملخص العربي
التنقيب في قواعد البيانات لتحديد چينات تحّمل الملوحة في الأرز

محمد عبد الحميد الاسكندراني[1]، عاطف شفيق سليمان[2]، حسام الدين محمد الوكيل[2]، أميرة محمد المغاوري[1]

[1]معهد الدراسات العليا والبحوث - جامعة لاسكندرية

[2]كلية الزراعة - سابا باشا- جامعة الاسكندرية

ان عملية التنقيب في قواعد البيانات من شأنها ان تميز علاقات وأنماط في مجموعة كبيرة من البيانات ليس بها اي وجه تشابه أومعقدة، وبما ان شرائح المصفوفات الدقيقة للـ DNA تعطي الكثير من بيانات التعبير الچيني لآلاف الچينات وعلي هذا فقد تطورت طرق التنقيب عن البيانات مؤخرا لمساعدة الباحثين لايجاد علاقات جديدة وهامة في هذه الشرائح وايضا الدور المتزايد لقواعد البيانات العامة لنماذج الكائنات المتاحة والمعلوماتية الحيوية في التنقيب عن البيانات وهذا قد قدم فرص جديدة وتحديات للباحثين لتطوير ادوات جزيئية للاكتشافات الوراثية. من العوامل المحددة لانتاج المحاصيل في العديد من البلدان الضغوط اللاحيوية والتي منها الجفاف، البرودة، الحرارة، نقص المعادن ،السمية والملوحة مما قد تتسبب في نقص الانتاجية بنسبة 50% علي مستوي العالم وحوالي 10 مليون هكتار من الاراضي لا تستخدم بسبب الملوحة. ولتقليل تأثير الملوحة علي انتاجية المحصول وتعظيم الاستفادة من الاراضي الملحية تستخدم تقنيات ادارة الماء والتربة أو التقنيات الحيوية. وحيث ان نبات الأرز يلعب دورا حيويا كثاني محصول حبوب في العالم، ومقدرته علي البقاء مغمورا تحت الماء لمدة طويلة وايضا لاكتمال معرفة التتابع الجيني للارز أصبح أهم النماذج النباتية للدراسة. وكخطوة لتحديد الجينات والبروتينات المسئولة عن تحمل الملوحة في الارز تم التنقيب في المكتبات الجينية لجذور نبات الارز المعاملة بكلوريد الصوديوم بتركيز 150 مللي عياري لأوقات مختلفة. حيث تم تحليل 1,418 قطعة جينية من النباتات المعرضة للاجهاد خلال فترات زمنية مختلفة لمدة 1، 3، 6 ساعات وايضا يوم وأسبوع، بالإضافة الي 1,106 قطعة من المكتبة الجينية للنبات غير معرض للملوحة. وهذه الدراسة ادت الي معرفة جينات جديدة خاصة لمواجهة الملوحة من خلال المكتبات الجينية للارز الموجودة في قواعد البيانات العالمية.