



## A computational system biology approach to construct gene regulatory networks for salinity response in rice (*Oryza sativa*)

SAMARENDRA DAS<sup>1</sup>, PRIYANKA PANDEY<sup>2</sup>, ANIL RAI<sup>3</sup> and CHINMAYEE MOHAPATRA<sup>4</sup>

Indian Agricultural Statistics Research Institute, Pusa, New Delhi 110 012

Received: 11 February 2015; Accepted: 26 June 2015

### ABSTRACT

Salinity is one of the most common abiotic stress which limits agricultural crop production. Salinity stress tolerance in rice (*Oryza sativa* L.) is an important trait controlled by various genes. The mechanism of salinity stress response in rice is quite complex. Modelling and construction of genetic regulatory networks is an important tool and can be used for understanding this underlying mechanism. This paper considers the problem of modeling and construction of Gene Regulatory Networks using Multiple Linear Regression and Singular Value Decomposition approach coupled with a number of computational tools. The gene networks constructed by using this approach satisfied the scale free property of biological networks and such networks can be used to extract valuable information on the transcription factors, which are salt responsive. The gene ontology enrichment analysis of selected nodes is performed. The developed model can also be used for predicting the gene responses under stress condition and the result shows that the model fits well for the given gene expression data in rice. In this paper, we have identified ten target genes and a series of potential transcription factors for each target gene in rice which are highly salt responsive.

**Key words:** Gene regulatory network, Multiple linear regression, Singular value decomposition, Target gene, Transcription factor

Rice (*Oryza sativa* L.) is the main staple food for more than half of the world's population and accounts for supply of 30-80% of daily calories consumed in Asia (Narisco *et al.* 2002). The production of rice is severely affected by salinity stress in soil or water due to the nature of rice plants to grow in swamp and fresh water marshes (Akbar and Ponnampereuma 1980). The detrimental effects of high salinity on plants can be observed as the death of plants and/or decrease in its productivity (Joseph and Mohanan 2013). Hence, salinity in soil or water is of much importance to agriculture as it limits rice production in about 30-50% of the rice growing area worldwide (Rahimi and Biglarifard 2011). The genes respond to different stresses at the transcriptional level and the products of these genes play role in the stress response and tolerance (Bray *et al.* 2000). The mechanism of salinity response in rice is quite complex and till now not known completely. Hence, the gene regulatory networks (GRN) can be used to understand the mechanism of salinity stress in rice.

In the post genomics era, modelling and construction

of GRN and gene interactions are important tasks in functional genomics and systems biology. The genome encodes thousands of genes which transcribed in proteins responsible for cell development and various cellular functions in response to diverse extracellular signals. These proteins interact with each other to comprise a structured regulatory network known as GRN (Wang *et al.* 2013). In system biology, GRN is a graphical representation in which nodes consist of genes and edges connecting between them show regulatory relationship. The regulation of gene expression by transcription factors (TF) is of keen importance as it binds to the promoter regions of their target genes (TG), thereby activating or inhibiting their expression. Deciphering and understanding TF-TG interactions are vital in the field of biology and medicines, ranging from the *in silico* modelling and construction of the GRN to the identification of new potential drug targets (Haury *et al.* 2012). Thus, mathematical modelling and computational algorithms for inferring the regulatory relations are indispensable. Infact, many mathematical models and algorithms have been proposed for the inference of GRN (de Jong 2002). For the discrete gene expression data (expressed: 1, unexpressed: 0), many models including Boolean Network model, Probabilistic Boolean Network model, multivariate Markov model etc. have been proposed (Akutsu *et al.* 2000, Boros *et al.* 1998, Ching *et al.* 2005, Hirose *et al.* 2006, Idekar *et al.* 2000, Noda *et al.* 1998,

<sup>1</sup>Scientist (email: sam@iasri.res.in); <sup>2</sup>Assistant Professor (email: ppl@nibong.ac.in), National Institute of Biomedical Genomics, Kalyani, West Bengal 741 251; <sup>3</sup>Principal Scientist and Head (CABin) (email: anilrai@iasri.res.in); <sup>4</sup>PhD Scholar (email: chinmayee09109@gmail.com), Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh 221 005

Shmulevich *et al.* 2002). These methods of discrete network construction fail, when there is large number of regulators and can't be applied to continuous gene expression data directly. For the continuous expression data, clustering algorithms, Bayesian, Dynamic Bayesian networks and Ordinary Differential Equations (ODE) based methods have been proposed for the network inference (Bansal *et al.* 2007). Bayesian method of network construction fails to consider the dynamic nature of the genes and Dynamic Bayesian method is limited to small networks due to high computational complexity. Yeung and Ruzzo (2002) used a linear model and singular value decomposition (SVD) to generate a family of candidate networks that are consistent with a given dataset, thus compensating for this deficiency in time points.

In this study, interest is in the continuous gene expression microarray data of salinity stress in rice. The difficulty of applying the ODE based models in such data sets lies in estimation of the interaction coefficients and the networks constructed based on this method may not satisfy the scale-free property of real biological networks (Barabasi and Albert 1999). These methods are often very complex and slow, do not consider the feature (TFs) selection as well as do not provide insight into the biological meanings of each parameter (Huang *et al.* 2010). In order to address the limitations of existing approaches for modeling and construction GRN for cross sectional gene expression data, an attempt has been made to develop an approach for construction and modelling of GRN which is based on multiple linear regression (MLR) and SVD. This technique also considers gene selection approach. This approach is applied to gene expression data in rice for construction and modeling of GRNs for salinity stress response.

## MATERIALS AND METHODS

For this study, the microarray data available at the NCBI Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/GEO>) generated by using Affymetrix platform is used, as it shows high degree of reproducibility and homogeneity across different laboratories when compared with other platforms. Affymetrix microarray data sets for salt responsive genes (Walia *et al.* 2009) are available in the above database (Cotsaftis *et al.* 2011, Pandit *et al.* 2010 and Jain *et al.* 2008) with accession number GSE14403, GSE16108, GSE6901, GSE13735. These data were generated from Affymetrix Rice Genome Array platform (GPL 2025), which contains 57381 probes and each probe corresponded to an individual gene. As there are 123 probe sets designed for control in this microarray. Therefore, after exclusion of these probe set, 57258 valid probe sets are obtained for further analysis.

Data analysis is performed using R (Ihaka and Gentleman 1995) with Bioconductor packages (Gentleman *et al.* 2004). Raw CEL files are processed using the Robust Multichip Average (RMA) algorithm available in the affy package (Gautier *et al.* 2004) with steps including

background correction, quantile normalization and summarization by the median polish approach (Irizarry *et al.* 2003). The log<sub>2</sub> scale data from RMA is used in statistical analysis. The expression level of the control conditions are taken as baseline. The gene expression level at each stressed conditions are subtracted from the baseline to get the relative expression level. The genes having at least 3-fold change in gene expression are selected and this procedure is continued for each experiment.

The set of genes whose expression level have at least 3-fold change and are commonly expressed in all the experiments, over different stages of growth, different tissues of rice cultivars are selected as TGs. For this study, TGs are identified which are commonly expressed in 4 out of 3 experiments over different tissues and growth stages of rice cultivars. The identified TGs having highly salt responsive nature are validated by published literature.

RiceSRTF data base, which is a good source of rice TFs, is explored to facilitate gene function analysis for abiotic stress responses in rice (Priya and Jain 2013) (<http://www.nipgr.res.in/RiceSRTFDB.html>). Total 934 TFs are selected from this data base, whose expressions are differentially expressed under salinity stress in rice. These TFs could be considered as the regulators of the TGs. One TG may also be considered as a regulator for another TG and *vice versa* (Huang *et al.* 2010). Hence, for a TG, the 934 TFs and the remaining TGs are considered to be regulators and this procedure is repeated iteratively for all TGs.

The candidate regulators from the RiceSRTF database are selected by correlation threshold of gene expression data, which is based on the assumption that there are possible correlations between TG and their upstream regulators (Chen *et al.* 2008). The Pearson's correlation coefficient is computed between the TGs and their own regulatory TFs separately and the correlation value is then used to identify the candidate regulators according to the assumption that the regulatory genes and TGs have a positively (or negatively) correlated with temporal relationship, if, the TG's expression profile is positively (or negatively) correlated with the regulatory genes profile.

A proper threshold limit for correlations between TFs and TG is chosen to select candidate transcription regulatory genes for each TG. For this purpose, we selected a random sample of 5000 genes from 57258 genes and computed their correlations by the Pearson's correlation co-efficient and as ranked in Fig 1. Further, by using the step of data randomisation procedure, the probability density function of parameter estimation showed the parameter estimations of regulatory genes by our threshold of correlation has a probability value less than 0.001.

$P [r_i \geq r_0] \leq 0.001$  where  $r_i$ : correlation co-efficient and  $r_0$ : threshold value

According to the Fig 1, 50% (0.7767 in correlation value) is selected as threshold value and the correlation values between the TGs and their regulators greater than threshold value are significant at 0.001 level of significance. In the next step, the candidate regulators are selected from

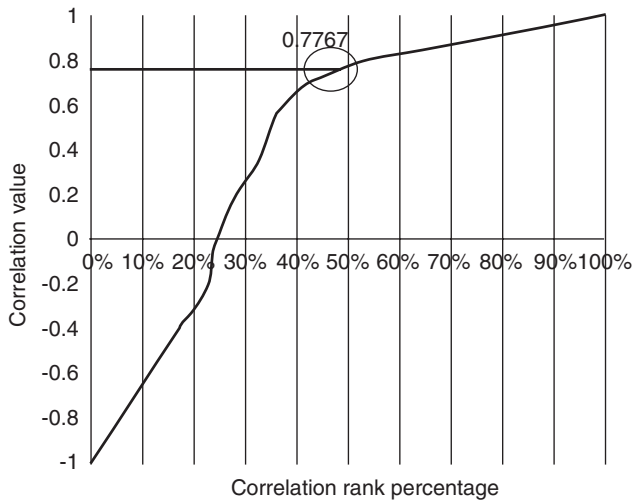


Fig 1 Distribution of correlation coefficients for selecting candidate regulators by correlation threshold method

the pool of regulators. This implies that, if, the correlation between a regulatory TF and the TG is more than threshold value, then the following regulatory TF is selected as a candidate regulator for that TG. After selecting candidate regulators by correlation threshold, these TGs and their candidate regulators are combined to fit the model, which is based on the MLR approach and the regression coefficients are estimated.

Zhang *et al.* (2010) considered the following model for time series data:

$$X_{t+1} = AX_t + \varepsilon_t$$

Here  $X_t$  is a vector describing the expression level of  $n$  different genes at time  $t$ .  $A$  is an  $n \times n$  matrix, where each entry of  $A$  models the regulatory ability of  $j$ -th gene to  $i$ -th gene.  $\varepsilon_t$  is used to model the noise at time  $t$ .

The equivalent model of the above model for the cross sectional time-series gene expression can be written as

$$Y B X + \varepsilon \tag{1}$$

Let's denote  $y'_i$  be the  $i$ -th row of  $Y$  and  $b'_i$  be the  $i$ -th row of  $B$ . Then by simple observation, we have

$$y'_i = b'_i x_i \text{ or } y_i = x'_i b_i \tag{2}$$

Adding error term in (2), this looks like the form of standard MLR equation and is given by: model

$$y_i = x'_i b_i + \varepsilon_i \quad \begin{matrix} i = 1, 2, \dots, m \\ j = 1, 2, \dots, n_i \\ k = 1, 2, \dots, l \end{matrix}$$

where,  $y_i = l \times 1$  vector of expression values for  $i$ -th TG,  $X_i = n_i \times l$  matrix of expression values of TFs for  $i$ -th TG,  $b_i = n_i \times l$  vector of regulatory strengths,  $\varepsilon_i = l \times 1$  vector of random errors and  $\varepsilon_i \sim N(0, \sigma^2 I)$ .

If the standard estimation procedure like Maximum Likelihood Method is applied to the above model, the estimate of  $b_i$  can be obtained as,

$$b^{\wedge}_i = (x_i x'_i)^{-1} x_i y_i \tag{3}$$

But the normal regression equation does not work as  $x'_i$  has rank smaller than or equal to  $(l-1)$  and is therefore singular. Therefore, the estimation of  $b_i$ , we considered SVD of the matrix,  $x'_i = U \Sigma V'$ , where  $U$  and  $V$  have  $u_i$  orthonormal columns (left singular vectors) and  $v_i$  (right singular vectors) and  $\Sigma$  is a diagonal matrix with diagonal entries are assumed to be arranged in descending order. Here  $U$  and  $V$  satisfy with

$$U U' = U U^{-1} = V V' = V V^{-1} = I \tag{4}$$

Applying SVD in the normal Equation (3), it becomes:

$$b^{\wedge}_i = V \Sigma^{-1} U' y_i \tag{5}$$

This procedure is repeatedly done for all TG. This model suits well when the TG have equal or unequal number of regulators. In equation (5),  $b_{ij}$  is the impact value of influence from gene  $j$  to gene  $i$ . The property of regression co-efficient can be used to construct networks with directivity.

$$b_{ij} = \begin{cases} > 0: j\text{-th TF acts as a upregulator for } i\text{-th target gene} \\ = 0: j\text{-th TF acts as a regulatory role for } i\text{-th target gene} \\ < 0: j\text{-th TF acts as a downregulator for } i\text{-th target gene} \end{cases}$$

Based on the above regulatory index, the GRN are constructed by using RCytoscape package (Shannon *et al.* 2013) of R.

The regulatory strengths of all the possible interactive candidate regulators on TG are found by using regression and SVD approach. But it is still unknown, exactly how significantly the regulatory ability can be regarded as a true potential regulator. In order to decide this, t- test is used for evaluating the significance of our model parameters to prune the unnecessary gene interactions.

$$t = \frac{b_i}{SE(b_i)} \sim t_{(n-2), \alpha}$$

where,  $b_i$ : regulatory strength for  $i$ -th target gene and  $S.E(b_i)$ : standard error of  $b_i$ .

The significant interactions between the regulatory TFs and TG were selected at 1% level of significance. The GRN is constructed by retaining all those significant interactions.

## RESULTS AND DISCUSSION

### Description of target genes

For this study, ten genes are identified as TGs. Among the selected TGs, 'LOC\_Os09g31019' encoding ubiquitin protein, plays an important role for resistant to salt stress in rice seedling (Liu *et al.* 2012), 'LOC\_Os06g36850' encoding Pyridoxal-phosphate dependent enzyme family protein responsible for salt tolerant in a halophytic bacteria *Exiguobacterium acetylicum* (Rajendran *et al.* 2009), 'LOC\_Os10g39360' which encodes Eukaryotic aspartyl protease family protein (EDPG) responsible to salt stress in *Arabidopsis thaliana* (Wang *et al.* 2013) and 'LOC\_Os11g18870' represents a phosphatidylethanolamine

binding family protein that was up regulated at different time points in *Hordeum vulgare* L. (Walia *et al.* 2005). The other TGs like 'LOC\_Os04g27980', 'LOC\_Os07g12240' and 'LOC\_Os01g36560.1' has regulatory roles in salinity stress in rice (Walia *et al.* 2005).

#### Construction of gene regulatory networks

Based on the 10 TGs and their potential candidate regulators, GRN is constructed at 1% levels of significance for salinity stress in rice. In these networks, directed graphs are considered with genes, represented by nodes and the edges are represented by different regulation types.

GRN is laid out by considering significant interactions between the genes at 1% level of significance, which has 110 nodes and 173 edges and has clearly 12 gene hubs (which includes ten TGs along with LOC\_Os08g09900 and LOC\_Os05g14370), these nodes are highly interconnected with each other. Further, to validate the genes (nodes) in *in silico* condition, gene ontology (GO) enrichment analysis is performed by using Database for Annotation, Visualization and Integrated Discovery (DAVID), a biological knowledgebase and analytic tool aimed at systematically extracting biological meaning from large gene lists (Huang *et al.* 2008). The gene enrichment results are listed in Table 1.

In molecular functions, the chosen genes are over represented in the categories of DNA binding, Transcription regulator activity, Transcription factor activity, Sequence-specific DNA binding and also another chosen gene sets are represented in the categories of zinc ion binding, Transition metal ion binding, metal ion binding, cation binding and ion binding all of which may be active due to the high ion concentration in salinity. In biological processes, regulation of transcription, regulation of RNA metabolic process, transcription are over represented

followed by auxin mediated signaling pathway, response to auxin stimulus, cellular response to hormone stimulus, hormone mediated signaling, response to hormone stimulus, response to endogenous stimulus, response to organic substance.

#### Testing for scale free property of biological network

The Power law distribution is fitted in the above network to test the scale free property of biological networks and the results were represented in Table 2.

The high value of co-efficient of determination ( $R^2$ ) as well as the slope value ( $b = -1$ ) indicated that the scale free property of biological network is satisfied to the given network data (Zhang and Horvath 2005). From the facts, it can be concluded that the GRN constructed by taking TGs and their potential regulators for salinity stress in rice satisfied the scale free property of biological networks.

#### Construction of refined gene regulatory networks

In a scale-free network, small-degree nodes are the most abundant, but the frequency of high-degree nodes decreases relative slowly (Barabasi and Albert 1999). The nodes which have degrees much higher than average degree are called as hubs in the network (Albert 2005). The node-degree distribution for the constructed GRN is

Table 2 Fitted value of parameter, Co-efficient of determination ( $R^2$ ) of the power law distribution

Networks	Estimated value of parameter		Co-efficient of determination ( $R^2$ )	Correlation co-efficients ( $r$ )
	a	b		
Network at 1% LS	67.11	-1.064	0.911	0.955

Table 1 GO term enrichment analysis of selected nodes by DAVID

GO Terms	Functions	Count	Ontology	P-Value	Benjamini
GO:0006355	Regulation of transcription	50	BP	5.50E-42	1.20E-40
GO: 0003677	DNA binding	49	MF	1.90E-26	6.00E-25
GO:0006355	Regulation of transcription, DNA-dependent	34	BP	2.50E-25	2.70E-24
GO:0051252	Regulation of RNA metabolic process	34	BP	2.90E-25	2.10E-24
GO:0030528	Transcription regulator activity	36	MF	4.70E-24	7.60E-23
GO:0003700	Transcription factor activity	29	MF	1.50E-21	1.60E-20
GO:0006351	Transcription	30	BP	3.40E-21	1.90E-20
GO: 0000156	Two component response regulator activity	3	MF	1.40E-02	8.90E-02
GO:0000160	Two component signal transduction system (Phosphorelay)	3	BP	1.90E-02	8.30E-02
GO: 0008270	Zinc ion binding	15	MF	2.00E-02	1.00E-01
GO: 0046914	Transition metal ion binding	15	MF	5.40E-01	9.60E-01
GO: 0046872	Metal ion binding	15	MF	8.50E-01	1.00E-1
GO: 0043169	Cation binding	15	MF	9.10E-01	1.00E-1
GO: 0043167	Ion binding	15	MF	9.10E-01	1.00E-1

“GO term” is the Gene ontology term. Ontology “BP” indicates Biological Process, Ontology “MF” indicates Molecular Function, and Ontology “CC” indicates Cellular Component. “Count” is the number of genes in set with annotation. “P-value” is the Fisher Exact P-value enrichment test. The smaller P-value is more enriched.

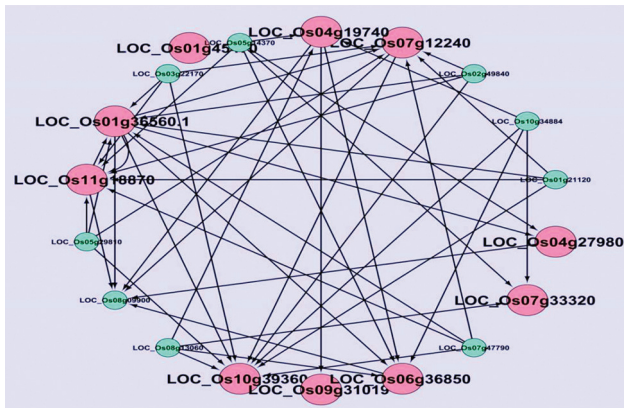


Fig 2 Refined GRN for salinity response in rice at 1% level of significance

obtained to identify perturbed hub gene interactions that induce salinity stress in rice. The nodes whose connection degree is greater than the average connection degree of the network may be considered as possible molecular target for breeding rice cultivar which may be tolerant in salinity stress condition. By taking the list of nodes whose connection degrees are greater than or equal to 4 (average connection degree), the refined GRN is constructed for salinity stress in rice.

In Fig 2, refined GRN at 1% level of significance contained 19 nodes with 48 edges and there were 8 gene hubs (LOC\_Os01g36560.1, LOC\_Os10g39360, LOC\_Os11g18870, LOC\_Os07g12240, LOC\_Os04g19740, LOC\_Os06g36850, LOC\_Os08g09900 and LOC\_Os05g14370) which were highly connected.

#### Predicting the target gene responses

To check the validity of the developed model, the responses of ten TGs are predicted by using information on regulators in microarray data with accession number GSE 21651 under salt tolerant and salt sensitive conditions and are represented in Fig 3.

The model based on multiple regression approach is used to generate the TG expression value via the gene expression profiles of the TFs from GSE 21651. The Fig 4 successfully showed that the above model predicted the responses of TGs under salt tolerant condition and salt sensitive with mean squared error 0.4578 and 0.6588 respectively.

In this work, the main aim is the construction and modelling of GRN of salinity stress response in rice. As there are 51 microarrays collected over four different microarray experiments in contrast to tens of thousands of probe sets, it is a great challenge to determine feature selection on this small sample, but high dimension data. Combining the Pearson's correlation selection algorithm and the t-test, a novel algorithm is created to select potential TFs from the pool of candidate TFs without loss of biological meaning. The small sample and large dimension problem narrows down the applicability of classical statistical techniques like regression in large genomic data

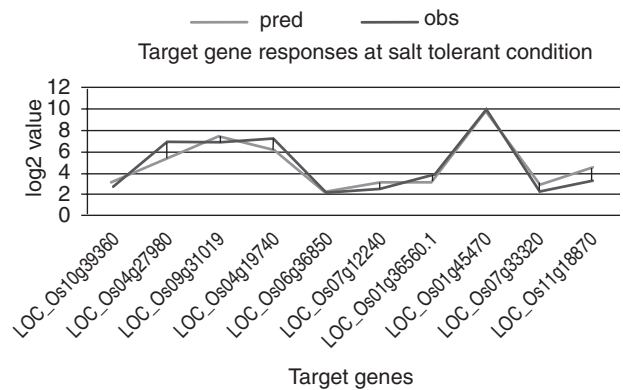
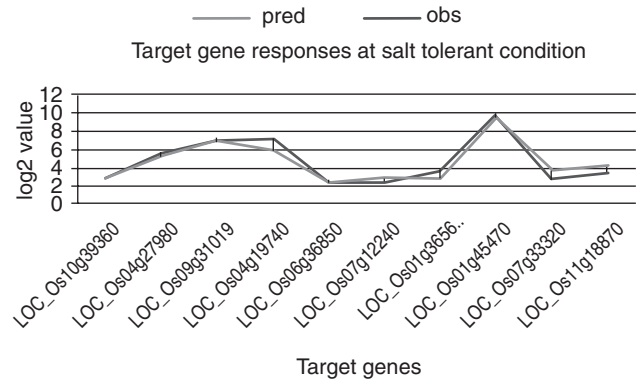


Fig 3 Predicted and observed responses (in log<sub>2</sub> value) of ten TGs under salt tolerant and sensitive condition

sets. This problem was tackled by using SVD along with MLR for estimating the regulatory strengths. The value of mean square error showed that the developed model can be successfully used for predicting the target gene responses under salt stress condition in rice. For testing the scale free property of biological networks, the Power law distribution is fitted to the constructed networks. The results showed that this approach of constructing GRN satisfied this property. For getting a clear idea about the salinity stress response mechanism in rice, the refined GRN is constructed by taking hub nodes. The refined networks did not satisfy the scale free property of biological networks, because the sub-networks of scale free networks are not scale free (Stumpf *et al.* 2005). From the GO analysis, it is found that most of the hub genes are involved in biological processes like transcription, regulation of transcription, DNA-dependent process, and regulation of RNA metabolic process. For the molecular function category, they are involved in DNA binding, transcription factor activity, transcription regulator activity, sequence-specific DNA binding and followed by DNA binding, zinc ion binding, ion binding, cation binding, metal ion binding, transition metal ion binding, which are due to high ion concentration in soil and water due to salt stress.

In this paper, a method is proposed for construction and modelling of gene regulatory network based on MLR and SVD approach. It is based on database information to construct salinity induced gene regulatory networks in rice.

It is also called a system biology approach for construction of GRN as we process the complex gene network of numerous genes and regulators from different sources at the same time. This approach of construction of gene network considers the reduction of the redundancy of the complex networks step by step. This approach of gene regulatory network construction satisfies the condition of scale free property of biological networks. The validity of the model over other salinity stress rice microarray data set is quite satisfactory. The hubs of gene interactions present in the stress induced gene networks in rice at different levels of significance need to be validated.

Our study may be helpful to set up hypotheses for researchers to design experiments for studying salinity stress response and some guidance for molecular breeders to improve traits. Since some target genes and their potential regulatory TFs have been identified in our study, which means researchers could conduct experiments to clone and validate these genes. The developed model may be used to predict the gene responses in case of missing observations. In our future research work, we will compare the performance our model with some proposed models such as Bayesian, Dynamic Bayesian and Boolean network model by using indices like AUC (Area under curve). Further, the developed computational approaches will be used to study the various stress mechanism in other plant species.

#### ACKNOWLEDGMENTS

Authors acknowledge for the help from ICAR and NIBMG (Government of India).

#### REFERENCES

- Akbar M and Ponnampereuma F N. 1980. Saline soil of South and Southeast Asia as potential rice lands, Rice Research Strategies for the Future. IRRI, Manila, Philippines, pp 265–81.
- Akutsu T, Miyano S and Kuhara S. 2000. Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics* **16**: 727–34.
- Albert R. 2005. Scale-free networks in cell biology. *Journal of Cell Science* **118**: 4 947–57.
- Allakhverdiev S I, Sakamoto A, Nishiyama Y, Inaba M and Murata N. 2000b. Ionic and osmotic effects of NaCl-induced inactivation of photosystems I and II in *Synechococcus* sp. *Plant Physiology* **123**: 1047–56.
- Bansal M, Belcastro V, Ambesi-Impiombato A and di Bernardo D. 2007. How to infer gene networks from expression profiles. *Molecular Systems Biology* **3**: 78.
- Barabasi A L and Albert R. 1999. Emergence of scaling in random networks. *Science* **286**: 509–12.
- Boros E, Ibaraki T and Makino K. 1998. Error-free and best-fit extensions of partially defined Boolean functions. *Information and Computation* **140**: 254–83.
- Bray E, Bailey S J and Weretilnyk E. 2000. Responses to abiotic stresses. (In) *Biochemistry and Molecular Biology of Plants*, pp 1158–1203. American Society of Plant Biologists.
- Chen B S, Yang S K, Lan C Y and Chuang Y J 2008. A systems biology approach to construct the gene regulatory network of systemic inflammation via microarray and databases mining. *BMC Medical Genomics* **1**: 46.
- Ching W, Fung E, Ng N and Akutsu T. 2005. On construction of stochastic genetic networks based on gene expression sequences. *International Journal of Neural Systems* **15**: 297–310.
- Cotsaftis O, Plett D, Johnson A A and Walia H. 2011. Root-specific transcript profiling of contrasting rice genotypes in response to salinity stress. *Mol. Plant* **4**(1): 25–41.
- de Jong H. 2002. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology* **9**: 69–103.
- Gautier L, Cope L, Bolstad B M and Irizarry R A. 2004. Affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**: 307–15.
- Gentleman R C, Carey V J, Bates D M, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y and Gentry J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**: R80.
- Hastie T J. 1992. Generalized additive models. (In) *Statistical Models*. Chambers J M and Hastie T J (Eds). Wadsworth & Brooks/Cole.
- Haury A C, Mordélet F, Vera-Licona P and Vert J P. 2012. TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. *BMC Systems Biology* **6**: 145.
- Hirose O, Nariai N, Tamada Y, Bannai H, Imoto S and Miyano S. 2006. Estimating gene networks from expression data and binding location data via Boolean networks. *Lecture Note in Computer Scineces* **3480**: 349–56.
- Huang D W, Sherman B T and Lempicki R A. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**: 44 – 57.
- Huang T, Liu L and Xie L. 2010. Using GeneReg to construct time delay gene regulatory networks. *BMC Research Notes* **3**: 142–52.
- Ideker T E, Thorsson V and Karp R M. 2000. Discovery of regulatory interactions through perturbation: Inference and experimental design. *Proceedings of Pacific Symposium Biocomputing* **5**: 302–13.
- Ihaka R and Gentleman R. 1996. R: A language for data analysis and graphics. *Journal of Computation Graphics and Statistics* **5**: 299–314.
- Irizarry R A, Hobbs B, Collin F, Beazer-Barclay Y D, Antonellis K J, Scherf U and Speed T P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249–64.
- Jain M, Tyagi A K and Khurana, J P. 2008. Genome-wide identification, classification, evolutionary expansion and expression of homeobox genes in rice. *FEBS Journal* **275**: 2 845–61.
- Joseph E A and Mohanan K V. 2013. A study on the effect of salinity stress on the growth and yield of some native rice cultivars of Kerala state of India. *Agriculture, Forestry and Fisheries* **2**(3): 141–50.
- Liu C W, Hsu Y K, Cheng Y H, Yen H C, Wu Y P, Wang C S and Lai C C. 2012. Proteomic analysis of salt-responsive ubiquitin-related proteins in rice roots. *Rapid Commun. Mass Spectrom* **26**(15): 1 649–60.
- Narciso J and Hossain M. 2002. *World Rice Statistics*. IRRI, Manila Philippines.
- Noda K, Shinohara A, Takeda M, Matsumoto S, Miyano S and Kuhara S. 1998. Finding genetic network from experiments

- by weighted network model. *Genome Informatics* **9**: 141–50.
- Pandit A, Rai V, Bal S and Sinha S. 2010. Combining QTL mapping and transcriptome profiling of bulked RILs for identification of functional polymorphism for salt tolerance genes in rice (*Oryza sativa* L.). *Molecular Genetics and Genomics* **284**(2):121–36.
- Priya P and Jain M. 2013. RiceSRTFDB: A database of rice TF-containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. *Database*, Article ID bat027, doi:10.1093/database/bat027.
- Rahimi A and Biglarifard A. 2011. Influence of NaCl salinity and different substrates on plant growth, mineral nutrient assimilation and fruit yield of strawberry. *Not Bot Horti Agrobo* **39**(2): 219–26.
- Rajendran N, Smith C and Mazhawidza W. 2009. Molecular and phylogenetic analysis of pyridoxal phosphate-dependent acyltransferase of *exiguobacterium acetylicum*. *Z. Naturforsch* **64**: 891–8.
- Shannon P T, Grimes M, Kutlu B, Bot J and Galas D J. 2013. RCytoscape: tools for exploratory network analysis. *BMC Bioinformatics*, **9**(14): 217.
- Shmulevich I, Saarinen A, Yli-Harja O and Astola J. 2002. Inference of genetic regulatory networks under the best-fit extension paradigm. (In) *Computational and Statistical Approaches To Genomics*. Zhang W and Shmulevich I (Eds). Kluwer, Boston, MA.
- Stumpf M P H, Wiuf C and May R M. 2005. Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proceedings of National Academy of Science USA* **102** (12): 4 221–4.
- Tanji K K. 1990. Nature and extent of agricultural salinity. *Agricultural Salinity Assessment and Management* **71**: 1–17.
- Walia H, Wilson C, Ismail A M and Close T J. 2009. Comparing genomic expression patterns across plant species reveals highly diverged transcriptional dynamics in response to salt stress. *BMC Genomics* **10**: 398.
- Walia H, Wilson C, Condamine P, Liu X, Ismail A M, Zeng L, Wanamaker S I, Mandal J, Xu J, Cui X and Close T J. 2005. Comparative transcriptional profiling of two contrasting rice genotypes under salinity stress during the vegetative growth stage. *Plant Physiology*. **139**(2): 822–35.
- Wang M, Verdier J, Benedito V A, Tang Y, Murray J D, Ge Y, Becker J D, Carvalho H, Rogers C, Udvardi M and He J. 2013. LegumeGRN: A gene regulatory network prediction server for functional and comparative studies. *Plos One* **8**(6): e64929. doi:10.1371/journal.pone.0064929.
- Wang X, Bojing D, Liu M, Sun N and Qi X. 2013. Arabidopsis transcription factor wrky33 is involved in drought by directly regulating the expression of *cesa8*. *American Journal of Plant Sciences* **4**: 21–7.
- Wu R and Garg A. 2003. Engineering rice plants with trehalose-producing genes improves tolerance to drought, salt and low temperature. ISBN News Report <http://www.isb.vt.edu>.
- Wu W S, Li W H and Chen B S. 2006. Computational reconstruction of transcriptional regulatory modules of the yeast cell cycle. *BMC bioinformatic*, **7**: 421.
- Yeung K and Ruzzo W. 2001. An empirical study on principal component analysis for clustering gene expression data. *Bioinformatics* **17**: 763–74.
- Zhang S Q, Ching W K, Tsing N K, Leung H Y and Guo D. 2010. A new multiple regression approach for the construction of genetic regulatory networks. *Artificial Intelligence in Medicine* **48**:153–60.
- Zhang B and Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Statistics, Applied Genetics Molecular Biology* **4**: 17–42.