Agri Gene 1 (2016) 126–134

Contents lists available at ScienceDirect

Agri Gene

journal homepage: www.elsevier.com/locate/aggene

A meta-analysis of potential candidate genes associated with salinity stress tolerance in rice

Sukhdeep Kaur [a], M.A. Iquebal [a], Sarika Jaiswal [a], Gitanjali Tandon [a], R.M. Sundaram [b], R.K. Gautam [c], K.P. Suresh [d], Anil Rai [a], Dinesh Kumar [a,*]

[a] Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110012, India
[b] Crop Improvement Section, ICAR- Indian Institute of Rice Research, Rajendranagar, Hyderabad 500030, Telengana, India
[c] Field Crop Improvement & Protection Division, ICAR-Central Island Agricultural Research Institute, Port Blair 744101, Andaman & Nicobar Islands, India
[d] ICAR-NIVEDI, Hebbal, Bengaluru 560024, Karnataka, India

ARTICLE INFO

Article history:
Received 29 March 2016
Received in revised form 25 July 2016
Accepted 1 August 2016
Available online 20 August 2016

Keywords:
Microarray
Rice
Salinity tolerance
Transcription factors
RNA-Seq data

ABSTRACT

Even though cultivated rice is highly sensitive to salinity, significant variability exists in the primary and secondary gene-pool of rice with respect to traits of salinity tolerance. Breeding salinity tolerance rice varieties is imperative due to climate change and increasing rice demand for global population. A meta-analysis of plethora of genomic data and published literature available on various genes/factors associated with response to rice salinity and tolerance can be used to enlist selected candidates genes affecting salinity. Such genes can be utilized to identify potential candidate salinity resistance genes from donor rice genotypes and facilitate their transfer to high yielding varieties of rice through marker-assisted breeding. This approach has tremendous advantage over transgenic approach as no bio-safety or regulatory issues are involved in exploiting the variability.

Meta-analyses were performed on three datasets viz., rice microarray data of 166 series comprising of 2586 samples, 1228 published research literature in the last one and half decades and RNA-Seq data of 454 and Illumina from Sequence Retrieval Archive (SRA) at NCBI. Among microarray dataset, six salinity related series were finally selected and multi experiment analysis revealed 2289 differentially expressed genes belonging to 44 gene families. Out of these, 13 families viz., AP2-EREBP, AUX/IAA, bZIP, C2H2, bHLH, C3H, HB, HSF, MYB, MYB-related, NAC, Tify and WRKY were selected. Applying various parameters on the published literature data, 13 genes were selected, of which five were common to the different microarray datasets. From RNA-Seq data, total of 751 differentially expressed genes were obtained from 21 gene families, out of which 11 genes were common with those obtained from microarray data and five genes, viz., AP2-EREBP/DREB, MYB, HSF, bZIP and NAC were common to all the three data sets. Based on the results obtained, a total of 31 meta-analyzed genes have been selected and recommended for use in genetic improvement programs aimed at salinity resistance in rice.

The meta-analysis of microarray, RNA-Seq and published literature has been successfully used to select 31 best salinity tolerance associated genes which can be exploited by candidate gene approach for targeted introgression through marker assisted breeding. This approach has multi-fold advantages, as it obviates statutory and ecological issues. Such endeavors are more warranted for combating the key abiotic stresses like salinity, whose effects are increasing due to a changing climate.

1. Introduction

Rice is one of the leading food crops in the world, grown over 154 million ha, 85% of total rice produced is consumed by humans. It provides 21% of global human per capita energy and 15% of per capita protein. Rice also provides minerals, vitamins, and fiber, although all constituents except carbohydrates are reduced by milling. Rice eaters and growers constitute the bulk of the world's poor

Abbreviations: TF, Transcription Factor; DEGs, Differentially Expressed Genes.

* Corresponding author.
E-mail addresses: sukhii.deep@gmail.com (S. Kaur), jiqubal@gmail.com (M.A. Iquebal), aijaiswal@gmail.com (S. Jaiswal), gitanjali.tandon@gmail.com (G. Tandon), rms@drricar.org (R.M. Sundaram), rgautampb@gmail.com (R.K. Gautam), sureshkp97@gmail.com (K.P. Suresh), anilrai645@gmail.com (A. Rai), dineshkumarbhu@gmail.com (D. Kumar).

http://dx.doi.org/10.1016/j.aggene.2016.08.001
2352-2151/

(UNDP Human Development Report, 1997), approximately 70% of the world's 1.3 billion poor people live in Asia, where rice is the staple food. The salt-affected soils are estimated to cover about 1.0 billion ha in the world (Massoud, 1974; Ponnamperuma, 1984). It has also been estimated that about 20% of all irrigated lands in the world are affected by excessive salts (Pitman and Läuchli, 2002). Salt affected soils encompass two types i.e. sodic and saline soils which together occupy about 6.73 m ha land in India and thus adversely affect crop productivity (Sharma et al., 2004). Though rice (*Oryza sativa* L.) is the most important crop and feasible option to start with crop cultivation and reclamation in such soils (Tyagi, 1998; Ismail et al., 2007; Singh et al., 2010), rice is also expected to suffer the most due to salt stress conditions particularly in countries with long sea shore line (Swaminathan and Kesavan, 2012).

It has been found that if root zone salinity exceeds from its threshold (i.e., lower to upper ranging from 3 dS/M to 11.3 dS/M, respectively), then plant growth ceases leading to yield loss of 12% per dS/M (Raes et al., 2012), with seedling and flowering stages being the most sensitive stages. Rice happens to be one of the most susceptible crop species especially to salinity (Grattan et al., 2002; Munns and Tester, 2008) and even with little amount of 50 mM NaCl (Yeo and Flowers, 1986) can affect adversely. Rice productivity in salt-affected areas is as low as <1.5 t/ha (Hairmansis et al., 2014).

The total rice requirement of world population is going to be 800 mT by 2025 (Kubo and Purevdorj, 2004), while in Asia, rice demand is expected to increase by 70% in coming 30 years coupled with population growth (Muthayya et al., 2014). This is likely to face twin facet problem, i.e., increase in the area under salinity and decrease in yield per unit area. Salinity reclamation has its own limitation, thus genetic improvement of rice for better salinity tolerance is imperative. Rice genome has been deciphered, and is recorded to possess >35 K genes (Goff et al., 2002). In order to exploit candidate gene approach, genes related to salinity response and tolerance must be enlisted and rationally prioritized. In the last decade, more than thousand research papers have been published across the globe related to salinity stress in rice and other crops and related data with respect to differential gene expression experiments are enriched in various public databases. Availability of such plethora of data in knowledge discovery research has led to enigmatic question that how many genes can be enlisted with prioritization indices in order so that they can be targeted for genetic improvement of rice for salinity tolerance through marker-assisted breeding.

There are two methods of genetic improvement viz., transgenic approach and candidate gene approach. Both the approaches have been used in rice salinity resistance breeding programs. In transgenic approach, for example e.g. SOD genes from yeast (Tanaka et al., 1999) and catalase genes (Motohashi et al., 2010) from E coli have been used to increase salinity resistance. Among candidate genes approach, for example, FL478 was used as a donor parent to introgress the Saltol QTL conferring salt tolerance into BT7 (Linh et al., 2012) and the candidate genes underlying the QTL have been identified and validated (Thomson et al., 2010). In case of candidate gene approach, there are no regulatory or other issues concerning environment, food and feed safety as the donor and recipients belong to the rice gene pool with hybridization being the key factor in mobilization of genes. There are well known salt tolerant donors varieties like Pokkali, Nona Bokra etc. (Moons et al., 1995). If a set of potential candidate salinity tolerance associated genes can be identified through meta-analysis, they can be validated and utilized in the rice breeding programs.

Meta-analysis, which is quantitative review of related but independent studies (Normand, 1999; Hunter and Schmidt, 2004) can be used as an important tool for knowledge discovery. Meta-analysis approach can resolve the issues related to identification of key of salinity tolerance associated genes and their prioritization so that marker assisted introgression programs can be gainfully be initiated. Though limited attempt of meta-analysis of abiotic stress in rice has been carried out for different traits like drought (Trijatmiko et al., 2014; Swamy et al., 2011; Khowaja et al., 2009; Shaik and Ramakrishna, 2013), biomarker search (Zimmermann et al., 2008), biotic stresses like rice blast (Ballini et al., 2008) and bacterial blight (Shaik and Ramakrishna, 2013), meta-analysis of rice salinity tolerance associated genes has not been attempted. A meta-analysis platform of rice is now available (McLaren et al., 2005). Thus there is a need to identify salinity tolerance related key candidate genes by meta-analysis using all existing datasets (both published literature and public databases).

The present work aims at meta-analysis of potential rice salinity tolerance associated candidate genes from publicly available experimental data (both microarray and RNA-Seq datasets) and published literature.

2. Materials and methods

Microarray data for salinity stress was retrieved from Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/). The microarray datasets from affymetrix platform Rice genome Array (GPL2025) were downloaded. Out of 166 series, 9 series were available for salt stress viz., GSE3053, GSE4438, GSE6901, GSE11175, GSE13735, GSE14403, GSE16108, GSE21651, GSE28209. The microarray dataset workflow is represented in Fig. 1. To maintain the uniformity of dataset, only 6 series datasets were used in the present study, excluding GSE3053, GSE4438 and GSE28209 from the working dataset (Table S1a). The RNA-Seq data (454 and Illumina) of rice salinity with accessions DRX000191, DRX000192, DRX000194, SRX145290, SRX144646, SRX144647, SRX145289, SRX144645, SRX145288, SRX144644, SRX145287 from Sequence Retrieval Archive (SRA) at NCBI were used, details of which are given (Table S1b). These were further used for meta-analysis of rice salinity, the workflow of which is illustrated in Fig. 2.

2.1. Identification of differentially expressed genes from microarray dataset

After data collection, data normalization was done using quantile method and for summarization step (Irizarry et al., 2003), median polish method was applied at RMAExpress (Bolstad, 2012). The RMA preprocessed files were used to identify differentially expressed genes using RankProd method in MultiExperiment Viewer (Saeed et al., 2003), which is a non-parametric method. In this method, genes were ranked according to their up and down regulation, fold change and p-values (Breitling et al., 2004). Rank product method combined the datasets from various studies that were carried out with different criteria like number of permutation set as 250, p-value cutoff <0.05, FDR (Benjamini and Hochberg, 1995) at 0.05 and fold change >2. Genes that satisfied all the cutoff thresholds were considered as differentially expressed genes (DEGs) (Hong and Breitling, 2008). Different number of DEGs were identified from each experiment and collected in single non redundant stress signal file. This file contains 2289 genes which were considered for further analysis.

2.2. Gene set enrichment analysis for microarray dataset

Functional enrichment analysis was done by AgriGO (Du et al., 2010b) followed by transcription factor database RiceSRTFDB (Priya and Jain, 2013). Enrichment analysis was done to characterize and combine specific pathways, domains in genes. Salinity responsiveness genes were searched. Gene ontology study was performed on the obtained 2289 genes in non-redundant stress signal file to find out enriched GO terms in molecular function, biological process and cellular components contributing in salinity response. To further prove salinity response we collected all salinity responsive transcription factors from RiceSRTFDB and comparatively analyzed them with the enriched GO IDs to annotate the all genes. Annotating genes and factors were further analyzed and selection was made for those with frequency/occurrence greater than five times. Analyzing 2289 genes, shows 44 types of