Regulatory variation in gene expression patterns in rice under drought stress: Identification of SNPs in promoter sequences of candidate genes.

Lachagari V. B. Reddy and Arjula R. Reddy*

Molecular Genetics and Functional Genomics Laboratory,
Dept. of Plant Sciences, University of Hyderabad, Hyderabad, 500 046, A.P., India and
***Department of Biotechnology, Yogi Vemana University, Kadapa, 516 003 A.P., India**

*arjulsl@yahoo.com

Abstract
Identification of genes involved in drought stress response process is a challenging task as the trait is controlled by many genes dispersed across the genome with complex regulation. In the post genomics era, identification of single nucleotide polymorphisms (SNPs) has become possible at many functionally linked genes governing the target trait. Further, SNPs having possible association with drought tolerance are invaluable to facilitate marker assisted selection (MAS). We attempted to identify such SNPs using genomic and molecular genetic approaches using eight rice cultivars with well characterized phenotype with reference to drought tolerance. A total of 200 genomic regions were sequence characterized and deposited in the public domain (dbSNP, NCBI, Handle ARR-VBREDDY). Screening of more than one million bases at the targeted sites revealed many informative sites which were mapped on to the rice genome. A majority of the SNPs i.e. 1203 are found in the promoter regions. With the SNPs identified at different targeted regions, we have attempted to develop allele sharing maps of candidate genes, and integrating with sequence and genetic map as well. Further, these SNPs were integrated with abiotic stress responsive QTLs and allele sharing maps. All such SNPs found in the promoter regions of the targeted genes were analyzed for their association with cis acting regulatory regions. Forty of these SNPs are found to be associated with various *cis* elements, and interestingly, 23 are found to be associated with abiotic stress response related elements such as *Myb, Myc, CRT/DRE, WRKY, ABRE, and CORE etc*. The implications of these observations will be discussed

Media Summary:
Identification of SNPs at gene loci, particularly, in regulatory elements, associated with drought tolerance will help in developing precise tools for molecular marker assisted selection in rice.

Keywords
Drought tolerance, SNPs, allelesharing maps, cis elements of candidate genes

Introduction
Rice, one of the most important food crop of the world, is a staple food for more than half of the world's population. It has become a model cereal crop for genomics because of its relatively small genome, availability of high density genetic and physical maps and a complete draft genome sequence. Various biotic and abiotic factors limit crop production in almost all rice growing areas. Of the abiotic stress factors, the yield losses caused by drought alone accounts to as much as 60% and reaching 100% during certain years in some areas

(Gorantla et al., 2007). Drought tolerance is a complex trait controlled by a large number of genes dispersed across the genome with a complex regulation. The performance of genotypes in under drought condition is largely attributed to genetic variations, mostly at the single nucleotide level. A crucial step towards understanding the molecular genetic basis of drought tolerance is identification, and functional study of candidate genes and their allelic variants, which requires large-scale genomic and genetic resources. In plants, particularly in grasses SNP analysis could provide valuable information on genetic control of agronomic traits with unprecedented pre*cis*ion. The abundance, ubiquity and interspersed nature of SNPs in plant genomes make them ideal candidates as molecular markers for marker-assisted plant breeding. Particularly, trait based dissection of genomic regions for allelic variations and further utilizing these variations in associating them to the trait of interest is a promising approach for complex traits. The present paper deals with SNPs at selected loci associated with regulation of drought stress response in rice.

Methods

Rice cultivars that were used in different breeding programs across India were obtained from various sources. The list represents genotypes which are phenotypically well characterized with reference to drought stress response, representing both indica and japonica cultivars. A total of 8 genotypes representing 3 highly drought tolerant, 2 moderately tolerant and 3 drought susceptible genotypes were used in SNP genotyping. A doubled-haploid line (DHL) mapping population of rice (IR68586) developed at IRRI and adopted for mapping of QTLs associated with drought resistance of rice along with parents was used in mapping. Candidate genes for drought tolerance were identified from the gene resource developed from drought stress induced seedlings of Oryza sativa cv Nagina22 (Reddy et al., 2002 and Babu et al., 2002) and microarray gene expression profiles of published literature and unpublished microarray studies (Markandeya et al., Unpublished). Rice genome sequence data was used to identify putative promoter region of targeted candidate genes. BAC/PAC clones sequence data of spanning targeted candidate gene region has been obtained by BLASTN program at NCBI and gene structure was predicted using FGENESH program. Targeting specific candidate regions of the promoter, amplification, sequencing, analysis was performed using standard bioinformatics and molecular biology techniques. Development of allelesharing maps, integration with genetic and sequence map was done using custom scripts utilizing the QTL data from Gramene.

Results

Screening of more than one million bases at the targeted sites enabled us to capture the informative sites which were categorized based on the type of nucleotide change observed. A total of 568 SNPs were identified between CT9993 and IR62266 and 494 SNPs were identified between Azucena and IR64. Interestingly the comparison of drought tolerant Nerica-1 with drought susceptible IR64 revealed 645 SNPs and with that of IR62266 revealed 673 SNPs where as the comparison of Nerica-1 with drought tolerant Azucena revealed only 337 SNPs. All the informative sites are mapped on to rice genome sequence map (IRGSP). SNP positions were identified and denoted following a unique nomenclature referring chromosome number followed by base position on the rice genome for the first time(Fig1). Though, the analysis revealed 2585 SNPs in the targeted regions, the application of Bayesian probabilistic algorithm with an appropriate cut-off (0.9 probability) limited us to identify 1640 SNPs. A total of 572 transitions, 446 transversions and 523 Indels were identified of which 310 are identified coding regions, 8 are in intronic regions and 20 are in

UTRs. Majority of the SNPs (1203) are found in promoter regions as our major focus was on 5' upstream region of the gene and this includes SNPs found in the regulatory regions of the promoters and surrogate SNPs as well. Most of the variations were observed to be a single nucleotide substitution or indels besides a few (47) multinucleotide indels (MNPs). Amongst single nucleotide changes majority of them are A/G (297) and T/C (272) changes. Rate of polymorphism (ROP) between genotypes ranged from 0.333 to 0.944. Interestingly, ROP was observed to be high between drought tolerant and susceptible varieties, when compared with in the panel suggesting tight association of the targeted regions with the trait. Further, it has been observed that a common pattern was observed with in the panel of genotypes in most of the targeted regions. We have also attempted to develop a genome wide haplotype and allele sharing map with the help of an in-house developed software tool called AS-map V1.0. This newly developed program enabled us to develop allele sharing maps of candidate genes and integrated with rice sequence map as well as genetic map. Further, in order to dissect and associate these SNPs we have integrated abiotic stress responsive QTLs to our allele sharing map(Fig2). This enabled us to identify haplotype tagged SNPs in the genomic regions associated with drought stress response, spanning abiotic stress responsive QTLs. A large number of SNPs in the QTL regions associated with relative water content (RWC), osmotic adjustment capacity (OA), leaf rolling, leaf drying, stomatal closure time, drought susceptibility index and dehydration tolerance beside large number of root related QTLs were identified (Fig3). The analysis of targeted promoter regions with SNP site at PLACE database revealed at least 40 SNPs associated with various *cis* elements. Interestingly, twenty three are found to be associated with stress response related *cis* acting elements such *as Myb, Myc, CRT/DRE, WRKY, ABRE, COR etc.*

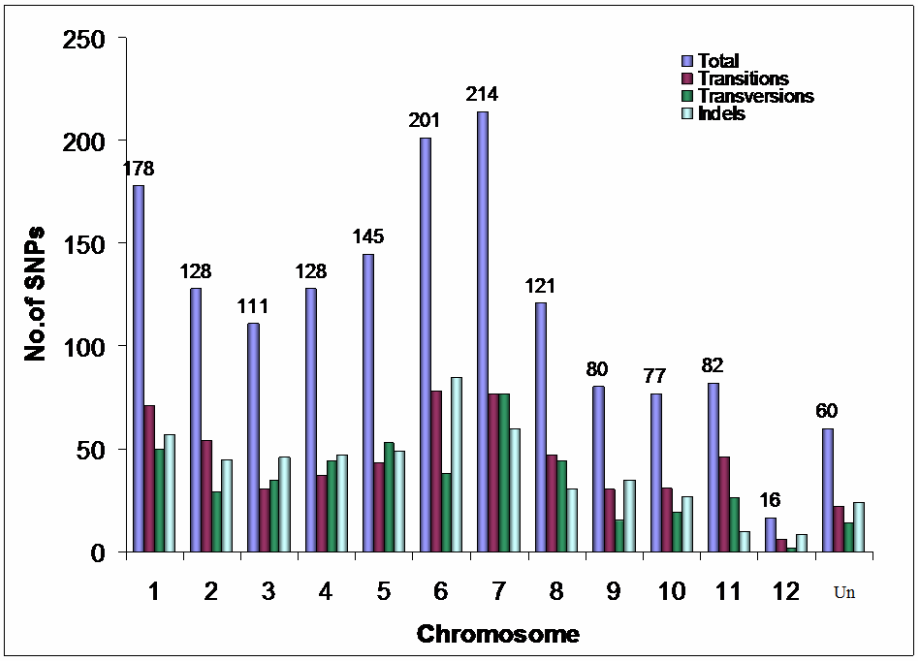


Figure 1: Classification and distribution of SNPs in different chromosomes of rice; Un-indicates SNP not mapped to any of the chromosomes

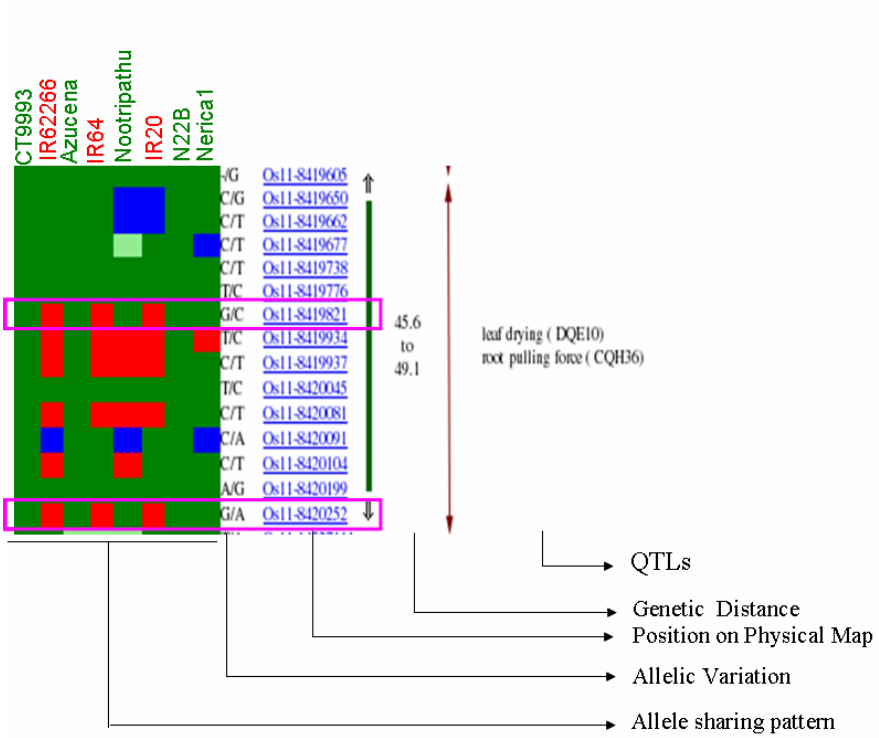
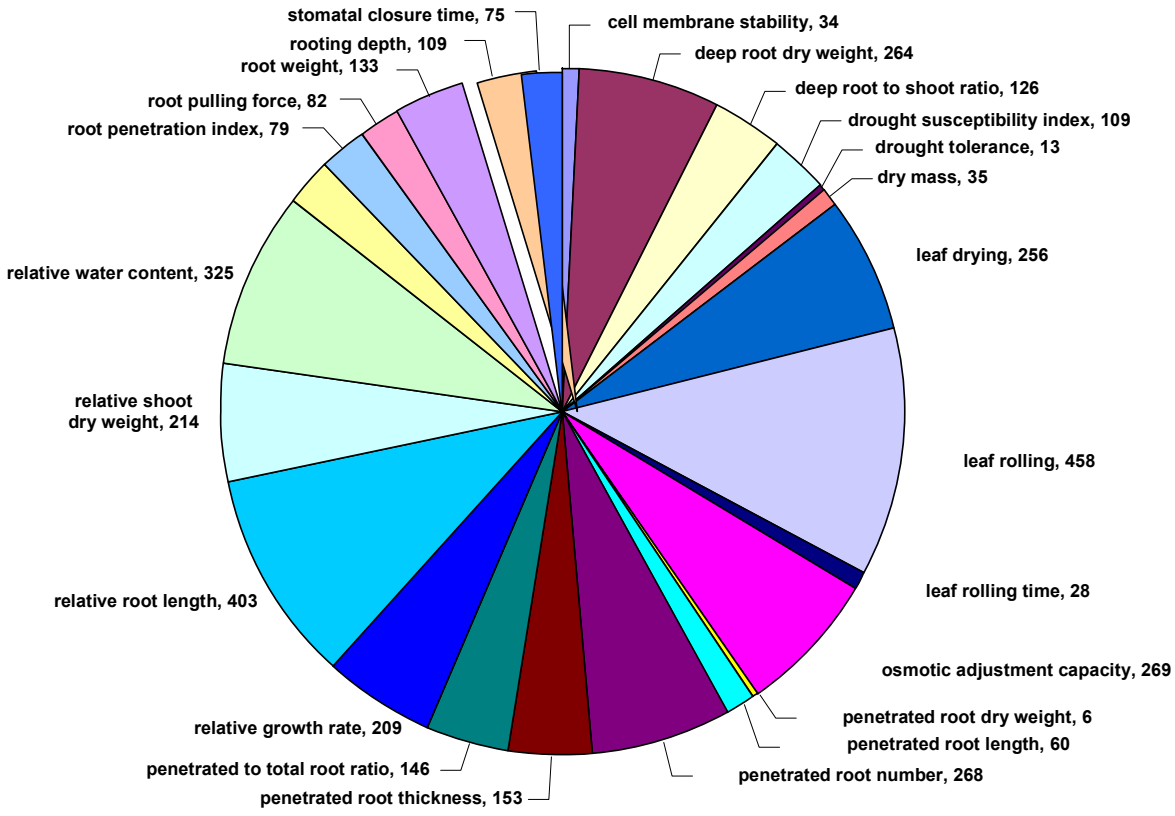


Figure 2: A snapshot of integrated allelesharing map: Poisition of SNP on IRGSP sequence map, genetic distance, and abiotic stress responsive QTLs.

Fig3. Single Nucleotide Polymorphisms identified in different abiotic stress responsive QTLs; The number after the QTL name represents number of SNPs in each QTL



(Financial support from Dept of Biotechnology, Govt. India and the Rockefeller Foundation, USA is gratefully Acknowledged)

References

Babu, P.R., Sekhar, A.C., Ithal, N., Markandeya, G., and Reddy, A.R. (2002). Annotation and BAC/PAC localization of nonredundant ESTs from drought-stressed seedlings of an indica rice. J Genet **81,** 25--44.

Gorantla, M., Babu, P.R., Lachagari, V.B.R., Reddy, A.M.M., Wusirika, R., Bennetzen, J.L., and Reddy, A.R. (2007). Identification of stress-responsive genes in an indica rice (Oryza sativa L.) using ESTs generated from drought-stressed seedlings. J Exp Bot **58,** 253--265.

Reddy, A.R., Ramakrishna, W., Sekhar, A.C., Ithal, N., Babu, P.R., Bonaldo, M.F., Soares, M.B., and Bennetzen, J.L. (2002). Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (Oryza sativa L. subsp. indica cv. Nagina 22). Genome **45,** 204--211.