Genome-wide Analysis of Putative *Erf*and *Dreb* Gene Families in Indica Rice (*o. Sativa* l. Subsp. *Indica)*

Hemalatha N (First Author)
Assistant Professor, AIMIT
St.Aloysius college
Mangalore
Karnataka, India
hemasree71@gmail.com

Rajesh M.K (Author)
Central Plantation Crops
Research Institute
Kasargod
Kerala India
mkraju_cpcri@yahoo.com

Narayanan N.K(Author)
School of Information Science
and Technology
Kannur University,Kannur
Kerala India
csirc@rediffmail.com

***Abstract*— Drought is a major constraint to rice production and its stability in rain-fed and poorly irrigated environments. Identifying genomic regions influencing the yield and its response to water deficits will aid in our understanding of the genetics of drought tolerance and development of more drought tolerant cultivars. Besides drought, the other major impediment to increased crop production is salt stress. In this context, identification of drought and salt-responsive genes assumes significance.**

In this paper we carried out genome-wide analyses to explore putative genes encoding ethylene responsive factor (ERF) and dehydration-responsive element binding proteins (DREB) in the genome of indica rice. Reference nucleotides of well established molecular function, representing each of the protein families investigated, were chosen as query sequences for searches in the indica rice genome database. Clones having genomic sequences similar to the related genes were taken and converted to amino acid sequences. Putative sequences were subjected to PROSITE and Pfam databases and 31 signature sequences related to ERF family and 30 sequences related to DREB were obtained. Proteins showing more than 30% identity were taken and phylogenetic trees were generated for each family. The results of this sudy provide basic genomic information about new *ERF* and *DREB* gene families in indica rice.

Keywords- Oryza sativa, genome wide analysis, DREB, ERF

I. INTRODUCTION

Rice is the staple food in India accounting for more than half of the calories consumed. It is also a major source of livelihood for farmers especially in villages with rice-based production systems providing the main income and employment for more than 50 million households in India. Rice production in developing country like India continuously faces the challenge of keeping pace with an annual population increase, while the area of fertile wetland (lowland) available for rice farming is steadily decreasing due to urbanization and industrialization. To satisfy the demand for increased production of rice in the next decades, the only alternative is to expand rice cultivation to marginal dry-land (upland) areas, where rice production is severely hampered by dehydration stress due to drought and salt stress.

Intensive research has been undertaken in the past few decades to identify drought and salt-responsive mechanisms in plants, both from a biological and genetic perspective. Transcription factors play a significant role in regulation of abiotic-stress responsive gene expression [1]. Transcription factors of the ethylene responsive factor (ERF) family contain a single AP2/ERF domain, and are sometimes further divided into two major subfamilies, the CBF (C-repeat/DRE-Binding Factor)/DREB (Dehydration Responsive Element Binding protein) subfamily and the ERF subfamily [1]. Genes in the CBF/DREB subfamily play a crucial role in the response of plants to abiotic stresses by recognizing the dehydration-responsive element (DRE) with a core motif of A/GCCGAC [2] [3].

Manifestation of several abiotic stresses such as drought, and salinity occurs via alteration of water status of plant cells and are result in cellular dehydration. Water movement across cellular membranes is regulated largely by a family of water channel proteins called aquaporins. Aquaporins contribute to drought/salt tolerance by facilitating water transport between different organs and at the cellular level by maintaining water homeostasis [4] [5]. The availability of complete genome sequence of indica rice enables genome wide analyses of gene families of abiotic-stress responsive gene expression. Several of putative proteins have not yet been assigned to any families or subfamilies in indica rice. To gain further information, we have carried out an analysis of the genomic sequences related to rice *ERF and DREB* gene families by searching the indica variety genome in public databanks. The purpose of this study was identification of new *ERF* and *DREB* genes that could play a major role in drought/salt tolerance in rice and their phylogenetic analyses.

II. MATERIALS AND METHODS

The reference sequences of well-established molecular function, representing drought and salt tolerant gene families chosen as query sequences for searches in the indica rice (*O. sativa*) genome databases were EU910896 (*Solanum lycopersicum* ethylene responsive factor *JERF3*) and EF583447 (*Glycine max* Dehydration Responsive Element Binding *DREB5*). Taking specific members of these families as query sequences, searches were made using the TBLASTN tool [16] against GenBank database non-redundant (NR), with search specifications for *O. sativa* L. subsp. *indica* cultivar. The BLAST server used was that of the National Center for Biotechnology Information

978-1-4244-9253-4 /11/$26.00 ©2011 IEEE

(http://www.ncbi.nlm.nih.gov/BLAST/). As selection criteria of BLAST hits for genomic sequences, a cut off e-value of e^{-10} was previously set. The genomic sequences found were used to predict putative genes contained within them. A mixed procedure was adopted combining *ab initio* gene prediction algorithms of genomic sequence alignments with similar sequences from expressed genes. The prediction algorithms used was GenScan (http:// genes.mit.edu/GENSCAN.html). Such expressed sequences were found by BLAST searches against EST and NR databases of GenBank, using the genomic sequence as query. The algorithm of choice for the multiple alignments of protein sequences was Clustal W2 [17]. All the proteins with greater than 30% identity, with at least one of the reference proteins used in the searches, were regarded as functionally similar (homologous) to the reference proteins [18] [19] [20] [21]. Those sequences that did not conform to this criterion were discarded.

Prediction of homology and signature sequences for the drought and salinity genes were carried out with PROSITE (http://www.ebi.ac.uk/InterProScan/) [22] and Pfam [23] databases. Sequences were included into families based on homology and presence of signature sequences. For topology prediction, HMMTOP [24] was used. Protein alignments obtained with ClustalW2 [17] were used as starting points for phylogenetic analysis.

III. RESULTS AND DISCUSSION

Understanding the mechanisms of plants tolerance to environmental stress has the potential to provide new tools and strategies to improve the tolerance to environmental stress. Until recently, functional gene dissection of plants was largely carried out by molecular characterization of individual genes from various species. With the availability of complete genome sequences, bioinformatic analyses have assisted unraveling information stored in these genomes and facilitated studies of gene evolution.

The availability of data of complete genome sequencing of indica rice has made it possible to search for new, putative drought/salt-stress responsive genes. To search for *ERF* and *DREB* family genes in indica rice, BLAST searches of indica rice genome databases was performed using *Solanum lycopersicum* ethylene responsive factor *JERF3 and Glycine max* Dehydration Responsive Element Binding *DREB5* as query sequences. After searching the databanks with TBLASTN, clones having genomic sequences to the related family were taken and converted to amino acid sequences. In each family, similar sequences were removed and the sequences were subjected to PROSITE and Pfam databases to see the presence of signature sequences for the corresponding families.

Thirty one full length cDNA clones in the genome of indica rice were identified as encoding ERF domains after these sequences were subjected to PROSITE (Table 1). All the 31 putative ERF proteins exhibited more than 30% identity with *Solanum lycopersicum* ethylene responsive factor JERF3 and were taken for the construction of phylogenetic tree . In an earlier study, [15] identified 139 ERF family genes in japonica rice (*Oryza sativa* L. subsp. *japonica*) using a comprehensive computational analysis. It has been demonstrated that the AP2/ERF proteins have important functions in the transcriptional regulation of a variety of biological processes related to growth and development, as well as various responses to environmental stimuli. Wang and Wu have demonstrated that the transcriptional regulation of JERF3 modulated the increased tolerance to drought, salt, and freezing in tobacco during germination and seedling development [13] [14]. The transcriptional regulation was two-fold: activation of the expression of osmotic stress genes and photosynthetic carbon assimilation/metabolism and oxidative genes establishing the involvement of JERF3 in a ROS-mediated regulatory module in transcriptional networks that govern plant response to stress.

Thirty full length cDNA clones in the genome of indica rice were identified as encoding DREB domains after these sequences were subjected to PROSITE (Table 2). All the 30 putative DREB proteins exhibited more than 30% identity with *Glycine max* Dehydration Responsive Element Binding DREB5 and were taken for the construction of phylogenetic tree. Increased expression of DREB genes have been implicated in increased tolerance to drought in Arabidopsis [1] [8], soybean [9], maize [11] and rice [10]. Each type of rice DREB protein is defined on the basis of demonstrating high homology with the corresponding DREB-type proteins in *Arabidopsis* in the AP2 domain [10]. Five *OsDREB* genes have been reported earlier by Dubouzet in japonica rice [10]; four of them being *DREB1*-related (*OsDREB1A-1D*) and the fifth one *DREB2*-related (*OsDREB2A*).

The detailed structural and phylogenetic analyses in this study provide insights into the presence and organization of putative *ERF* and *DREB* gene families in indica rice. The outcome of the present study provides basic genomic information for the gene families involved in response to abiotic stresses and will pave the way for elucidating the precise role of these genes in plant growth and development in the future. The unraveling of roles of individual members of these gene families in abiotic stresses tolerance will require a concerted effort by adoption of diverse approaches, including molecular genetic analysiswhen proofreading spelling and grammar:

REFERENCES

[1] Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K. 2002. DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochemical and Biophysical Research Communications 290, 998–1009.

[2] Yamaguchi-Shinozaki K, Shinozaki K. A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature or high-salt stress. Plant Cell 1994; 6:251–264.

[3] Thomashow MF. 1999. PLANT COLD ACCLIMATION: freezing tolerance genes and regulatory mechanism. Annual Review of Plant Physiology and Plant Molecular Biology 50, 571–599.

[4] Tyerman SD, Niemietz CM, Bramley H. 2002. Plant aquaporins: multifunctional water and solute channels with expanding roles. Plant, Cell and Environment 25, 173–194.

[5] Kjellbom P, Larsson C, Johansson II, Karlsson M, Johanson U. 1999. Aquaporins and water homeostasis in plants. Trends in Plant Science 4, 308–314.

[6] Sarda X, Tousch D, Ferrare K, Cellier F, Alcon C , Dupuis JM, Casse F, Lamaze T. 1999. Characterization of closely related delta-TIP genes encoding aquaporins which are differentially expressed in sun over roots upon water deprivation through exposure to air. Plant Molecular Biology 40, 179±191.

[7] Suga S, Komatsu S, Maeshima M. 2002. Aquaporin isoforms responsive to salt and water stresses and phytohormones in radish seedlings. Plant and Cell Physiology 43, 1229±1237.

[8] Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1998) Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought and low-temperature-responsive gene expression, respectively, in Arabidopsis. Plant Cell 10:1391–1406

[9] Li XP, Tian AG, Luo GZ, Gong ZZ, Zhang JS, Chen SY (2005) Soybean DRE-binding transcription factors that are responsive to abiotic stresses. Theor Appl Genet 110:1355–1362

[10] Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Os- DREB genes in rice, Oryza sativa L, encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J 33:751–763

[11] Kizis, D. and Pages, M. (2002) Maize DRE-binding proteins DBF1 and DBF2 are involved in rab17 regulation through the drought-responsive element in an ABA-dependent pathway. Plant J. 30, 679-689.

[12] Yanhui Peng, Wuling Lin, Weiming Cai and Rajeev Arora (2007) Over expression of a Panax ginseng tonoplast aquaporin alters salt tolerance, drought tolerance and cold acclimation ability in transgenic Arabidopsis plants PlantaVolume 226, Number 3 / August, 2007729-740

[13] Lijun Wu, Zhijin Zhang2, Haiwen Zhang, Xue-Chen Wang, and Rongfeng Huang Transcriptional Modulation of Ethylene Response Factor Protein JERF3 in the Oxidative Stress Response Enhances Tolerance of Tobacco Seedlings to Salt, Drought, and Freezing Plant Physiology, December 2008, Vol. 148, pp. 1953–1963

[14] Wang H, Huang Z, Chen Q, Zhang Z, Zhang H, Wu Y, Huang D, Huang R (2004) Ectopic over expression of tomato JERF3 in tobacco activates downstream gene expression and enhances salt tolerance. Plant Mol Biol 55: 183–192

[15] Toshitsugu Nakano, Kaoru Suzuki, Tatsuhito Fujimura, Hideaki Shinshi (2006).Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiology Preview January 2006 as DOI:10.1104/pp.105.073783

[16] Altschul S.F.,W. Gish, W. Miller, E.W. Myers, D.J. Lipman (1990).Basic local alignment search tool, J. Mol. Biol. 215 403–410.

[17] Thompson J.D., D.G. Higgins, T.J. Gibson (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice Nucl. Acids Res. 25 4673–4680.

[18] Orengo C.A., D.T. Jones, J.M. Thornton (1994). Protein super families and domain super folds, Nature 372 631–634.

[19] Rost B., A. Valencia (1996). Pitfalls of protein sequence analysis Curr. Opin. Biotechnol. 7 457–461.

[20] Rost B. (1999). Twilight zone of protein sequence alignments Protein Eng. 12 85–94

[21] Todd A.E., C.A. Orengo, J.M. Thornton (2001). Evolution of function in protein super families, from a structural perspective, J. Mol. Biol. 307 1113–1143.

[22] Hofmann K., P. Bucher, L. Falquet, A. Bairoch (1999). The PROSITE database and its status, Nucl. Acids Res. 27 215–219.

[23] Bateman A., E. Birney, R. Durbin, S.R. Eddy, K.L. Howe, E.L. Sonnhammer (2000). The Pfam protein families database, Nucl. Acids Res. 28 263–266.

[24] Tusnady G.E., I. Simon (1998). Principles governing amino acid composition of integral membrane proteins: application to topology prediction, J. Mol. Biol. 283 489–506.

ERF gene families found in indica rice genome orthologous to *Solanum lycopersicum JERF3* and their Genbank accession numbers.(* indicates presence of full length cDNA)

Sequence name	GenBank accession number	Full length cDNA	% identity with query
OsJERF3.1	CT834246.1	*	40%
OsJERF3.2	CT832042.1	*	41%
OsJERF3.3	CT830564.1	*	68%
OsJERF3.4	CT832041.1	*	71%
OsJERF3.5	CT831646.1	*	66%
OsJERF3.6	CT837654.1	*	58%
OsJERF3.7	CT829047.1	*	62%
OsJERF3.8	CT832402.1	*	83%
OsJERF3.9	CT837611.1	*	61%
OsJERF3.10	CT828880.1	*	44%
OsJERF3.11	CT833424.1	*	58%
OsJERF3.12	CT833260.1	*	75%
OsJERF3.13	CT833261.1	*	75%
OsJERF3.14	CT831330.1	*	67%
OsJERF3.15	CT837590.1	*	69%
OsJERF3.16	CT832482.1	*	65%
OsJERF3.17	CT832723.1	*	70%
OsJERF3.18	CT829445.1	*	55%
OsJERF3.19	CT832234.1	*	42%
OsJERF3.20	CT834242.1	*	58%
OsJERF3.21	CT831165.1	*	66%
OsJERF3.22	CT835379.1	*	37%
OsJERF3.23	CT829864.1	*	64%
OsJERF3.24	CT835592.1	*	54%
OsJERF3.25	CT831975.1	*	36%
OsJERF3.26	CT828886.1	*	48%
OsJERF3.27	CT837872.1	*	55%
OsJERF3.28	CT836522.1	*	32%
OsJERF3.29	CT828346.1	*	46%
OsJERF3.30	CT834499.1	*	56%
OsJERF3.31	CT829643.1	*	48%

Table 1.

DREB gene families found in indica rice genome orthologous to *Glycine max* Dehydration Responsive Element Binding *DREB5* and their Genbank accession numbers. (* indicates presence of full length cDNA)

Sequence name	GenBank accession number	Full length cDNA	% identity with query
OsDREB5.1	CT829047.1	*	68%
OsDREB5.2	CT830564.1	*	80%
OsDREB5.3	CT831646.1	*	83%
OsDREB5.4	CT837611.1	*	74%
OsDREB5.5	CT834246.1	*	63%
OsDREB5.6	CT837654.1	*	62%
OsDREB5.7	CT833424.1	*	59%
OsDREB5.8	CT832042.1	*	61%
OsDREB5.9	CT832041.1	*	62%
OsDREB5.10	CT828880.1	*	62%
OsDREB5.11	CT832234.1	*	63%
OsDREB5 .12	CT829445.1	*	62%
OsDREB5.13	CT832723.1	*	59%
OsDREB5.14	CT831330.1	*	65%
OsDREB5.15	CT832482.1	*	57%
OsDREB516	CT837590.1	*	65%
OsDREB5.17	CT832402.1	*	61%
OsDREB5.18	CT834242.1	*	62%
OsDREB5.19	CT833261.1	*	59%
OsDREB520	CT835379.1	*	61%
OsDREB521	CT833260.1	*	60%
OsDREB522	CT831165.1	*	60%
OsDREB523	CT829864.1	*	59%
OsDREB524	CT831975.1	*	39%
OsDREB525	CT835592.1	*	53%
OsDREB526	CT837872.1	*	51%
OsDREB527	CT836522.1	*	32%
OsDREB528	CT829643.1	*	41%
OsDREB529	CT828886.1	*	57%
OsDREB530	CT834499.1	*	56%

Table 2.