Indian J. Genet., 77(1): 32-41 (2017)
DOI: 10.5958/0975-6906.2017.00005.0

In silico characterization of QTL regions associated with drought tolerance traits in rice (*Oryza sativa* L.)

Sweta Sinha[1]*, A. S. Kotasthane and S. B. Verulkar

Department of Plant Molecular Biology and Biotechnology, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Krishak Nagar, Raipur 492 001, Chhattisgarh

(Received: June 2016; Revised: December 2016; Accepted: December 2016)

Abstract

An attempt was made to carry out *in silico* characterization of 14 putative QTL regions identified on chromosomes 1, 2, 4, 5, 6 and 8 associated with drought tolerance. A total of 354 BAC/PAC clones were detected in the underlying regions of putative QTLs mapped. These BAC/PAC clone nucleotide sequences were used for mining SSRs and primer designing. The tri-nucleotide motif was the most abundant class which was followed by tetra, di, penta and hexa nucleotide repeats. A total of 1849 class I SSR markers were designed from QTL regions. The underlying regions of putative QTL displayed different functional categories of drought tolerance genes, transcription factors (45.65 %), protein kinases (33.94 %), stress proteins (10.11 %), phytohormones (9.75 %) and aquaporins (0.72 %). The massively parallel signature sequence (MPSS) analysis of candidate genes showed higher expression for transcription factor having WRKY and zinc finger super family (LOC_Os02g08440) in stem, immature panicle, developing seeds, mature leaves, mature roots, merismatic tissues, cold, drought and salt stressed young roots transcript tissue library. The functional validation identified putative genes based on MPSS signatures will facilitate to unravel the complexity of drought tolerance at molecular level and further will accelerate the genetic improvement of drought tolerance in rice.

Key words: Candidate gene, drought, *In silico*, MPSS, QTL, rice

Introduction

Rice is known as the grain of life as it feeds more than three and half billion people worldwide. It also serves as a model monocot plant for genetic and genomic study. It is cultivated in 154 million hectares globally in different agro-ecological conditions. Rice is susceptible to drought stress due to its shallow root distribution and limited capacity to absorb water from deep soil (Kondo et al. 2003). Drought acts as a serious limiting factor in agriculture production by preventing a crop from reaching the genetically determined theoretical maximum yield. Traits contributing to enhance tolerance to drought in rice, primary traits *viz*., root traits (root thickness, root dry weight per tiller, maximum rooting depth, root volume, root/shoot ratio and root/shoot weight ratio) are the key because they are associated with acquisition of moisture, nutrients and transmitting drought perception signals (Gowda et al. 2011; Rogers and Benfey 2015). However, the genetic improvement of rice for drought tolerance has been a major challenge to plant breeders due to the complexity of the inheritance of drought tolerance, G x E interaction and difficulties faced in the screening of plant materials for drought tolerance (Khush 2001). These complexities have led to limited success in developing drought tolerant plants or improving crop yields. In this context, molecular marker technology gained prominance in improving the efficiency of conventional plant breeding for drought tolerance during last two decades through QTL mapping (Nguyen et al. 1997). Many QTLs have been reported in rice for traits that are putatively associated with performance under drought. Divyapriya et al. (2015) evaluated backcross improved lines in white Ponni background carrying QTLs from Apo displayed a good degree of drought tolerance. However, it is still unclear about genes underlying the QTLs for drought tolerance and therefore their characterization is important to understand the quantitative variation and genes

*Corresponding author's e-mail: bablysweta@gmail.com
[1]Present address: Department of Molecular Biology and Genetic Engineering, Bihar Agricultural University, Sabour, Bhagalpur, Bihar
Published by the Indian Society of Genetics & Plant Breeding, F2, First Floor, NASC Complex, PB#11312, IARI, New Delhi 110 012
Online management by indianjournals.com; http://epubs.icar.org.in/journal/index.php/IJGPB

underlying these regions for strategizing rice breeding programs. Hazen et al. (2005) and Degenkolbe et al. (2009) identified candidate genes for QTL related to drought resistance in rice by differential expression patterns. QTL mapping eventually provide as entry point to utilize massive amounts of biological data for *in silico* analysis. Now with the availability of *in silico* expression profiling tools include microarray, serial analysis of gene expression (SAGE) and massively parallel signature sequencing (MPSS) for transcriptome analysis will provide the opportunity for digital expression profiling of the putative candidates genes (Brenner et al. 2000; Mochida and Shinozaki, 2010) and which can be then taken for further functional validation. The use of MPSS for the identification of gene expression has been reported by several workers (Nakano et al. 2006; Banerjee et al. 2010; Dubey and Chandel 2010; Channamallikarjuna et al. 2010; Chandel et al. 2011; Sonah et al. 2016). Therefore, taking the advantage of publicly available genomic data for rice an attempt was made to *in silico* characterize the QTL region associated with drought tolerance traits to provide the basis for the development of candidate drought tolerance gene markers for fine mapping and effective use in marker-assisted breeding.

Materials and methods

Target QTL region for in silico analysis

Fourteen putative QTLs identified by Sinha et al. (2015) for drought tolerance traits in a F_7 RIL population derived from the parent Danteshwari and Dagad Deshi were selected for the *in silico* study. The identified putative QTLs was 2 for each root volume (*qrv4-1*, *qrv6-1*), root dry weight (*qrdw4-1*, *qrdw8-1*), shoot length (*qshl1-1*, *qshl1-2*), shoot fresh weight (*qsfw4-1*, *qsfw5-2*) and total plant length (*qtpl1-1*, *qtpl4-1*), 1 for root/shoot fresh weight ratio (*qrsfw2-1*) and 3 for root/shoot dry weight ratio (*qrsdw4-1*, qrsdw8-*1, qrsdw8-2*). The physical position of each flanking QTL interval marker was obtained by using BLAST search in RGP web site. The entire BAC/PACs clone underlying QTL interval and their nucleotide sequences were downloaded from TIGR Pesudomolecule Build 5 in FASTA formatted text files (http://rice. Plantbiology .msu. edu/annotation_pseudo_current.shtml).

SSRs mining and primer designing using the BAC/PACs clone underlying QTL region

The nucleotide sequences of BAC/PACs clone obtained were used for mining of SSR motifs using a program Batchprimer3 (http://probes.pw.usda.gov/cgi-bin/batchprimer3/). The SSR motifs identified from the BAC/PACs were analyzed for occurrence of di-, tri-, tetra-, penta- and hexa- motifs. The SSR motifs were classified into different repeat categories and repeat motifs AT/TA, AC/CA, TG/GT, AG/GA, GC/CG and TC/CT were group in the same class. Based on the length of the SSR motifs, they were classified into class I ($\geq$20 nucleotides) and class II (12-19 nucleotides) (Temnykh et al. 2001). The tri-nucleotide motifs were categorized on the basis of amino acid coded by them.

Primers were designed using BatchPrimer3 software. Primer designing was done by putting the individual BAC/PACs sequences in FASTA format with desired specifications for GC content (50-60%), annealing temperature (50-60^{o}C), primer length (18-20bp) and length of amplified fragments (100-300bp) with other parameters as default setting. Primer pairs generated were also classified into class I and class II.

Mining of putative genes within putative QTL regions

The genes on sequences of BAC/PACs encompassing putative QTL intervals were downloaded from TIGR websites. Based on publically available information the genes on sequences encompassing QTL intervals were categorized for drought tolerance (http://rice. plantbiology.msu.edu).

In silico expression analysis of putative drought tolerance genes underlying the putative QTL regions

Massive Parallel Signature Sequence (MPSS) tags corresponding to drought related genes in rice were identified. MPSS is a sequence based technology for transcriptional profiling. It uses a unique method to quantify gene expression levels by generating millions of short sequence tags per library by sequencing 16-20 bp from the 3' side of cDNA using a micro bead array. The abundance of each signature represents and measures the gene expression levels in the sample tissue (Brenner et al. 2000). MPSS provides quantitative description of gene expression. The MPSS tags sequences are derived from specific tissue expression library and abundance of these tags in a library is depicted in terms of its TPM values. Therefore TPM value of co-localized MPSS tags in corresponding expression library will provide an indirect estimate of the levels of expression of that putative gene. The rice MPSS database made up of

comprehensive set of libraries and these MPSS libraries derived from diverse tissues root, leaves, stem, panicle, germinating seeds and abiotically stressed (cold, drought and salt) tissues have been generated for japonica cultivar Nipponbare grown under different conditions as light and dark, different developmental stages and several biological replicates (http:/mpss.udel.edu/rice). TIGR locus identifier for each gene was used as query to obtain all annotated or non-annotated MPSS tags. The output of each search was recorded and analyzed to obtain a meaningful result. The search resulted in 17 nt long MPSS tags, tag sequence, chromosome coordinate position, tissue library information and transcript abundance values such as TPM (transcripts per million) value, normalized abundance in different steps and p value displayed in table format. The transcript number under 'Norm Abund' category was considered mainly to draw a conclusion about abundance of transcript and henceforth level of expression of gene in the particular tissue type. Nakano et al. (2006) in a study of MPSS tag based characterization of expression pattern in *Arabidopsis* have reported that TPM<5 corresponds to normalized housekeeping genes and TPM<15 indicated very weak expression (Meyers et al. 2004). Hence only those tissue were considered in this study that showed TPM>15.

Results and discussion

In silico analysis of the target QTL region

The fourteen putative QTLs detected on chromosomes 1, 2, 4, 5, 6 and 8, underlying this region 354 BAC/PAC clones were worked out (Table 1). The obtained nucleotide sequences of BAC/PAC clones were further used for *in silico* analysis.

SSRs mining using BAC/PAC clone sequences present within putative QTL region

The obtained nucleotide sequences of BAC/PAC clones were used for mining of SSRs. Of the total 23431 putative SSRs were found in BAC/PAC sequences, 2604 belongs to class I and 20827 belongs to class II SSRs (Table 2). The tri-nucleotide motif (9808) were the most abundant class which was followed by tetra- (5635), di- (4726) nucleotides repeats, penta- (1932) and hexa-(1326) nucleotides repeats.

The AT/TA was most common di-nucleotide motif 42.30% followed by AG/GA (28.85 %); TC/CT (13.64

Table 2. Analysis of BAC/PACs sequences present within the putative QTL regions for the presence of simple sequence repeat motifs

Chr.	SSR class		Nucleotide motif				
	I	II	Di-	Tri-	Tetra-	Penta-	Hexa-
1	337	2765	588	1300	748	276	190
2	311	2304	425	1133	707	213	137
4	895	7519	1679	3550	2051	663	471
5	46	231	46	126	50	28	23
6	612	4625	1198	2077	1202	481	279
8	403	3383	790	1622	877	271	226
Total	2604	20827	4726	9808	5635	1932	1326

Chr. = Chromosome

Table 1. BAC/PACs encompassing putative QTLs region

QTLs detected	Chromosome	Common or overlapping marker interval used for *in silico* analysis	Size of QTL spanning genomic region (Mb)	No. of BAC/ PAC present within the putative QTL region
QTL for Shoot length (*qshl1-1*, *qshl1-2*), Total plant length (*qtpl1-1*)	1	HvSSR01-80-HvSSR01-87	4.1	37
QTL for root/shoot fresh weight ratio (*qrsfw2-1*)	2	HvSSR2-12-HvSSR2-23	3.1	30
QTL for root volume (*qrv4-1*), shoot fresh weight (*qsfw4-1*), total plant length (*qtpl4-1*), root dry weight (*qrdw4-1*), root/shoot dry weight (*qrsdw4-1*)	4	RM3471-HvSSR04-32	16.4	136
QTL for shoot fresh weight (*qsfw5-2*)	5	HvSSR05-12-HvSSR05-13	0.3	3
QTL for root volume (*qrv6-1*)	6	HvSSR06-30-HvSSR06-44	9.3	83
QTL for root/shoot dry weight (*qrsdw8-1, qrsdw8-2*), root dry weight (*qrdw8-1*)	8	HvSSR08-24-HvSSR08-29	6.6	65

%), GC/CG (10.13 %), TG/GT (2.74 %) of the total number of di-nucleotide identified in putative QTL region of BAC/PAC sequences. The least abundant di-nucleotide motif was AC/CA and was only 2.34 % of the total number of di-nucleotide identified in putative QTL region of BAC/PAC sequences (Table 3). The (AT)n motif is the most abundant di-nucleotide motifs in rice as well as in other plant species was also described by Cardle et al. (2000) and Temnykh et al. (2001).

The tri-nucleotide repeat motifs were the most abundant class of microsatellites in eukaryotes especially cereal species as well as in Arabidopsis as described by Parida et al. (2006). They reported that the proportion of trinucleotide containing unigenes ranged from 60% in maize to 80% in rice. The higher frequency of the tri-nucleotide repeat motifs than other classes due to the selection against frameshift mutations that limits expansion of nontriplet microsatellites as reported by Metzgar et al. (2000).

Table 3. Di-nucleotide repeat motifs identified in BAC/PAC sequences present within the putative QTL regions

Motif	Chromosome						Total
	1	2	4	5	6	8	
AT/TA	183	181	685	21	620	295	1985
AC/CA	28	34	70	3	40	33	208
AG/GA	136	74	375	12	186	154	937
TG/GT	29	30	77	2	54	33	225
GC/CG	64	21	106	1	67	70	329
TC/CT	148	85	366	7	231	205	1042

In the present investigation observation among the tri-nucleotide repeats, the motifs AGA/CGA/CGC/CGG/CGT/AGG coding for amino acid arginine were most abundant within the putative QTL region followed by GCA/GCC/GCG/GCT coding for alanine; GGA/GCG/GGT coding for glycine; CCA/CCG/CCT coding for proline (Table 4). The proportion of proline (freq.

Table 4. Tri-nucleotide repeat motifs identified in BAC/PACs sequences coding for different amino acids

Tri-nucleotide motif identified in BAC clone sequences	Amino acid	Chromosome						Total
		1	2	4	5	6	8	
GCA, GCC, GCG, GCT	Alanine	231	177	661	19	346	276	1710
AGA, CGA, CGC, CGG, CGT, AGG	Arginine	255	271	802	32	471	331	2162
AAC	Asparagine	3	16	12	-	11	15	57
GAC, GAT	Aspartic acid	28	20	81	1	49	36	215
TGC, TGT	Cysteine	40	16	61	1	23	15	156
GAA, GAG	Glutamic acid	42	32	112	6	84	56	332
CAA, CAG	Glutamine	18	17	78	1	30	42	186
GGA, GGC, GGT	Glycine	172	129	372	15	245	188	1121
CAT, CAC	Histidine	20	15	41	2	27	17	122
ATA, ATC, ATT	Isoleucine	16	25	64	2	31	27	165
TTA, TTG, CTA, CTC, CTG, CTT	Leucine	82	86	229	4	120	111	632
AAG, AAT	Lysine	26	15	57	-	37	36	171
ATG	Methionine	3	1	15	-	9	8	36
TTC	Phenylalanine	4	12	26	1	22	21	86
CCA, CCG, CCT	Proline	155	136	387	30	223	183	1114
TAA, TAG, TGA	Stop	13	16	60	1	37	27	154
AGC, AGT, TCA, TCC, TCG, TCT	Serine	90	64	212	8	148	99	621
ACA, ACC, ACG, ACT	Threonine	28	19	91	-	45	34	217
TGG	Tryptophan	20	14	44	1	42	19	140
TAC, TAT	Tyrosine	22	21	38	-	29	27	137
GTA, GTC, GTG, GTT	Valine	32	31	107	2	48	54	274

1114), glycine (freq. 1121), alanine (freq. 1710) and arginine (freq. 2162) were represented as 11.36 %, 11.43 %, 17.43 %, 22.04% respectively and were represented in all six chromosomes (BAC/PACs) of putative QTL interval region. The abundance of small/ hydrophilic amino acid repeat motifs like alanine is present in the unigenes of cereals and Arabidopsis due to strong selection pressure it replaces the codon repeats encoding hydrophobic/other amino acids (Katti et al. 2001).

Primer designing for the development of novel SSR markers

Identified SSRs within the putative QTL region are used for primer designing to develop SSR markers. Of the total 16755 primer pair's designed, 1849 were defined as Class I and 14906 as Class II (Table 5).

Table 5. Frequency distribution of novel SSR markers identified in BAC/PAC sequences

Chr.	SSR class		Primer pairs designed	
	Class I	Class II	Class I	Class II
1	337	2765	255	1977
2	311	2304	232	1725
4	895	7519	643	5518
5	46	231	32	176
6	612	4625	418	3130
8	403	3383	269	2380
Total	2604	20827	1849	14906

Chr. = Chromosome

Cho et al. (2000) and Temnykh et al. (2001) found that class I SSR markers were more polymorphic than the class II SSR and there was a positive correlation between SSR length and polymorphism in rice. The novel SSR markers developed in the study are in direct relevance to drought tolerance traits. After validation under wet lab and analysis for association with drought related traits, the novel class I SSRs identified within QTL regions will allow faster genotyping with mapping populations for drought related trait as well as cross transferability across related cereal genomes.

Putative genes identified within the QTL regions

Despite many decades of research, drought tolerance continues to a major challenge because of the complexity of the traits. Accumulating evidence suggests that plant response to drought tolerance is controlled by more than one gene and is highly influenced by environmental variation. In order to narrow down the target QTL regions, putative candidate genes were mined for drought tolerance and genes were selected based on Hadiarto and Tran (2011) publically available information. In the present study 277 (4.44 %) putative candidate genes related to drought tolerance traits were identified out of 6233 genes downloaded for 354 BAC/PAC clones distributed across the detected putative QTL regions (Table 6).

Table 6. Frequency distribution of putative genes identified in BAC/PAC clones underlying the putative QTL regions

Chromo-some	Gene displayed by BAC/PAC	Putative drought tolerance gene	% gene
1	654	65	9.94
2	482	56	6.22
4	2556	72	5.32
5	54	5	9.26
6	1440	51	3.54
8	1047	28	2.67

The putative QTLs regions are found to harbor key drought tolerance related genes namely, transcription factors, protein kinases, stress proteins, phytohormones and aquaporins (Figs. 1 and 2). The present study showed the occurrence of key signaling genes namely, serine/threonine protein kinase and calcium dependent protein kinases in the QTL region. Reddy et al. (2002) reported these protein kinases are either stress inducible or up regulated by dehydration. The putative QTL region was found to contain key transcription factors namely zinc finger, WRKY, MYB (myeloblastosis), AP2 (Apetala 2), helix loop helix, bZIP (basic-domain leucine zipper), DREB (drought-responsive element binding factor) and NAC domain (NAM-no apical meristem). Many workers also revealed that genes encoding transcription factors occupy a considerable fraction of the genomes of rice and other higher plants (Riechmann et al. 2000; Gao et al. 2006; Ramya et al. 2010). Maurel et al. (2002) reported for plant growth and development, osmoregulation and physiological processes like water uptake and flow across cellular membranes, aquaporins are essential proteins that involved in water transport. Rice cytochrome P450 genes and Heat shock proteins were also identified in the putative QTL regions but have not been fully characterized in rice (Wang et al. 2014). Earlier studies on expression analysis of rice AP2

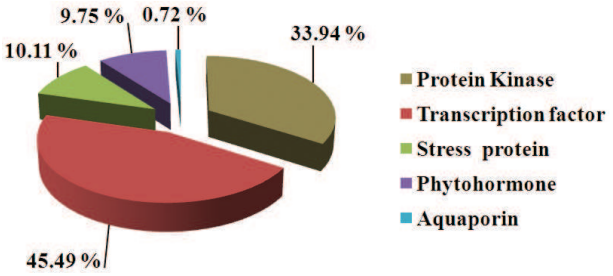

Fig. 1. Functional categorization of putative drought tolerance genes detected within the putative QTLs region associated with drought tolerance traits

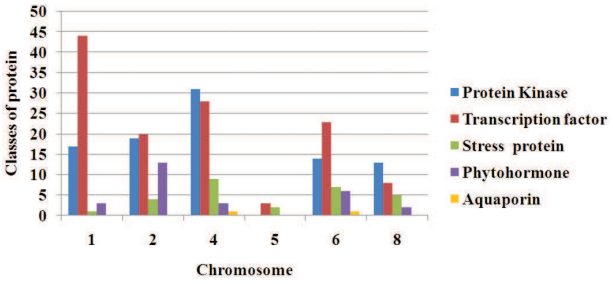

Fig. 2. *In silico* functional categorization of putative drought tolerance genes detected within the putative QTLs region associated with drought tolerance traits on six chromosomes

domain (Matsukura et al. 2010), WRKY (Ramamoorthy et al. 2008), NAC (Nuruzzaman et al. 2010), zinc finger (Huang et al. 2009b), MYB family (Yang et al. 2012), basic helix-loop-helix (Kiribuchi et al. 2005), cytochrome P450 (Gorantla et al. 2007; Degenkolbe et al. 2009; Tamiru et al. 2015), protein kinases (Xiong and Yang, 2003) and aquaporins (Grondin et al. 2016) also indicated the involvement of these genes in drought conditions. Further characterization of these genes may provide a basis for map based cloning of the target region for drought tolerance rice breeding.

In silico expression analysis of putative drought tolerance genes underlying the putative QTL regions

To unravel in depth into the expression analysis of drought tolerance genes underlying the putative QTL region, in silico based MPSS analysis was done. The analysis was done by putting locus identifiers of individual genes as query and search resulted in 17 bp long MPSS tags, tag sequence, chromosome coordinate position, tissue library information and transcript abundance values such as TPM value, which indicates expression level of corresponding gene quantitatively. Here, TPM values of MPSS tags varied from 1 to 145163. A total of 481MPSS tags (17 bp) were found corresponding to putative QTL region.

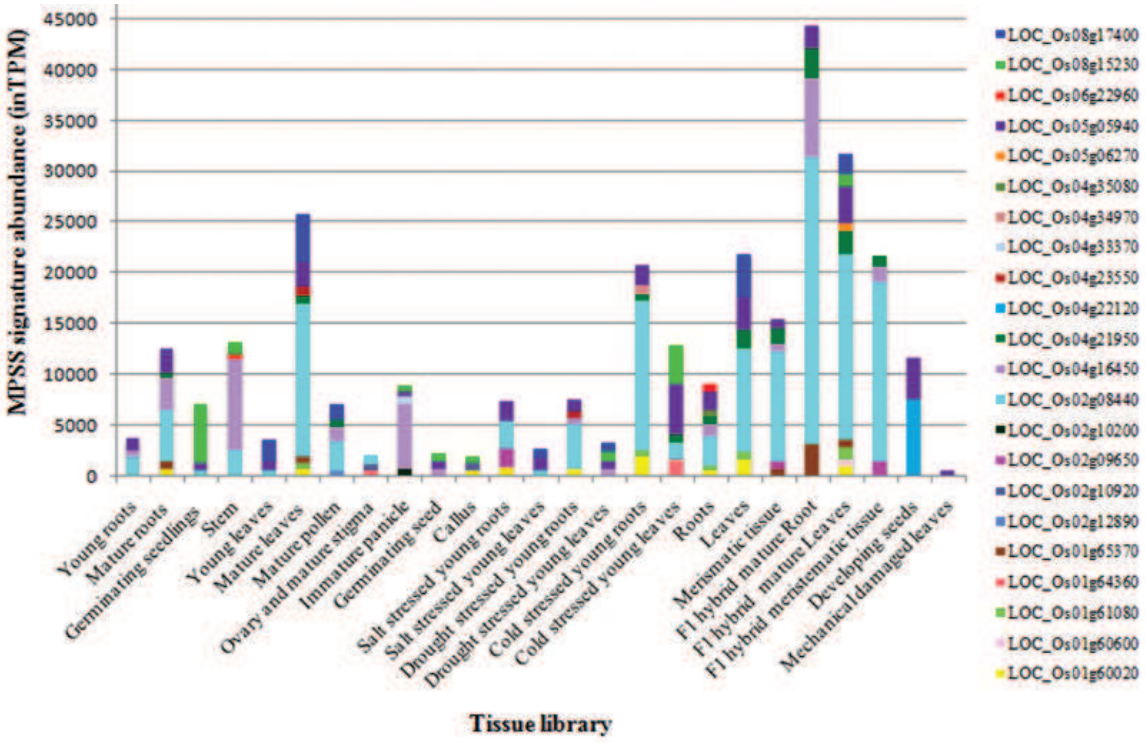

Fig. 3. *In silico* expression pattern in different tissue libraries for 22 putative drought tolerance genes underlying the QTL regions based on MPSS signature abundance

Maximum number of signature tags belonged to class I (within exon, same strand) and rest belonged to III, II, V and VI class of MPSS signature tags (Table 7). Dubey and Chandel (2010) also reported maximum number of signature of Class I. Class I MPSS signature tags have been used for the characterization of several developmental libraries and stress transcriptomes by many workers (Goring et al. 2004).

Table 7. Number of 17 bp MPSS signatures of different classes corresponding to putative drought related gene present in the QTL regions

Class of MPSS tags	Position of MPSS tags	No. of signature tags on different chromosomes						Total
		1	2	4	5	6	8	
1	Within exon, same strand	52	37	48	6	43	18	204
2	Within 500 bp potential 3_UTR	27	24	22	2	28	10	113
3	Antisense to exon	24	28	37	1	27	5	122
4	Unannotated	-	-	-	-	-	-	-
5	Within intron, sense strand	6	3	7	-	9	2	27
6	Within intron, antisense strand	4	1	5	1	2	2	15
7	Spans an exon/intron splice site	-	-	-	-	-	-	-

Based on the frequency of TPM, out of 277 locus, 80 locus were containing MPSS tag of >500 TPM value. The loci LOC_Os01g60020, LOC_Os02g08440, LOC_Os04g16450, LOC_Os05g05940, LOC_Os06g19680 and LOC_Os08g15230 on chromosome 1, 2, 4, 5, 6 and 8 have higher MPSS signature abundance respectively. The tissue library wise information showed higher expression in stem, mature leaves, mature roots, immature panicle, developing seeds, meristematic tissues, cold, drought and salt stressed young roots (Fig. 3). MPSS analysis also indicated 22 putative genes were identified as highly expressed drought tolerance loci having TPM> 500 in most of the tissue library (Meyers et al. 2004) (Table 8). The locus identifier LOC_Os02g08440 (MPSS tag value 1,45,163.00) identified in QTL region contributing toward the root/shoot fresh weight ratio was functional gene in general which code for transcription factor (WRKY super family, OsWRKY71) is more promising candidate responsible for drought tolerance in rice.

The loci showed comparatively higher expression based on MPSS signature abundance in the library comprising of drought stress. These MPSS tag may serve as a source for identification, analysis and functional characterization of genes controlling drought stress. These putative drought tolerant loci can also be used in the development of functional markers for drought tolerance.

The genomics based information along with advanced computational tools and databases has greatly influenced and redefined crop breeding practices as genomics assisted breeding. In this study, a total of 354 BAC/PAC clones, 2604 class I SSR motifs, 1849 class I SSR markers *in silico* designed, 277 drought responsive putative genes and 481 MPSS tags (17 bp) were detected in the underlying regions of putative QTLs mapped on chromosomes 1, 2, 4, 5, 6 and 8 associated with drought tolerance traits. The developed novel class I SSR markers within QTL regions with repeat lengths of 20 bp or higher is expected to be more polymorphic due to high length dependent replication slippage and will help in saturation mapping of QTL regions for better resolution of the rice genetic maps. The putative QTL region contains genes associated with drought namely transcription factors (WRKY, MYB, AP2, helix loop helix, bZIP and NAC domain), protein kinases, stress proteins, heat shock proteins, phytohormones and aquaporins. Furthermore, *in silico* expression analysis of these genes revealed that WRKY and zinc finger super family (LOC_Os02g08440) have higher expression. The identified putative candidate genes to QTL regions with mapped phenotypes provide a basis for map based cloning of this region and the MPSS expression information gave putative expression pattern which will help in selection of genes for the improvement of drought tolerance. Thus, further validation of these genes that are identified in this study can lead to unravel the molecular mechanism

Table 8. Expression pattern of 22 putative drought tolerance genes underlying the putative QTL regions based on MPSS signature abundance (TPM> 500)

Gene name	Locus identifier	Signature tag	MPSS signatures abundance (TPM)
Chromosome 1			
NAC domain transcription factor, putative, expressed	LOC_Os01g60020	GATCATGCACGAGTACC	8429
WRKY DNA-binding domain containing protein, expressed	LOC_Os01g60600	GATCCTACATGTGGTTC	646
OsWRKY24 - Superfamily of TFs having WRKY and zinc finger domains, expressed	LOC_Os01g61080	GATCACCACCTACGAGG	3835
MYB family transcription factor, putative, expressed	LOC_Os01g64360	GATCCATAACCGTGGAC	2145
MYB family transcription factor, putative	LOC_Os01g65370	GATCAAGAACTACTGGA	5928
Chromosome 2			
Cytochrome P450, putative, expressed	LOC_Os02g12890	GATCGGGCCACGGGCGT	535
Zinc finger family protein, putative, expressed	LOC_Os02g10920	GATCGAGTCGTGTAACC	1324
AP2 domain containing protein, expressed	LOC_Os02g09650	GATCCGCGACCCGGCGA	3899
Zinc finger A20 and AN1 domain-containing stress-associated protein, putative, expressed	LOC_Os02g10200	GATCTAAGGCGGCTAGG	636
OsWRKY71- Superfamily of TFs having WRKY and zinc finger domains, expressed	LOC_Os02g08440	GATCATCGGCGGCGGCC	143586
Chromosome 4			
Aquaporin protein, putative, expressed	LOC_Os04g16450	GATCTTCTGGGCGGGGC	33935
OsWRKY51-Superfamily of TFs having WRKY and zinc finger domains, expressed	LOC_Os04g21950	GATCCGCGTGCCGGCGA	13863
Protein kinase, putative, expressed	LOC_Os04g22120	GATCGCTATCTTTAGTC	7570
Basic helix-loop-helix family protein, putative, expressed	LOC_Os04g23550	GATCACATGGATAGTGT	1435
Cytochrome P450, putative, expressed	LOC_Os04g33370	GATCTGCCCCGGGGTGG	721
AP2 domain containing protein, expressed	LOC_Os04g34970	GATCCTGGCGGGGGCGC	885
Protein kinase domain containing protein, expressed	LOC_Os04g35080	GATCACCACCACCACCA	682
Chromosome 5			
Stress-related protein, putative, expressed	LOC_Os05g05940	GATCGCCTTGTTGGGTC	37307
Zinc finger, C3HC4 type domain containing protein, expressed	LOC_Os05g06270	GATCACCCGAGGGCTCA	762
Chromosome 6			
Aquaporin protein, putative, expressed	LOC_Os06g22960	GATCGAACACTTCTGTT	1246
Chromosome 8			
Heat shock protein-related, putative, expressed	LOC_Os08g15230	GATCGCCGCCGCCGCCG	14949
WRKY DNA-binding domain containing protein, expressed	LOC_Os08g17400	GATCTGCACGGTTACAG	17658

of drought tolerance and development of functional markers to accelerate breeding programs in rice.

Authors' contribution

Conceptualization of research (SS, SBV); Designing of the experiments (SS); Contribution of experimental materials (SBV); Execution of field/lab experiments and data collection (SS); Analysis of data and interpretation (ASK, SBV); Preparation of manuscript (SS).

Declaration

The authors declare no conflict of interest.

Acknowledgements

The authors are highly thankful to Indira Gandhi Krishi Vishwavidyalaya and Department of Plant Molecular Biology and Biotechnology for providing computational laboratory facilities to carry out this work. The financially support by the Department of Information Technology (DIT), Government of India, New Delhi is gratefully acknowledged.

References

Banerjee S., Sharma D. J., Verulkar, S. B. and Chandel, G. 2010. Use of In-silico and semiquantitative RT-PCR approaches to develop nutrient rich rice (*Oryza sativa* L.). Indian J. Biotechnol., **9:** 203-212.

Brenner S., Johnson M., Bridgham J., Golda G., Lloyd D. H., Johnson D., Luo S., McCurdy S., Foy M. and Ewan M. 2000. Gene expression analysis of massively parallel signature sequencing (MPPS) on microbead arrays. Nat. Biotechnol., **18:** 630-634.

Cardle L., Ramsay L., Milbourne D., Macaulay M., Marshall D. and Waugh R. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics, **156:** 847-854.

Chandel G., Dubey M., Samual P. and Meena R. 2011. *In silico* mining and characterization of novel SSRs and candidate genes within QTLs controlling grain protein contents using MPSS signatures and micro array analysis in rice (*Oryza sativa* L.). POJ., **4**(3): 126-135.

Channamallikarjuna V., Sonah H., Prasad M., Rao G. J. N., Chand S., Upreti H. C., Singh N. K. and Sharma T. R. 2010. Identification of major quantitative trait loci qSBR11-1 for sheath blight resistance in rice. Mol. Breed., **25**(1): 155-166.

Cho Y. G., Ishii T., Temnykh S., Chen X., Lipovich L., Park W. D., Ayres N., Cartinhour S. and McCouch S. R. 2000. Diversity of microsatellites derived from genomic libraries and GeneBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet., **100:** 713-722.

Degenkolbe T., Do P. T., Zuther E., Repsilber D., Walther D., Hincha D. K. and Kohl K. I. 2009. Expression profiling of rice cultivars differing in their tolerance to long-term drought stress. Plant Mol. Biol., **69:** 133-153.

Divyapriya D., Kalamani A., Raveendran M. and Robin S. 2016. Evaluation of drought tolerance in backcross inbred lines of rice genotypes based on selection indices. Indian J. Genet., **76**(3)**:** 290-298. DOI: 10.5958/0975-6906.2016.00044.4

Dubey M. and Chandel G. 2010. *In silico* survey and characterization of Resistance Gene Analogues (RGAs) in the genomic regions encompassing gall midge resistance genes *Gm4* and *Gm5* in rice (*Oryza sativa* L.). POJ., **3**(5): 140-148.

Gao G., Zhong Y., Guo A., Zhu Q., Tang W., Zheng W., Gu X., Wei L. and Luo J. 2006. DRTF: a database of rice transcription factors. Bioinformatics, **22:** 1286-1287.

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.

Goring D. R., Salt J. N., Stone S. L., Shiu S. H. and Mudgil Y. 2004. A large complement of the predicted Arabidopsis ARM repeat proteins are members of the U-Box E3 ubiquitin ligase family1**.** Plant Physiol., **134:** 59-66.

Gowda V. R. P., Henry A., Yamauchi A., Shashidhar H. E. and Serraj R. 2011. Root biology and genetic improvement for drought avoidance in rice. Field Crops Res., **122:** 1-13.

Grondin A., Mauleon R., Vadez V. and Henry A. 2016. Root aquaporins contribute to whole plant water fluxes under drought stress in rice (*Oryza sativa* L.). Plant, Cell and Environment, **39:** 347-365.

Hadiarto T. and Tran L. P. 2011. Progress studies of drought-responsive genes in rice. Plant Cell Rep., **30**(3): 297-310.

Hazen S. P., Pathan M. S., Sanchez A., Baxter I., Dunn M., Estes B., Chang H. S., Zhu T., Kreps J. A. and Nguyen H. T. 2005**.** Expression profiling of rice segregating for drought tolerance QTLs using a rice genome array. Funct. Integr. Gen., ***5*:** 104-116.

Huang X. Y., Chao D. Y., Gao J. P., Zhu M. Z., Shi M. and Lin H. X. 2009b. A previously unknown zinc finger protein, DST, regulates drought and salt tolerance in rice via stomatal aperture control. Genes Dev., **23:** 1805-1817.

Katti M. V., Gupta P. K. and Ranjekar V. S. 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Biol. Evol., **18:** 1161-1167.

Khush G. S. 2001. Green revolutions: the way forward. Nature Rev. Genet., **2:** 815-822.

Kiribuchi K., Jikumaru Y., Kaku H., Minami E., Hasegawa M., Kodama O., Seto H., Okada K., Nojiri H. and Yamane H. 2005. Involvement of the basic helix-loop-helix transcription factor RERJ1 in wounding and drought stress responses in rice plants. Biosci., Biotechnol. Biochem., **69**(5): 1042-1044.

Kondo M., Pablico P. P., Aragones D. V., Agbisit R., Abe J., Morita S. and Courtois B. 2003. Genotypic and environmental variations in root morphology in rice genotypes under upland field conditions. Plant Soil, **255:** 189-200.

Matsukura S., Mizoi J. and Yoshida T. 2010. Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. Mol. Genet. Genomics, **283:** 185-196.

Maurel C., Javot H., Lauvergeat V., Gerbeau P., Tournaire C., Santoni V. and Heyes J. 2002. Molecular physiology of aquaporins in plants. Int. Rev. Cytol., **215:** 105-148.

Meyers B. C., Vu T. H., Tej S. S., Ghazal H., Matvienko M. and Agrawal V. 2004. Analysis of the transcriptional complexity of *Arabidopsis thaliana* by massively parallel signature sequencing. Nat. Biotechnol., **22:** 1006-1011.

Metzgar D., Bytof J. and Wills C. 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res., **10:** 72-80.

Mochida K. and Shinozaki K. 2010. Genomics and Bioinformatics Resources for Crop Improvement. Plant Cell Physiol., **51**(4): 497-523.

Nakano M., Nobuta K., Vemaraju K., Tej S. S., Jeremy W. S. and Meyers B. C. 2006. Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. Nucleic Acids Res., **34:** 731-735.

Nuruzzaman M., Manimekalai R., Sharoni A. M., Satoh K., Kondoh H., Ooka H. and Kikuchi S. 2010. Genome-wide analysis of NAC transcription factor family in rice. Gene, **465:** 30-44.

Nguyen H. T., Babu R. C. and Blum A. 1997. Breeding for drought tolerance in rice: physiology and molecular genetics considerations. Crop Sci., **37:** 1426-1434.

Parida S. K., Kumar K., Vivek D., Singh N. K. and Mohapatra T. 2006. Unigene derived microsatellite markers for the cereal genomes. Theor. Appl. Genet., **112:** 808-817.

Ramamoorthy R., Jiang S. Y., Kumar N. and Venkatesh P. N. 2008. A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments. Plant Cell Physiol., **49:** 865-879.

Ramya M., Raveendran M., Sathiyamoorthy S. and Ramalingam J. 2010. *In silico* analysis of drought tolerant genes in rice. Int. J. Biol. Med. Res., **1**(3): 36-40.

Reddy A. R., Ramakrishna W., Chandra 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.

Riechmann J. L., Heard J., Martin G., Reuber L., Jiang C., Keddie J., Adam L., Pineda O., Ratcliffe O. J., Samaha R. R., Creelman R., Pilgrim M., Broun P., Zhang J. Z., Ghandehari D., Sherman B. K. and Yu G. 2000. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science, **290:** 2105-2110.

Rogers E. D. and Benfey P. N. 2015. Regulation of plant root system architecture: implications for crop advancement. Curr. Opinion Biotechnol., **32:** 93-98.

Sinha S., Verulkar S. B. and Kotasthane A. S. 2015. Identification of QTLs for root and shoot traits in RIL population of rice (*Oryza sativa* L.). Indian J. Genet., **75**(1): 49-56.

Sonah H., Chavan S., Katara J., Chaudhary J., Kadam S., Patil G. and Deshmukh R. K. 2016. Genome-wide identification and characterization of Xylanase Inhibitor Protein (XIP) genes in cereals. Indian J. Genet., **76**(2): 159-166.

Tamiru M., Undan J. R., Takagi H., Abe A., Yoshida K., Undan J. Q., Natsume S., Uemura A., Saitoh H., Matsumura H., Urasaki N., Yokota T. and Terauchi R. 2015. A cytochrome P450, OsDSS1, is involved in growth and drought stress responses in rice (*Oryza sativa* L.). Plant Mol. Biol., **88**(1): 85-99.

Temnykh S., Declerk G., Lukashover A., Lipovich L., Cartinhour S. and McCouch S. R. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations and genetic marker potential. Genome Res., **11:** 1441-1452.

Wang H., Lin S., Song Q., Li K., Tao H., HuangJ., Chen X., Que S. and He H. 2014. Genome-wide identification of heat shock proteins (Hsps) and Hsp interactors in rice: Hsp70s as a case study. BMC Genomics, **15:** 344.

Xiong L. and Yang Y. 2003. Disease resistance and abiotic stress tolerance in rice are inversely modulated by an abscisic acid inducible mitogen-activated protein kinase. Plant Cell, **15:** 745-759.

Yang A., Dai X. and Zhang W. 2012. A R2R3-type MYB gene, OsMYB2, is involved in salt, cold, and dehydration tolerance in rice. J. Exp. Bot., **63:** 2541-2556.