Bioinformatics International

Case Report
Published: 05 Jun, 2020

A Computational Network Biology Approach to Understand Salinity Stress Response in Rice (*Oryza sativa L.*)

Samarendra Das[1,2,3], Swarnaprabha Chhuria[4,5], Eric C Rouchka[3,6] and Shesh N Rai[2,3,7,8,9]*

[1]*Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute, PUSA, India*

[2]*Biostatistics and Bioinformatics Facility, JG Brown Cancer Center, University of Louisville, USA*

[3]*School of Interdisciplinary and Graduate Studies, University of Louisville, USA*

[4]*Department of Horticulture, Institute of Agricultural Sciences, SOA University, India*

[5]*Department of Horticulture, College of Agriculture, Odisha University of Agriculture and Technology, India*

[6]*Computer Engineering and Computer Science, Speed School of Engineering and Bioinformatics and Biomedical Computing Laboratory, University of Louisville, USA*

[7]*Alcohol Research Center, University of Louisville, USA*

[8]*Hepatobiology and Toxicology Center, University of Louisville, USA*

[9]*Department of Bioinformatics and Biostatistics, University of Louisville, USA*

Abstract

Rice (*Oryza sativa L.*), the major staple food for more than half of world's population, is being seriously affected by salinity stress worldwide. Salinity tolerance in rice is governed by many genes, identification of these stress responsive key genes as well as understanding the underlying cellular mechanisms is of paramount importance for developing salt tolerant varieties. In this study, meta-analysis was performed to combine gene expression gene expression datasets related to the identification of salinity stress responsive genes. A two-stage filtering approach was used to initially identify relevant genes. Then, a weighted gene co-expression network analysis was performed to detect the various gene modules associated with salinity stress in rice followed by DHGA approach to detect hub genes and unique hub genes. Moreover, other bioinformatics tools and techniques like Gene Ontology, motif analysis, protein structure prediction and protein-protein interactions were used to understand the salinity stress response mechanism in rice. Through the hub gene detection approach, 167 and 178 hub genes were identified in salinity stress and normal condition respectively, where 121 hub genes were common to both the conditions and 46 were unique to salinity stress condition. The functional enrichment analysis of hub genes further revealed their involvement in various processes linked with the salinity stress in rice. The 46 salinity stress genes were further analyzed with QTL, protein-protein interaction, gene ontology and motif analysis. These identified genes and mechanisms will add to the understanding of salinity response and its regulation in rice.

Keywords: Gene; Gene co-expression Network; Salinity; Hub gene; Rice

OPEN ACCESS

Correspondence:
Shesh N Rai, Department of Bioinformatics and Biostatistics, University of Louisville, Louisville 40202, KY, USA, Tel: 15024260016; E-mail: shesh.rai@louisville.edu

Received Date: *07 May 2020*
Accepted Date: *03 Jun 2020*
Published Date: *05 Jun 2020*

Citation:
Das S, Chhuria S, Rouchka EC, Rai SN. A Computational Network Biology Approach to Understand Salinity Stress Response in Rice (Oryza Sativa L.). Bioinform Int. 2020; 1(1): 1003.

Introduction

Rice (*Oryza sativa L.*) is the most important cereal crop and the major food source for more than half of world's population. Further, the productivity of rice has been increasingly affected by salinity stress worldwide [1]. The salinization problem affects at least one-third of the world's arable land and many more areas will be added to the list in future due to climate change and global warming [2,3]. Rice is preferably grown in submerged condition, where high salinity stress is highly prevalent [4]. Salinity occurs when there is high concentration of soluble salts found in soil/water. Salinity stress causes damage to plants in two ways *i.e.* (i) osmotic stress occurs due to relatively high salt ion concentration in soil/water that poses a threat to plant tissue and (ii) ion specific stresses resulting mostly from altered ratios of sodium and potassium ions (Na^+/K^+) and sodium and chlorine (Na^+/Cl^-) ions which causes damage to plant physiology [5]. At the cellular level, the mechanism of salinity response in rice is not completely known. Therefore, it is necessary to use the available network biology and bioinformatic tools to understand the underlying mechanism of

salinity stress response in rice. In addition, there is a need to identify salinity responsive genes in rice, which can further be used for the breeding of salinity tolerant rice varieties.

With the advancement of genomic technologies, a huge volume of high throughput and high dimensional data has been generated through various experiments across the world. A Gene Expression (GE) study is one such approach and the data generated from such technology are available in the public domain databases such as NCBI, Array Express, *etc*... over the years. There are both challenges and advantages in analyzing crop GE data. For crops, there are typically limited experimental datasets available, usually generated over varying experimental conditions and little bioinformatics work has been done on these data so far. Hence, integration and analysis of data generated by GE experiments for the same stress or related conditions is crucial in systems biology to enhance the sensitivity of the hypothesis under consideration for drawing valid conclusions [6]. For instance, meta-analysis of microarray data pertaining to different experiments in rice over different locations [7], soybean and *Arabidopsis* revealed the presence of highly connected key genes that are central to the plant defense system under various biotic and a biotic stresses [8,9].

Application of network theory in GE data analysis leads to the discipline of network biology and Gene Co-expression Network (GCN) analysis is one such key approach. In other words, GE data is given as input to network analysis tools to analyze and extract informative knowledge [7,9,10]. In this direction, Weighted Gene Co-Expression Network Analysis (WGCNA) is used to decipher co-expression patterns of genes across samples [11]. More specifically, WGCNA identifies modules using GE levels that are highly correlated across samples [12]. WGCNA has been successfully applied to detect co-expression modules in *Arabidopsis*, rice, maize, soybean and poplar [8-10,13]. In rice, it has been used to identify gene modules associated with drought stress tolerance used the gene network analysis technique to functionally annotate the rice genes [9,6,14]. Used this technique to identify the associated gene co-expression modules as well as consensus modules for drought and bacterial stress in rice *via* meta-analysis. WGCNA leads to construction of GCN, a scale free network, where, genes are represented as nodes and edges depict associations among genes [12]. In such a network, highly connected genes are called hub genes, which are expected to play an important role in understanding the biological mechanism of response under stress/disease [8,15]. Identification of such genes will also help in mitigating the stress response in crop plants through genetic engineering.

In this study, an attempt has been made to combine GE datasets for the identification salinity stress responsive genes and gene modules in rice. Further, a two-stage approach was used to detect the influential genes out of the large number of genes. Then, various hub genes and unique hub genes were identified for two differential GCNs constructed under salinity stress and control conditions using the available DHGA approach [7]. Here, we also used the WGCNA approach to detect the gene modules for salinity stress and further their functional enrichment analysis revealed the underlying mechanism of such stress response in rice [8]. Moreover, the findings were enriched with motif analysis and protein-protein interactions analysis.

Materials and Methods

Data collection

The rice microarray data was collected from GE Omnibus with platform GPL2025 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL2025), which contains 3,384 samples and 208 series of *Oryza sativa L.* This platform consists of experimental samples on 57,381 probes generated by using the affymetrix gene chip rice genome array. Out of 3,384 samples only few are related to salinity stress. From these series, datasets pertain to GSE13735, GSE14403, GSE21651, GSE28209, GSE16108 and GSE6901 are selected to study the salt stress response mechanism in rice through meta-analysis. The detailed descriptions about the selected GE samples are given in Supplementary Data S1.

Data pre-processing

The raw CEL files of salinity stress related samples were initially processed using the Robust Multichip Average (RMA) algorithm that involves background correction, quantile normalization and summarization through the median polish approach [16,17]. To remove the outlier samples, meta-analysis was performed. This involves with selection of samples with mean ≥ 5.2 and standard deviation ≤ 2.8, as they are observed to be highly homogeneous at these parameters setting irrespective of their experimental conditions (Figure S1 in Document S1). In this process only 70 samples were selected (Document S1) and the log2 scale transformed expression data from the RMA for these selected experimental samples were used for further statistical analysis. Further, the analytical steps undertaken in this study are shown in Figure 1.

Selection of influential genes

A two-stage filtering procedure was applied to identify the genes that are differentially expressed in salinity stress conditions compared to normal condition. In the first stage, t-test was employed and the genes which were found to be significant at 5% level of significance were retained for the next stage of filtering. The test statistic for testing the significance of the *i-th* GE profile is given by:

$$t_i = \frac{\bar{y}_i - \bar{x}_i}{\sqrt{\frac{s^2 y_i}{ny} + \frac{s^2 x_i}{nx}}} \quad (1)$$

Where, $\bar{y}_i$ and $\bar{x}_i$ are the means, $s^2 y$ and $s^2 x_i$ are the variances, n_y and n_x are the sample sizes under salinity stress and normal conditions respectively. In the second stage of filtering, Fold Change (FC) measure was computed for the genes selected at the first stage. The genes having at least 1.5 FC in their expression level was selected for further analysis. For *i-th* GE profile, the FC measure was computed as:

$$FC_i = \log_2 \bar{y}_i - \log_2 \bar{x}_i \quad (2)$$

Weighted gene co-expression network analysis (WGCNA)

GCN was constructed by using gene co-expression similarity measure. The gene co-expression similarity measure s_{ij} between *i-th* and *j-th* gene was computed by using the absolute value of Pearson's correlation co-efficient (Horvath and Dong 2008) and given as:

$$s_{ij} = \left| cor\,(x_i, x_j \right| \quad (3)$$

Where, x_i and x_j, are the expression profiles of *i-th* and *j-th* gene respectively. Then, the adjacency score, a_{ij}, between *i-th* gene and *j-th* gene was computed as:

$$a_{ij} = s_j^{\beta} \quad (4)$$

Where, β (≥ 1) is soft threshold power determined based on the concept of scale free property of biological networks. The detail methodology for determination of the soft threshold power has been discussed in detail by [11]. The use of weighted co-expression network along with soft threshold has advantage over unweighted networks or classical network approach [11], because the continuous nature of the gene co-expression information is preserved. Also, the results of weighted network approach are highly robust with respect to the choice of the soft threshold parameter.

Hub gene identification

Hub genes are referred as highly interacting genes present in the GCNs [12]. In network theory, the node is defined as hub node, if its connection degree is greater than the average connection degree of the network [18]. From a biological point of view, the identification of such informative genes is highly desirable as it can throw light on the underlying stress response mechanism in plants and further can be used for breeding stress resistant cultivars. For this purpose, the statistical approach, *i.e.* DHGA, developed by was used and can be briefly explained as [8]:

Let the Weighted Gene Score (WGS) for *i-th* gene in terms of weighted gene connectivity is written as

$$WGS_i = \sum_j a_j \quad (5)$$

This WGS_i gave relative importance of *i-th* gene based on its connections to all other genes in the co-expression network. For the purpose of hub gene detection, we test following hypothesis:

$H_0 : WGS_i \leq \mu$ *i.e. i-th* gene in the co-expression network is not a hub gene (6)

$H_1 : WGS_i > \mu$ *i.e. i-th* gene in the co-expression network is a hub gene

Where, μ is average connection degree of the complete network model. To identify the hub genes for salinity stress in rice, we executed the dhga package implemented on the R language [8].

Identification of gene modules

To identify the gene modules (the genes which are closely related with each other), the average linkage hierarchical clustering was employed and Topological Overlap Matrix (TOM) based dissimilarity was used [12]. The TOM dissimilarity between *i-th* and *j-th* gene was computed as:

$$d_{ij} = \sum_u 1 - w_j \quad (7) \text{ where,}$$

$$w_{ij} = \frac{l_j + a_j}{\min(k_i, k_j) + 1 - a_j}; \quad l_{ij} = \sum_u a_{iu} a_{ju} \quad \text{and} \quad k_i = \sum_u k_{iu}$$

(connectivity of the *i-th* gene).

To implement this approach, the *Block Wise Modules* function available in *WGCNA* package of R-software was executed [20].

Gene ontology and motif analysis

Gene ontology analysis of the genes present within each module under salinity stress condition was carried out using Database for Annotation, Visualization and Integrated Discovery (DAVID), an annotation tool for researchers to understand biological meaning behind the list of genes (version 6.7) [19]. Further, to determine whether the genes in modules are transcriptionally co-regulated, a motif analysis was performed to find conserved motif regulatory elements in their promoters. Here, MEME was performed to predict motifs on the upstream region of 2,000 bps from the translation start site of the genes [20]. Then FIMO was used to conduct a Chi-squared test on the significance of these motifs in comparison to randomly selected genes in the genome.

Table 1: R packages used in this study.

Methods	Tools/R Package	Parameters
Genomic platform of R	Bioconductor	
Data preprocessing	Affy	Default setting
Network analysis	WGCNA	Default value given in the package
Differential hub gene analysis	dhga	Default setting
Motif detection (EM algorithm)	MEME	Default value given in the package
3D Protein structure	Swissprot	Default setting
Influential gene selection	R-code written	

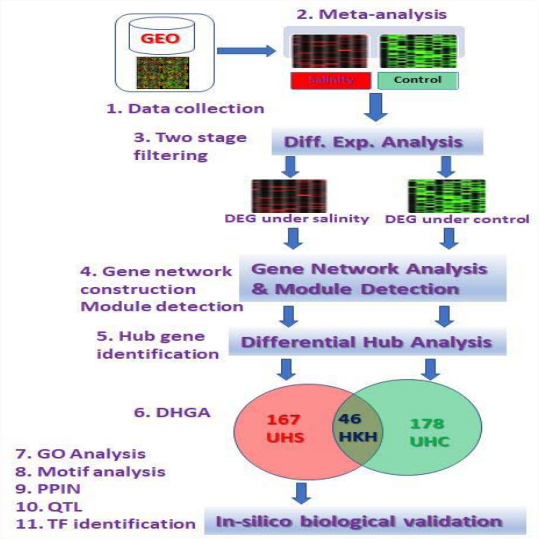

Figure 1: Analytical steps in flowchart form.
DEG: Differentially Expressed Genes; DHGA: Differential Hub Gene Analysis; UHS: Unique Hub gene in Salinity; HKH: House Keeping Hub genes; UHC: Unique Hub gene in Control condition; GO: Gene Ontology; PPIN: Protein-Protein Interaction Network; QTL: Quantitative Trait Loci; TF: Transcription Factor

Software and tools

For this present study, R software (*v.* 3.6.1) was used. It is an open source programming language and software environment for statistical computing and graphics [21]. The R language is widely used among statisticians and data miners for developing statistical and data mining models. For this study, different R packages and tools are used, which are listed in Table 1.

Results

Selection of influential genes

Using complete GE profiles, the expected differentially expressed genes were selected through a two-stage filtering procedure. In the first stage a t-test was used, where 5,036 number of genes were selected with p-values <0.05, which comprises 8.8% of total number of genes.

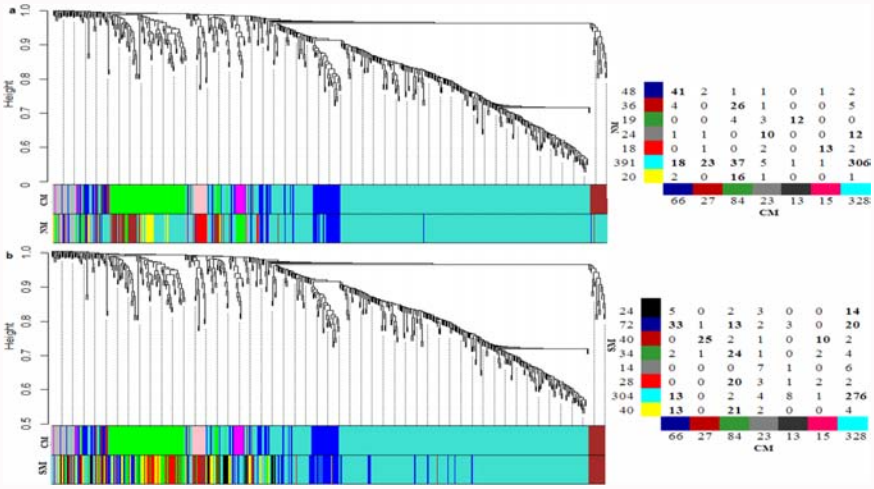

Figure 2: Clustering dendrogram of genes and various gene modules found in various conditions.
The correspondence between consensus modules and modules found in (a) normal and (b) salinity stress condition are represented. The extent of crosstalk between the Consensus Modules (CM) and stress modules (SM) and Normal Modules (NM) are shown in matrix form. Each row of the Table corresponds to modules under SM and NM conditions (labeled by color as well as text along with the number of genes in the modules) and column corresponds to consensus modules. Numbers in the Table indicate gene counts in the intersection of the corresponding modules. The figures in various colors in the Table showed the highest values.

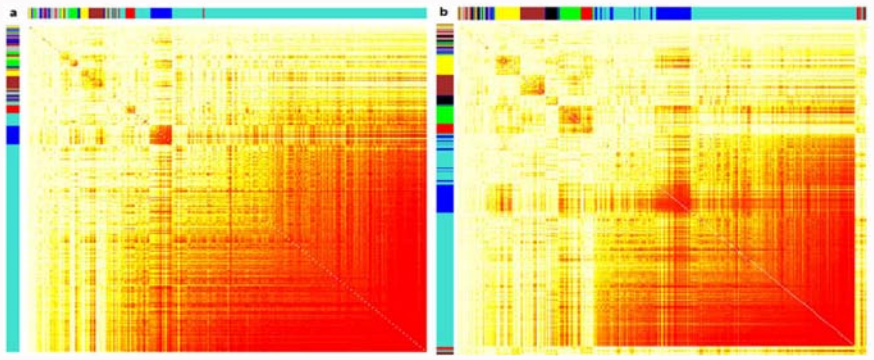

Figure 3: Dendrograms and heat maps of genes divided into tightly co-expressed modules under various conditions. The heat map depicts the correlations among the 556 genes from microarray GE profiling under (a) stress and (b) normal condition. The intensity of deep red color in the heat map shows the strong correlation among genes present in the module.

In the second stage, out of 5036 number of genes, 556 genes (0.97% of total number of genes) with FC value ≥ 1.5 were selected. Further, the selected probes are mapped to corresponding genes using MSU rice genome browser [22]. The detail descriptions about the selected 556 genes along with their genomic locations are given in Table S1.

Identification of gene modules for salinity stress condition

For both salinity stress and normal conditions, the value of β was observed as 6 with better approximation to scale free topology (R^2=0.85) and mean connectivity (K=30) (Figures in Document S2). Using this soft threshold parameter, 556 salinity responsive genes were divided into 8 and 7 modules for salinity stress and normal condition respectively (Figure 2). For this purpose, other parameters like module size, deep split level and tree merge cut height was set at 30 to 60.3 and 0.25 to 0.35 respectively. In both the conditions, the module represented by turquoise color contained maximum number of genes, hence designated as the largest module for either condition (Figure 2). The module memberships (number of genes in each module) of genes are listed in Table 2. Seven consensus modules between salinity and normal condition were also identified, where the module represented by turquoise color contained maximum number of genes (Figure 3b).

The matching of various modules in terms of their colors in either condition revealed the similar co-expression patterns, which can be well visualized from Figure 2. The module represented by brown color in consensus module diagram completely matched with that of salinity stress condition (Figure 2b). Other modules like blue and turquoise under salinity condition also partially matched with that of consensus condition (Figure 2b). Further, the extent of crosstalk between the various modules of consensus and stress conditions can be observed in matrix form (Figure 2). Similarly, for normal *vs.* consensus module diagram, the modules represented by turquoise and blue color under normal condition matched with that of consensus condition (Figure 2a).

The dendrograms and heatmaps of the selected genes divided into tightly co-expressed modules are represented in Figure 3 for both stress and normal conditions respectively by using TOM. The long branch in the dendrogram and high intensity of the red color in the heat map showed that the genes belong to the same module have higher degree of co-expression as compared to the genes outside of the module (Figure 3).

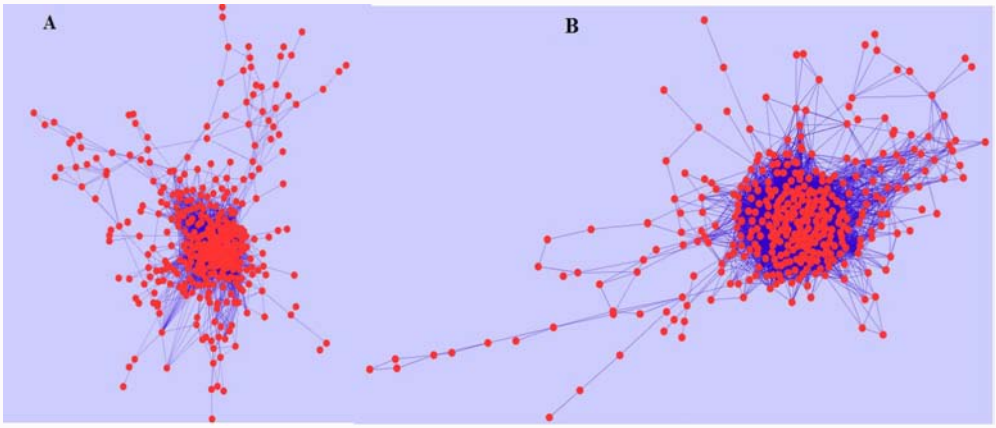

Figure 4: Gene co-expression networks for two differential conditions in rice. The GCNs are constructed for salt stress (A) and normal (B) conditions respectively. The nodes are represented as genes and edges are shown as the association among the genes.

Table 2: Module description along with gene and hub gene membership in salinity stress condition.

Modules	#G	#SRTF	#HG	#UHG	Functions
Black	24	1	0	0	Unknown
Blue	72	2	26	15	Peptidase activity (GO:0070011), Proteolysis (GO:0006508), Endopeptidase activity (GO:0004175), Ion transport (GO:0006811)
Brown	40	1	0	0	Cation binding (GO:0043169), Ion binding (GO:0043167)
Green	34	3	0	0	Unknown
Grey	14	1	0	0	Oxidation reduction (GO:0055114), Iron ion binding (GO:0005506), heme binding (GO:0020037), Tetrapyrrole binding (GO:0046906)
Red	28	2	0	0	Unknown
Turquoise	304	12	141	31	Oxidation reduction (GO:0055114), Heme binding (GO:0020037), Tetrapyrrole binding (GO:0046906), Electron carrier activity (GO:0009055), Cation binding (GO:0043169), Ion binding (GO:0043167), Metal ion binding (GO:0046872)
Yellow	40	3	0	0	Metal ion binding (GO:0046872), Cation binding (GO:0043169), Ion binding (GO:0043167)

Modules, identified gene modules represented by color; G, represents number of genes present in the module; SRTF, number of salinity responsive transcription factors; family, SRTF family; HG, represents number of hub genes in the module; UHG, number of unique hub genes in the module; Functions, molecular functions associated with each module.

Out of 556 genes, 56 were found to be Salinity Responsive Transcription Factors (SRTFs) in rice as per RICESRTF database [23] (Table 2). This database is available at http://www.nipgr.res.in/RiceSRTFDB.html and has 938 SRTF which are differentially expressed under salinity stress condition. The module membership of these SRTFs is given in Table 2. The largest turquoise module contained highest numbers of SRTFs belong to diversified of TF families followed by blue module (Table 2). The WRKY families of SRTFs are over-represented in blue module.

Identification of hub genes for salinity stress

For both salinity stress and normal conditions, the genes with *p-value* ≤ 0.0001 were considered as hub genes. A total of 167 and 178 hub genes were chosen for stress and normal conditions respectively (Table 3). The lists of genes and hub genes along with their locations and brief description are given in Table S1 and S2. From the module membership of the hub genes, it was observed that all the identified hub genes under salinity stress condition belong to two modules namely turquoise (141) and blue (26) (Table 2). From differential co-expression network analysis (analysis of networks for two contrasting conditions, *i.e.* salinity *vs.* control) unique hub genes (UHGs) were identified. In other words, the hub genes in the GCN, which are unique to either, stress (disease) or control condition is referred as UHG. It was found that 46 and 57 are UHG for salinity stress and normal conditions respectively, whereas 121 hub genes are common to both the conditions. Out of 46 UHGs expressed under salinity condition 31 are from turquoise and 15 from blue module (Table 2). A brief description about these 46 UHGs are provided in Table S3. The GCNs constructed for the two differential conditions are shown in Figure 4. Further, large numbers of gene-gene interactions in the GCNs are found in normal condition as compared to salt stress in rice (Figure 4).

Table 3: Comparison of DHGA and existing approaches in terms of predicted hub genes.

Data description	dhga Approach					
	p-value ≤ 1E-04		p-value ≤ 1E-08		p-value ≤ 1E-12	
	# HG	% HG	# HG	% HG	# HG	% HG
Salinity stress	167	30.03	135	24.28	82	14.17
Control	178	32.01	140	25.21	87	15.65

Functional analysis of modules under salinity stress

The GO analysis of the genes presents in the modules revealed the underlying molecular functions associated with the modules and are given in Table 2. In the largest turquoise module, the genes are over-represented in the category of oxidation reduction activity, heme binding, tetrapyrrole binding, electron carrier activity, cation binding, ion binding and metal ion binding. Similar interpretations are also made for all other modules (Table 2). Further, no significant

Table 4: GO term enrichment analysis of hub genes identified in salinity stress condition by using DAVID.

GO terms	Ontology	Description	#Genes	%Genes	p-value	Fisher Exact Value
GO:0055114	P	Oxidation reduction Activity	16	11.2	0.00053	2.00E to 04
GO:0005506	F	Iron ion binding Activity	12	8.4	0.0013	4.00E to 04
GO:0009055	F	Electron carrier activity	12	8.4	0.0014	4.40E to 04
GO:0020037	F	Heme binding	9	6.3	0.004	1.10E to 03
GO:0046906	F	Tetrapyrrole binding	9	6.3	0.0045	1.30E to 03
GO:0005509	F	Calcium ion binding	8	5.6	0.0055	1.40E to 03
GO:0043169	F	Cation binding	26	18.2	0.0096	5.90E to 03
GO:0043167	F	Ion binding	26	18.2	0.0096	5.90E to 03
GO:0006979	P	Response to oxidative stress	5	3.5	0.012	2.00E to 03
GO:0004601	F	Peroxidase activity	5	3.5	0.013	2.30E to 03
GO:0016684	F	Oxido-reductase activity, acting on peroxide as acceptor	5	3.5	0.013	2.30E to 03
GO:0046872	F	Metal ion binding	24	16.8	0.017	1.10E to 02
GO:0016209	F	Antioxidant activity	5	3.5	0.018	3.60E to 03
GO:0008171	F	O-methyltransferase activity	3	2.1	0.024	1.80E to 03
GO:0015101	F	Organic cation transmembrane transporter activity	2	1.4	0.04	1.10E to 03

Ontology ''P'' indicates Biological Process and Ontology ''F'' indicates Molecular Function. ''# of genes'' is the number of genes in the query gene list. "% of genes" is the percentage of genes in the query gene list. P-value represents the statistical significance of the gene enrichment test. "Fisher Exact Value" is the significance value obtained from Fisher's test.

Table 5: *Arabidopsis* homolog corresponds to unique hub genes for salinity stress in rice.

Gene ID	Query Species	Homolog ID	Homolog Species	BBMH score	BBMH e-value
Os01g07530	*Oryza sativa*	AT5G40390.1	*Arabidopsis thaliana*	959	0
Os02g19820	*Oryza sativa*	AT4G10850.1	*Arabidopsis thaliana*	228	3.00E to 60
Os02g37040	*Oryza sativa*	AT3G21670.1	*Arabidopsis thaliana*	596	1.00E to 170
Os08g43390	*Oryza sativa*	AT3G61880.2	*Arabidopsis thaliana*	582	2.00E to 166
Os09g16030	*Oryza sativa*	AT2G16890.2	*Arabidopsis thaliana*	312	4.00E to 85
Os09g38130	*Oryza sativa*	AT5G65980.1	*Arabidopsis thaliana*	479	2.00E to 135
Os11g34460	*Oryza sativa*	AT1G68050.1	*Arabidopsis thaliana*	874	0

Figure 5: Predicted 3D structure of two key proteins under salt stress in rice. The three-dimensional structures of two key proteins (a) LOC_Os01g63810.1 (b) LOC_Os01g62830.1, which are overlapped with salinity tolerant QTL regions are predicted.

GO Molecular Function (MF) terms were associated with black, green and red color modules (Table 2) and it may be inferred that role of the genes present in these modules under salinity stress are still largely unknown.

Functional analysis of hub genes under salinity stress

Further, the GO analysis of the identified hub genes under salinity stress condition was carried out by using DAVID database and the results are shown in Table 4.

From GO term enrichment analysis the hub genes were found to be mostly over-represented in the ontology category like oxidation reduction (redox) activity (GO: 0055114) and response to oxidative stress (GO: 0006979) (Table 4). The role of the genes in the redox activities are related to electron transport that balances the charges during ion transport [5]. The redox activities are also related to reactive oxygen intermediates that are produced in response to oxidative stress due to water deficit during salinity stress [24]. Under MF taxonomy, the chosen hub genes were over-represented in the categories like cation binding, ion binding, metal ion binding (Table 4), which might be due to the high ion concentration present in the soil or water because of salinity stress. The other members of chosen hub gene set were also found to be involved in other MFs like iron ion binding activity, electron carrier activity, tetrapyrrole binding, calcium ion binding, heme binding, *etc.* (Table 4). The role of the selected hub genes in tetrapyrrole binding activity is related to the osmotic adjustment for salt tolerance in crop. The biosynthesis of

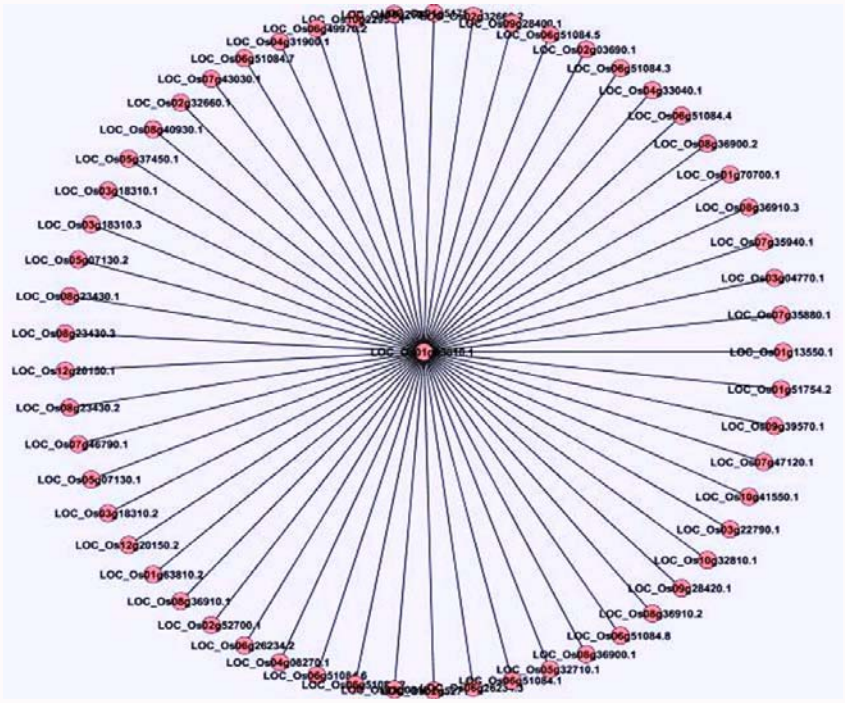

Figure 6: Protein interaction network for the protein LOC_OS01G63810.1.LOC_Os01g63810.1 is the central node and interacts with 57 other proteins in the constructed protein-protein interaction network.

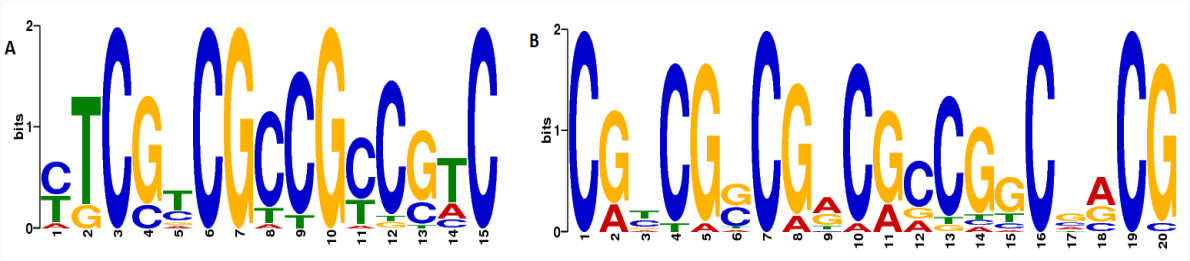

Figure 7: Detected motifs in the upstream sequences. The detected motifs are shown for the blue module (a) and turquoise module (b).

proline in plants also plays an important role in salinity tolerance, as it binds the free ions, thus stabilizes the cell structure. Further, the proline synthesis may also play role in nitrogen metabolism in plants which is related to salinity tolerance in plants [25].

Unique hub gene analysis

Here, the identified UHGs under salinity stress condition (Tables 2, S3) were further validated by using QTL information. Further, 17 salinity stress responsive Quantitative Trait Loci (QTLs) were obtained from the Gramene QTL database (http://www.gramene.org/qtl/). These QTLs and hub gene regions were then mapped using Gramene annotation of rice genome obtained from MSU Rice Genome Annotation (Osa1) Release 7 [22]. Interestingly, two UHGs *Os01g0856900* (from blue module) and *Os01g0847100* (from turquoise module) are found to be perfectly overlapped with QTLs (AQEM001and AQEM008) region, which indicates their key role in salinity stress tolerance in rice (Table S3). These two genes encode *LOC_Os01g63810.1* and *LOC_Os01g62830.1* proteins respectively, the 3D structures of which were predicted by using SWISS MODEL and shown in Figure 5.

The protein-protein interaction data for the protein *LOC_Os01g63810.1* was extracted from database of interacting proteins in *Oryza sativa* available at http://csb.shu.edu.cn/dipos/. Further, this information was used to construct the Protein Interaction Network (PIN) only for the protein *LOC_Os01g63810.1*. The PIN was plotted using *RCytoscape* and is shown in Figure 6 [26]. From the PIN, it is observed that the protein interacts with 57 other proteins (Figure 6). Furthermore, identified 46 UHGs were mapped to their *Arabidopsis* orthologs by using GreenPhylDB and are shown in Table 5 [27]. The results showed that seven UHGs in rice have unique orthologs corresponding to *Arabidopsis* (Table 5).

Motif analysis of unique hub genes

The identified 46 UHGs under salinity stress mostly belong to blue (15) and Turquoise (31) gene modules. To determine whether the UHGs in both the modules are transcriptionally co-regulated or not, motif analysis was performed on both the gene modules. The nucleotide sequence data of the UHGs present in the both the modules were taken from the Gene database of NCBI (www.ncbi.nlm.nih.gov/gene) and further used in motif analysis separately for both the modules.

At the transcription level, 31 and 15 UHGs present in Blue and

Turquoise modules are transcriptionally co-regulated by examining if the promoter regions of these genes share conserved motifs as the regulatory elements. Initially, three candidate motifs are predicted separately for both the modules of length 15 and 20 for blue and turquoise modules respectively and only one motif is validated by sequence comparison with known *cis* regulatory motifs in the place database. Further, the results are shown in Figure 7. The results indicated that the detected motif for the blue module is TCG-rich motif, while the detected motif for the turquoise module is rich in GC-content (Figure 7). The motif analysis and co-expression analysis provide some support that the identified gene module is likely to be co-regulated (Figure 7).

Discussion

Understanding salinity stress response mechanism in rice is of paramount importance for plant breeders to develop salinity stress tolerant cultivars. In public domain databases, there are few samples available related to salinity stress in rice, which have been generated over varying experimental conditions by multiple studies. Thus, meta-analysis was performed to combine these datasets and the meta-data was used for further statistical analysis. Then, developed and existing techniques were applied on the Meta GE data to identify the responsible key genes and modules to understand salinity stress response mechanism in rice.

The chromosome annotation of the hub genes which are uniquely expressed under salinity stress condition showed that they mostly lie on the chromosome 1, 2, 3, 6, and 9. The hub gene *Os05g0439000* (*p-value* = 6.32×10^{-84}), which encodes for zinc finger protein was found to have regulatory role in drought and salt tolerance [19]. The hub gene *Os02g0715000* was found to be involved in the cysteine proteinase activity that is responsible for plant growth and for conferring tolerance to abiotic stress like salinity [28]. The hub gene *Os11g0591800* encodes wound induced protein involved in cell wall structure, which is responsible for salinity tolerance in plants [29]. The hub gene, *Os01g0856900* encodes starch binding domain protein that may offer tolerance to salinity stress by accumulating carbohydrates such as sugars (*e.g.* glucose, fructose, fructans and trehalose) and starch occur under salt stress [30].

Hub gene *Os06g0271000* with significance value 8.55×10^{-06} involved in glycosyl transferase activity, which has been reported to have biological role in abiotic stress tolerance mechanism in *Arabidopsis* [31]. The hub gene *LOC_Os10g05400* was found to be involved in protein kinase activity that plays an important role in plant salinity stress tolerance by means of regulation of ionic and osmotic homeostasis [32]. The hub gene *Os11g0245100* encodes glutathione S-transferase zeta protein, which has been reported to have important role in salinity and oxidative stress in *Arabidopsis* [33]. The hub gene *Os03g0195900* with *p-value* = 5.98×10^{-70} involved in sulphate transfer activities that plays significant role in salinity and drought tolerance in plants. This transporter may play role in plant leaves in controlling their early response to water stress in strong connection with abscic acid biosynthesis [34]. The hub gene *Os02g0770800* involved in nitrate reductase activities that involves in salinity tolerance mechanism as it affects the synthesis of both proline and other free nitrogenous compounds, which might be utilized in osmotic adjustment [35]. The hub gene *LOC_Os09g09230* with *p-value* 6.40×10^{-84} involved in Dihydroflavonol-4-Reductase (DFR), which plays a key role in flavenoid biosynthesis in purple sweet potato that supports the protective function of anthocyanins of enhanced scavenging of reactive oxygen radicals in plants under salinity stress conditions [33]. Hub gene *Os04g0687900* with significance value 6.32×10^{-84} encodes OsFBT6 - F-box protein, which confers tolerance to abiotic stress in chickpea [36]. The hub gene *Os05g0515600* (*p-value* = 1.63×10^{-74}) involved in O-methyltransferase activity, which is associated with sodium sequestration excessively found under salinity stress condition [4]. The hub gene *Os06g0195800* with *p-value* 6.32×10^{-84} encodes DnaJ protein, which provides tolerance to abiotic stress by lowering the accumulation of ROS [37,38]. Among these uniquely expressed hub genes, the information about some genes and their role in abiotic stress tolerance could not be traced back by using available resources. However, the unreported genes may have some role in salinity tolerance in rice, which needs be studied in detail.

Conclusion

In this study, meta-analysis was performed to combine the publicly available GE datasets related salinity stress in rice. Using the available WGCNA and dhga approaches, a number of key genes with high connectivity in the GCN for both normal and salinity conditions were identified. Functional enrichment analysis of these key genes revealed the associated molecular functions under salinity stress. These identified genes may act as potential candidates for salinity stress response engineering in rice. Moreover, the identified genes were validated with QTL, protein-protein interaction, protein structure prediction, motif analysis etc. Such type of analysis provided information on various molecular mechanisms like biosynthesis of secondary metabolites and stress specific roles of certain plant products revealed. Further, this study will surely add on to the understanding of salinity response and its regulation in rice. Besides, two key proteins were predicted and mapped to QTLs, which can be used by breeders to clone these genes for salinity stress response engineering in rice.

Acknowledgement

This study was supported in part by National Institute of General Medical Sciences of National Institute of Health (NIH), USA (Grant No: P20GM113226 and P20GM103436). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. Authors are thankful to Indian Council of Agricultural Research (ICAR), New Delhi, India for providing Netaji Subhas-ICAR International Fellowship to Samarendra Das (OM No. 18(02)/2016-EQR/Edn.). Also Dr. S. N. Rai was partially supported by Wendell Cherry Chair in Clinical Trial Research.

References

1. Wu R, Garg A. Engineering rice plants with trehalose producing genes improves tolerance to drought, salt and low temperature. ISBN News Report. 2003.
2. FAO Soils Portal. 2016.
3. Dasgupta S, Hossain MM, Huq M, Wheeler D. Climate change and soil salinity: The case of coastal Bangladesh. Ambio. 2015;44(8):815-26.
4. Das S, Meher PK, Pradhan UP, Paul AK. Inferring gene regulatory networks using Kendall's tau correlation coefficient and identification of salinity stress responsive genes in rice. Curr Sci. 2017;112(6):1257-63.
5. Wang J, Chen L, Wang Y, Zhang J, Liang Y, Xu D. A Computational systems biology study for understanding salt tolerance mechanism in rice. PLoS One. 2013;8(6):e64929.
6. Tseng GC, Ghosh D, Feingold E. Comprehensive literature review and statistical considerations for microarray meta-analysis. Nucleic Acids Res.

2012;40(9):3785-99.

7. Rodrigo G, Carrera J, Ruiz-Ferrer V, del Toro FJ, Llave C, Voinnet O, et al. A meta-analysis reveals the commonalities and differences in Arabidopsis thaliana response to different viral pathogens. Plos One. 2012;7(7):e40526.
8. Das S, Meher PK, Rai A, Bhar LM, Mandal BN. Statistical approaches for gene selection, hub gene identification and module interaction in gene co-expression network analysis: An application to Aluminum stress in Soybean (Glycine max L.). Plos One. 2017;12(1):e0169605.
9. Shaik R, Ramakrishna W. Genes and co-expression modules common to drought and bacterial stress responses in Arabidopsis and Rice. Plos One. 2013;8(10):e77261.
10. Childs KL, Davidson RM, Buell CR. Gene co-expression network analysis as a source of functional annotation for rice genes. Plos One. 2011;6(7):e22196.
11. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
12. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Molec Biol. 2005;4:17.
13. Weston DJ, Karve AA, Gunter LE, Jawdy SS, Yang X. Comparative physiology and transcriptional networks underlying the heat shock response in Populus trichocarpa, Arabidopsis thaliana and Glycine max. Plant Cell Environ. 2011;34(9):1488-506.
14. Zhang L, Yu S, Zuo K, Luo L, Tang K. Identification of gene modules associated with drought response in rice by network-based analysis. Plos One. 2012;7(5):e33748.
15. Chen BS, Yang SK, Lan CY, Chuang YJ. A systems biology approach to construct the gene regulatory network of systemic inflammation via microarray and databases mining. BMC Med Genomics. 2008;1:46.
16. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy-analysis of Affymetrix Gene Chip data at the probe level. Bioinformatics. 2004;20(3):307-15.
17. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249-64.
18. Stumpf MPH, Porter MA. Critical truths about power laws. Science. 2012;335(6069):665-6.
19. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.
20. Bailey T, Bodén M, Buske F, Frith M, Grant C. MEME SUITE: Tools for motif discovery and searching. Nucleic Acids Res. 2009;37:W202-8.
21. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2015.
22. Kawahara Y, De la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice. 2013;6:4.
23. Priya P, Jain M. Rice SRTFDB: A database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. Database. 2013:2013.
24. Miller G, Shulaev V, Mittler R. Reactive oxygen signaling and abiotic stress. Plant Physiol. 2008;133(3):481-9.
25. Averina NG, Bayzaei Z, Shcherbakov RA, Usatov AV. Role of nitrogen metabolism in the development of salt tolerance in Barleys. Russ J Plant Physiol. 2014;61(1):97-104.
26. Shannon PT, Grimes M, Kutlu B, Bot JJ, Galas DJ. RCytoscape: Tools for exploratory network analysis. BMC Bioinformatics. 2013;14:217.
27. Rouard M, Guignon V, Aluome C, Laporte MA, Droc G. GreenPhylDB v2.0: Comparative and functional genomics in plants. Nucleic Acids Res. 2011;39:1095-102.
28. Grudkowska M, Zagdańska B. Multifunctional role of plant cysteine proteinases. Acta Biochim Pol. 2004;51(3):609-24.
29. Byun M, Kwon HB, Park S. Recent advances in genetic engineering of potato crops for drought and saline stress tolerance. In Advances in molecular breeding toward drought and salt tolerant crops. 200;713-37.
30. Miller G, Shulaev V, Mittler R. Reactive oxygen signaling and abiotic stress. Plant Physiol. 2008;133(3):481-9.
31. Sun YG, Wang B, Jin SH, Qu XX, Li YJ. Ectopic expression of Arabidopsis glycosyltransferase ugt85a5 enhances salt stress tolerance in Tobacco. Plos One. 2013;8(3):e59924.
32. Golldack D, Li C, Mohan H, Probst N. Tolerance to drought and salt stress in plants: Unraveling the signaling networks. Front Plant Sci. 2014;5:151.
33. Sharma R, Sahoo A, Devendran R, Jain M. Over-expression of a rice tau class glutathione s-transferase gene improves tolerance to salinity and oxidative stresses in Arabidopsis. Plos One. 2014;9(3):e92900.
34. Gallardo K, Courty PE, LeSignor C, Wipf D, Vernoud V. Sulfate transporters in the plant's response to drought and salinity: Regulation and possible functions. Front Plant Sci. 2014;5:580.
35. Viegas RA, Roberto A, de Melo B, da Silveira JAG. Nitrate reductase activity and proline accumulation in cashew in response to NaCl salt shock. Rev Bras Fisiol Veg. 1999;11(1):21-8.
36. Gupta S, Garg V, Kant C, Bhatia S. Genome-wide survey and expression analysis of F-box genes in chickpea. BMC Genomics. 2015;16(1):67.
37. Kong F, Deng Y, Zhou B, Wang D, Wang Y, Meng Q. A chloroplast-targeted DnaJ protein contributes to maintenance of photo system II under chilling stress. J Exp Bot. 2014;65(1):143-58.
38. Shaik R, Ramakrishna W. Genes and co-expression modules common to drought and bacterial stress responses in Arabidopsis and Rice. Plos One. 2013;8(10):e77261.