Asian Journal of Plant Sciences

ISSN 1682-3974

science alert

ANSInet
an open access publisher
http://ansinet.com

Asian Journal of Plant Sciences 7 (3): 268-275, 2008
ISSN 1682-3974

Data Mining of *Sub*QTL Region on Chromosome 9: Dissecting Gene Structure and Protein Function

[1]V. Ruanjaichon, [2]S. Tragoonrung and [1,3]A. Vanavichit
[1]Rice Gene Discovery Unit, National Center for Genetic Engineering and Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Nakorn Pathom 73140, Thailand
[2]DNA Technology Laboratory, National Center for Genetic Engineering and Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Nakorn Pathom 73140, Thailand
[3]Department of Agronomy, Kasetsart University, Kamphaeng Saen Campus, Nakorn Pathom 73140, Thailand

Abstract: The complete set of genes and their genomic locations could be identified via genomic sequencing with an acceptable investment. A 1.35 Mb contiguous region consisting of 14 BAC/ PAC clones were anchored on the long arm of chromosome 9 corresponding to about 15.5 cM of the genetic map using sequence of DNA markers on the genetic map as well as the EST markers. Candidate genes underlying major quantitative trait loci (QTL) for submergence tolerance on chromosome 9 were identified 1.35 Mbp from *O. sativa* subsp. *japonica* cv. Nipponbare. We predicted a total of 228 CDS features on the submergence tolerance region, of which 104 are related to transposable elements (TEs). A putative function could be assigned to 21 genes, with another 51 genes annotated as expressed, leaving 71 that encode hypothetical proteins. On the basis of these finding, the gene density with none transposable element-related gene in the critical region of submergence tolerance is about 10.8 Kb/gene.

Key words: Gene annotation, genome sequencing, repetitive sequence, retrotransposon, submergence tolerance, *Sub*QTL9

INTRODUCTION

Rice (*Oryza sativa*) has been chosen as a model crop to be sequenced by an international sequencing consortium, the IRGSP (International Rice Genome Sequencing Project) (Sasaki and Burr, 2000). Because of rice's genome is the smallest among crops and available information of linkage and physical map have already been established (Arumuganathan and Earle, 1991; Harushima *et al*., 1998). Moreover, over 100,000 expressed sequence tags (ESTs) and full-length cDNA have been reported (Yamamoto and Sasaki, 1997) and mostly mapped. Completion of the rice genome sequence and accumulation of related genomic resources present an excellent opportunity to solve intractable problems in rice that are governed by quantitative inheritance.

Submergence tolerance is characterized by complex phenotype-associated traits which can be generalized as the ability to survive and continue growing after several days in submerged conditions. Recently genetic linkage between submergence tolerance and shoot elongation was clearly shown by QTL analysis in segregating recombinant inbred line (Siangliw *et al*., 2003). Both submergence tolerance and suppression of elongation were coincidentally mapped on chromosome 9 (Toojinda *et al*., 2003) quantitative trait loci (QTLs) analysis for submergence tolerance had been identified *Sub*QTL9 as a major QTL. Together with genomics analysis and a major QTL are important keys to discover those candidate genes that regulate the tolerant plant for submergence trait on chromosome 9 in rice.

The advent of large-scale genomic sequencing has been conducted on a major QTL response for submergence tolerance from chromosome 9 resulting in a significant increase in gene sequence information. However, our knowledge of the function and interaction is lacking of these newly discovered genes. In recently years, the genomic sequence around the major QTL of submergence tolerance has been completely sequenced (http://www.ncbi.nlm.nih.gov/GenBank/index.html). Therefore, the investigation of structures and functions of responsible genes is essential for understanding the genetic mechanisms controlling submergence tolerance in rice.

Corresponding Author: Apichart Vanavichit, Rice Gene Discovery Unit, National Center for Genetic Engineering and Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Nakorn Pathom 73140, Thailand
Tel: (66) 034 355194 Fax: (66) 034 355197

In this study, sequence data of *Sub*QTL9 region were collected by anchoring sequence-based genetic markers to the sequence map. We then identified genes as candidates with respect to position, which located on the major QTL of submergence tolerance. We also presented the characterization of all the putative open reading frames and its representative sequence on the critical region of major QTL for submergence tolerance, a contig with a sequence localized around 1.35 Mbp of the long arm of rice chromosome 9.

MATERIALS AND METHODS

Submergence tolerance region: Over the past few years, a major QTL controlling submergence tolerance was mapped to a 15.5 cM region of chromosome 9. It was cloned and sequenced at Rice Gene Discovery Unit during 2001-2003 (The Rice Genome Project in Thailand). This region is flanked by S10709 and RZ698s, which were shown a map distance of approximately 15.5 cM (CentiMorgan) of three mapping populations (Siangliw *et al*., 2003; Toojinda *et al*., 2003). The first was the F1-derived doubled haploid lines (DHL) from a cross between IR49830-7-1-2-2 [IR49830; a submergence tolerant breeding line from the International Rice Research Institute (IRRI)] and CT6241-17-1-5-1 [CT6241; a submergence intolerant line from Centro International de Agriculura Tropical (CIAT)]. This population consists of 65 DHLs and was developed using an anther culture method (Lentini *et al*., 1995). The second population derived from the cross between FR13A (an Indian land race cultivar and one of the most submergence tolerant lines) and CT6241 consists of 172 Recombinant Inbred Lines (RILs). This population was developed via Single Seed Descent (SSD) at International Rice Research Institute (IRRI). The third population consists of 188 F2 plants developed from a cross made at Kasetsart University between Jao Hom Nin (a black rice with moderate tolerance to submergence) and Khaow Dawk Mali 105 (KDML105; an aromatic traditional cultivar from Thailand that commercially important but intolerant of submergence).

Anchoring sequence-based genetic markers to the sequence map: In order to retrieve BAC clones aligned in the region proximal to *Sub*QTL9, the electronic polymerase chain reaction (ePCR) approach was used in this study (Schuler, 1997). The primers for the analysis were obtained retrieved from either the Gramene database (www.gramene.org) (Ware *et al*., 2002) or the Japan Rice Genome Project (RGP) database (Sasaki, 2001). Markers were correlated with the physical map by ePCR, run against the rice genome BAC sequences, retrieval from the Gramene database and by BLAST (Altschul *et al*., 1990) alignment searches of the rice BAC sequences.

Sequencing data: We downloaded an available sequence data found in GenBank (www.ncbi.nlm.nih.gov) and from the rice databases of The Institute for Genomic Research (TIGR; www.tigr.org). The complete sequences of the critical region of submergence tolerance bear the accession number as P0651G05, P0645D04, P0603H10, OSJNBa0009H03, OJ1190B07, P0663H05, P0453B09, B1151D08, B1106B03, B1054C11, B1043F11, OSJNBa0044K01, P0592C05 and OJ1381_H04 were subjected to analysis.

Annotation: Annotation involved both DNA and protein database searches and gene prediction program. BLAST (Basic Local Alignment Search Tool) (Altschul *et al*., 1990) search and gene prediction programs were used to determine the genomic sequence of PACs for protein-coding genes. BlastX and Blastn programs were employed to search for non-redundant and EST databases at the NCBI (National Center for Biological Information) and TIGR (The International Genomic Research). BLOSUM 62 was utilized in the BLAST program as a default matrix. The homology was considered significant if an E-value was greater than -15 for at least 95% identity for 250 nucleotide tract. Potential coding regions were predicted by GenScan (Arabidopsis) (Burge and Karlin, 1997) GeneMark HMM (rice) (Lukaskin and Borodovsky, 1998) and Grail (Arabidopsis) (Uberbacher and Mural, 1991). Predicted protein sequences were investigated against a non-redundant amino acid database using blastp (Altschul *et al*., 1997). A database for plant *cis*-acting regulatory DNA elements (PLACE) was utilized in the analysis of the DNA sequences (Higo *et al*., 1999).

Analysis of DNA sequence and the putative open reading frames: Repetitive sequences were masked using Repeat-masker and tRNA Scan-SE (http://ftp.genome.wash-ington.edu/RM/RepeatMasker.html) (Lowe and Eddy, 1997) Annotated genes were categorized according to the homology level. Each of putative open reading frames (ORFs) were analyzed or scanned by several sequence motif searches including PROSITE (Hofmann *et al*., 1999), BLOCKs (Henikoff *et al*., 1999), ProDom (Corpet *et al*., 1999), PRINTS (Attwood *et al*., 1999) and Pfam (Bateman *et al*., 2004) (http://www.sanger.ac.uk/ Software/ Pfam). Proteome analysis database (http://www.ebi.ac.uk/ pro-teome/) and ExPASy Molecular Biology server (http://www.expasy.ch/) were used as a resource to identify the functional classification of proteins in this region.

RESULTS

Alignment of the rice genetic map onto the physical map: Chromosome 9, the second smallest segment in rice, has the total map distance approximately 22 cM. The sub-centromeric region was chosen as the entry point for genome sequencing based on several interesting features for both structural and functional genomics. Nearby its centromere, the major QTL for submergence tolerance was mapped into a 15.5 cM region where two molecular markers S10709 and RZ698 were located (Fig. 1). Several traits at this major QTL were characterized with plant survival, plant elongation, induced shoot elongation, visual tolerance score and leaf senescence. The *Sub*QTL9 was detected consistently in experiments across all years and in the genetic backgrounds of all three mapping populations. The candidate sequence interval spanning the candidate QTL flanked by S10709 to RZ698 was then

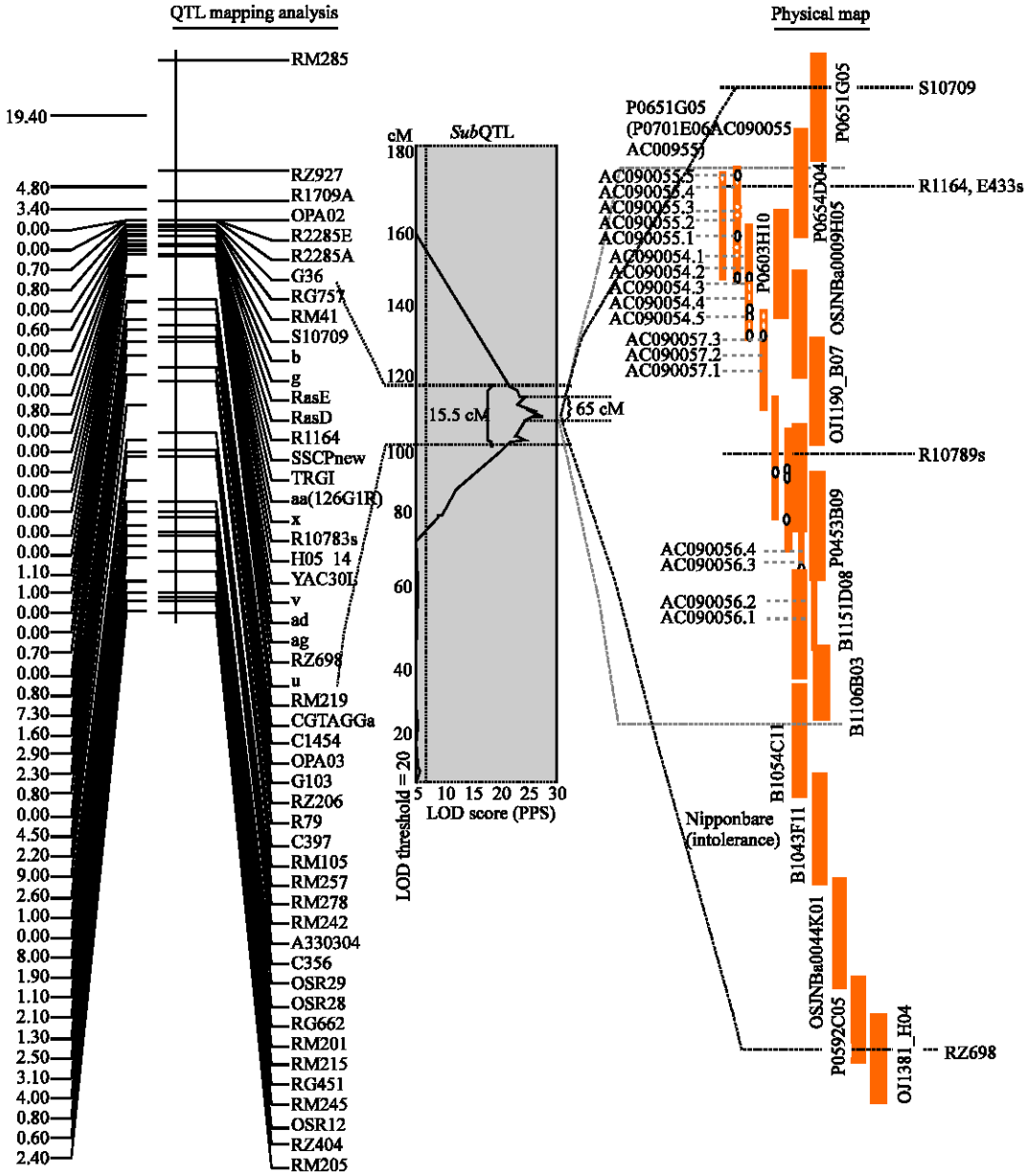

Fig. 1: A 1.35 Mb contiguous region consisting of 14 BAC/PAC clones were anchored on the long arm of chromosome 9 corresponding to about 15.5 cM of the genetic map. The sequence of DNA markers on the genetic map as well as the EST markers on the YAC physical map used for anchoring the BAC/PACs are also shown

Table 1: Feature of 14 BAC/PAC contigs along the major QTL of submergence tolerance region on chromosome 9

Clone name	Accession No.	Size (kp)	CDS feature (Candidate genes)	Genetic markers
P0651G05	AC090055	143.1	24-(7)	S10709
P0645D04	AC090054	139.0	21-(5)	R1164
P0603H10	AC090057	127.3	19-(1)	-
OSJNBa0009H03	AP005818	151.7	24(1)	-
OJ1190B07 P0663H05	AP005562	140.8	21-(5)	-
P0453B09	AC090056	161.9	25-(5)	R10783s
B1151D08	AP005705	173.5	23	-
B1106B03	AP005907	152.8	17 (3)	-
B1054C11	AP006449	155.5	16	-
B1043F11	AP006464	141.5	13 (1)	-
OSJNBa0044K01	AP006156	167.3	22 (2)	-
P0592C05	AP005839	160.6	18 (6)	-
OJ1381_H04	AP004756	150.4	26 (12)	-
Total	AP004011	130.9	20 (5)	RZ698
		1.35 Mbp[A]	228 (53)	15.5 cM

[A]The 1.35 Mbp region was represented a total of non-overlapping contig from 14 BAC/PAC clones between the molecular marker of S10709 and RZ698

identified. The assembly of fourteen BAC/PAC clones in this interval yielded a contiguous sequence region of 1.35 Mbp within which we identified 228 predicted gene structures. The complete sequence of the critical region of submergence tolerance bears the accession number shown in Table 1.

Gene characterization and classification: In general, the functional genome annotation is based on the idea that some sequence similarities detected between two proteins mean that they are homologous. They may come from the same ancestor and share the same biochemical function. Here, we present the characterization of all the putative open reading frames and its repetitive sequences in the critical region, a contig of which sequences localized around 15.5 cM of submergence tolerant genes of rice chromosome 9.

A total of 228 CDS features were found on the submergence tolerance region. Of these, a function could be assigned to 28 (12.2%), while 25 (10.9%) were annotated as encoding an expressed protein including a full-length cDNA(s) and 175 (76.7%) were predicted as encoding a hypothetical protein without similarity to an entry in the public databases. In total, 53 potential protein-coding genes were identified, along with a 1.35 Mbp region as predicted through computational search by the packages indicated in Materials and Methods. These genes were named SUB 1-53 as a working nomenclature (Table 2). The average length of the gene model was 2,789 bp (from start to stop codon) and contains 6 exons with an average size of 487 base pair (Table 3). The average predicted coding region was 1,171 bp (range 399-5127). Almost all genes contained intron (s), with an average intron (s) size of 569 nucleotides. Of the gene with introns, the average number of introns in a gene was 4 (range 1-24). Within the 1.35 megabase-region, all of genes were predicted and confirmed by homology search (Table 5). The putative function was analyzed to find the presence of functional domains. The results showed that the biological function of the ORFs was related to essential processes of gene regulation and gene signaling that might be involved in transcription regulation, signal transduction pathway, ubiquitilation and proteolysis based on their amino acid motif. According to promoter analysis, we found that many types of the DNA sequence elements were classified as a hormonal signaling in various plants. At least in the case of submergence tolerance, a gene underlining the QTL might be driven by a mediator molecule of signal transduction pathways involved in plant hormones such as gibberellins, ethylene, ABA and auxin (Table 4). The roles of these regulatory proteins in submergence tolerance are elucidated in the next experiments.

Analysis of intergenic region: Altogether, the 228 predicated genes, including exons and introns, account for around 48.93% of the 1.35 Mbb contig. In other words, about half of the 14 BAC/PAC of *Sub*QTL9 sequence are intergenic regions. The overall GC content of the contig is 43.50%, with an average content of 61.89% in exons and an average content of 39.94% in other region (introns plus intergenic region).

Benefits of molecular marker as simple sequence repeats (SSRs) have been well known not only to detect genetic variations within or between species but also to develop molecular markers tightly linked to agronomically important traits in breeding programs. Our analysis of 1.35 Mbp *Sub*QTL9 region revealed a total of 65 SSRs using RepeatMasker and Censor with default parameters. Most of them were classified into 64% of a di-nucleotide with lengths ranging from 6 to 46 bp. While other classes of simple repeats were found as tri-nucleotide (25%), tetra-nucleotide (8%) and mononucleotide (4%). The

Table 2: The characteristics of the 53 open reading frames (ORFs) found in submergence tolerance region

Sub No.	CDS (bp)	Exon	Aa.	Putative characteristics	Predicted localization	Biological function
1	720	6	239	Rhodanese-like family protein	Mitochondrial	-
2	600	6	245	Nucleic acid binding	Chloroplast	RNA recognition motif (RRM)
3	687	7	696	Isomerase activity	-	ATP binding protein
4	1080	12	359	Oxidoreductase activity	Membrane	Electron transport, metabolism
5	1404	11	467	Transcription factor	Nuclei	HBP-1b(c1)
6	624	10	207	Mnd1	Nuclei	
7	573	3	190	Unknown	Plasma membrane	4TM
8	870	7	289	Splicing factor 4	Nuclei	Spliciosome
9	681	2	226	GTP-binding protein Rab 11d	Cytoplasm	
10	1323	2	440	Unknown	Membrane	-
11	5223	8	1740	Receptor like-kinase	Membrane	TRAP170
12	921	2	306	Transparent 1	Chloroplast	WIP5 protein
13	747	2	248	Ankyrin-like protein	Chloroplast	ANK4
14	1218	5	405	Ubiquitin conjugating enzyme 7	Nuclei	UbE3
15	924	8	307	Serine/threonine phosphatase	Cytoplasm	Metallophos
16	468	2	155	Unknown	Cytoplasm	-
17	909	6	302	Chaperone GrpE type 2	Mitochondrial	GrpE2
18	1401	12	466	Enhancer of polycomb-like protein	Nuclei	CKS
19	2559	2	852	Unknown	Nuclei	-
20	270	0	89	Unknown	-	-
21	1608	23	535	NHE-8	Plasma membrane	7TM
22	699	0	232	Transcription factor	Nuclei	C-repeat/DRE-binding factor
23	751	1	250	Transcription factor	Nuclei	AP2/EREBP
24	591	2	196	Antigen receptor-like protein	Membrane	3TM
25	2022	12	673	Epstein-Barr virus EBNA-1-like protein	Nuclei	ULP_PROTEASE
26	2067	1	688	Far-red impaired response protein	Nuclei	FAR1 family, SWIM, WRKY
27	819	5	272	Unknown	Plasma membrane	F-box protein
28	372	2	123	Unknown	Nuclei	Unknown
29	270	3	89	Unknown	Cytoplasm	Unknown
30	1185	1	394	Unknown	Cytoplasm	Kelch repeat, F-box
31	243	2	80	Unknown	Microbody	2TM
32	447	2	148	Ubc2 enzyme	Cytoplasm	Ubiquitilation
33	3645	1	1214	LRR receptor kinase	Plasma membrane	Phytosulfokine
34	1278	10	425	Unknown	Nuclei	Cis-trans isomerase
35	2748	19	915	Aspartate kinase-homoserine dehydrogenase	Mitochondrial	Bifunctional enzyme
36	525	2	174	Serine/threonine protein kinase	Chloroplast	-
37	1302	8	433	UQ_con, Ubc2	Nuclei	Ubc2
38	1488	1	495	Unknown	Nuclei	OsNAC protein
39	327	1	108	Unknown (Xs domain)	Cytoplasm	Xs protein
40	1698	10	565	Unknown	Mitochondrial	NTPase domain
41	588	1	195	Unknown	1TM-outside	-
42	1401	5	466	Unknown	Cytoplasm	Protease
43	1464	1	487	Unknown	Plasma membrane	Enzyme activity
44	966	5	321	Chlorophyll a/b-binding	Plasma membrane	2TM
45	2391	13	796	WD40, G beta repeat	Chloroplast stroma	-
46	1449	9	482	Ferrochelatase	Plasma membrane	1TM
47	447	4	148	Ubiquitin-conjugating enzyme	Mitochondrial	Ubi-con
48	1554	1	517	Monosaccharide transporter 6	Plasma Membrane	10TM
49	1227	9	408	Phosphoenolpyruvate/phosphate translocator	Plasma Membrane	7TM
50	525	2	174	GPI-anchored protein	Plasma Membrane	2TM
51	618	1	205	Unknown	Plasma Membrane	1TM
52	1857	19	618	Phosphofructokinase	Cytoplasm	PFK
53	1503	9	300	ADP-glucose pyrophosphorylase small subunit	Mitochondrial	-

number of n ranged was showed at 5 to 29 bp. These are typical microsatellite motifs. They were found upstream of the 5'-UTR of several genes and sometimes in the ORF of specific genes.

Transposable elements (TEs): One feature that might contribute to high recombination hotspot is the high number of transposon and retrotransposon flanking genes. The TEs were predominated with unclassified (38%), CACTA, En/Spm sub-class (30%), Ty1-gypsy sub-class (25%), Ty1-copia sub-class (7%) and mutator sub-class (1%). Most of them were long terminal repeats (LTRs) of retrotransposon with two kinds of direct repeats and inverted repeats as Ty3-gypsy sub-class. Surprisingly, according to the analysis by tRNAscan-SE, *Sub*QTL9 region has no tRNA gene in the 1.35 Mbp

Table 3: Statistics of rice chromosome 9 at a critical region of mapping QTL controlling submergence tolerance

Feature	Statistic
Total number of BACs/PACs	14
Total BAC length	1.35 (Mbp)
GC content	43.50%
Average length of gene model	2789 (bp)
Total number of SSRs	85
Total number of genes	228
Known/putative genes	28 (12.2%)
Expressed genes	25 (10.9%)
Hypothetical genes	175(76.7%)
Transposable elements	61
Gene density	5.92 (Kb)
Average CDS length	1171 (bp)
Average exon size	487 (bp)
Average number of exon/gene	6
Average intron size	1981 (bp)
Average number of intron/gene	5
Integrated genetic markers	24

Table 4: Analysis of promoter region of candidate genes associated with the major QTL was identified on rice chromosome 9

Domain type	Gene family function	Cis-acting elements	Genetically characterized factor
MYB	Secondary metabolism, Cellular morphogenesis, signal transduction in plant growth, abiotic and biotic stress responses, circadian rhythm and dorsoventrality	MRE-like sequences	AtMYB2,ATR1,CCA, CPC, GL1,LHY,WER
AP2/EREBP	Flower development, cell proliferation, secondary metabolism, abiotic and biotic stress responses, ABA response and ethylene response	GCC box	ABI4, ANT, AP2, CBF1-3, DREB1A-C,DREB2A, ERF1
bZIP	Seed-storage gene expression, photomorphogenesis, leaf development, flower development, defense response, ABA response and GA biosynthesis	G-box, Dof, OCS	ABI5, HY5, PAN
Z-C_2H_2	Flower development, flowering time, seed development and root nodule development	GARE	FIS2, SUP,HRT
WRKY	Defense response	W Boxes	TDBA12

fragment. However, sequences with significant similarities to rice mitochondrial protein (HGWP repeat containing protein, rhodanese-like family protein and chaperone heat shock protein) were found. The average physical genetic distance of 87 kb per cM was surprisingly high considering the sequencing area is proximal to the centromere. On the basis of these findings, we found that the gene density in the gene-rich region with none-TEs-related gene is approximately 5.92 kb/gene. High gene density with small physical to genetic distance makes this subcentromeric region particularly gene-rich and hotspot for recombination.

DISCUSSION

In this study, we anchored rice QTL map to the rice physical map to identify submergence stress tolerance candidate genes based on previously reports of QTL position. We chose a sequence of molecular markers for our attempt to link submergence tolerance genotype to phenotype using bioinformatics tool. Because of strong combined genetic evidence for the existence of a major effect QTL for stress tolerance and available rice genome sequences in the region. A set of candidate genes of known or related function were identified in *Sub*QTL9 region using computational tools and rice gene annotation.

Gene Identification and Density: In fact, abiotic stress tolerance, particularly submergence, is the priority target trait in Southeast Asia dealing with rice plants. Among cereals, rice has the smallest genome with a size of only 430 Mbp. Consequently, the complete set of genes and their genomic locations could be identified via genomic sequencing with an acceptable investment. Chromosome 9 is the second smallest chromosome among the total of 12 chromosomes. At the short arm of chromosome 9 located the only Nucleolar Organizer Region (NOR) of the rice genome is 22 Mb or 6.3% of the rice genome. However, a major QTL controlling submergence tolerance was mapped at the long arm. Several evidences suggested that reduced plant elongation was a critical morphological change for rice to survive during submergence (Siangliw *et al*., 2003; Toojinda *et al*., 2003). This critical region might be associated with the regulation of their plant adaptation under flooding.

The candidate gene approach is one such tool as a promising method of merging QTL analysis with the extensive data available on the cloning and characterization of genes. Of the 53 genes identified by the prediction programs in the 1.35 Mbp contig, only two lacks intron while the others contain from 1 to 53. The gene density with non-TEs related protein coding sequence is about one gene per 10.8 kb, similar to those obtained from other rice BAC/PAC clones sequenced,

resulting one gene per 9.9 kb (International Rice Genome Sequencing Project). However, our result of gene density with TE-related gene, 7.31 kb, is consistent with the gene density reported for the entire *indica* genome (Yu *et al*., 2005). This report excluded TE-related gene was effected one gene per every 7.2 kb.

Retrotransposons in the *Sub*QTL 9: Retrotransposons, a group of mobile elements that transpose via the RNA intermediate, are important components of the eukaryotic genomes (Boeke and Corces, 1989). The structure of retrotransposons resembles that of integrated retroviruses, with long terminal repeats (LTRs) and an internal domain encoding a group-specific antigen and a polyprotein (Pol). The Pol region has conserved domains characteristic of protease, reverse transcriptase, integrase and RNase H genes and this region is present in both retrotransposons of 1.35 Mbp-region. All BAC/PAC clones containing retrotransposons are currently in the annotated rice BAC/PAC database. About half of them are LTR-retrotransposons, as those present in this region. Many of the rice retrotransposons including hypothical protein gene, such as the two described here, are pseudogenes because they are interrupted by stop codons or no initiation codon in CDS, no termination codon in CDS and frameshifts. Thus, the rice genome indeed contains many retrotransposons, although many are inactive.

Functional gene analysis: Under flooding, plant limits oxygen supply and gas diffusion that affect directly to anaerobiosis of submerged plant parts or tissues. Low oxygen supply impedes mitochondrial respiration, since di-oxygen is the final electron acceptor in the respiration chain. As a consequence, submergence-induced elongation is mediated by the interplay between three phytohormones, ethylene, abscisic acid and gibberellin. Most of plant responses to abiotic stress are controlled by a regulator of transcription or a sensing molecule in hormonal signaling pathway. In total of 53 candidate genes in the *Sub*QTL region, 14 of them are a nuclear localization protein, 12 of them are a functional protein in mitochondria and chloroplast, while 8 of them are cytoplasm protein. The remainders are predicted for biological function of 16 transmembrane proteins and only one of microbodi protein. Based on biological function and similarity search using computational tools, these predicted genes might control a part of plant adaptation under submergence stress.

To collect further supporting evidence for the function of the predicted genes in the region, EST sequence and a full-length cDNA database were aligned by BLAST against candidate gene sequence. These analyses reveal that 51 candidate gene predictions match expressed sequence in several libraries. Regulatory protein of their candidate genes were clustered into three distinct *Sub*QTL9 surrounded by large number of retrotransposons forming three gene-dense islands. These islands might be essential for maintaining submergence tolerance in several mechanisms of signal transduction pathway, gene regulation of transcription factors and ubiquitilation pathway. According to promoter analysis result, a gene underlining the QTL plays important roles in signal transduction pathways that involve in plant hormones, gibberellin, ethylene, ABA and auxin. The roles of these candidate genes in submergence tolerance will be elucidated in the further perspective model.

ACKNOWLEDGMENTS

We would like to express their sincere thanks to the Rockefeller Foundation and Rice Gene Discovery Unit, National Center for Genetic Engineering and Biotechnology and DNA Technology Laboratory, National Center for Genetic Engineering and Biotechnology and Center of Agricultural Biotechnology (CAB) for their support of this research.

REFERENCES

Altschul, S.F., G. Warren, M. Webb, W.M. Eugene and J.L. David, 1990. Basic local alignment search tool. J. Mol. Biol., 215 (3): 403-410.

Altschul, S.F., T.L. Madden, A.A. Axhaffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman, 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res., 25 (17): 3389-3402.

Arumuganathan, K. and E.D. Earle, 1991 Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep., 9 (3): 208-218.

Attwood, T.K., D.R. Flower, A.P. Lewis, J.E. Mabey, S.R. Morgan, P. Scordis, J. Selley and W. Wright, 1999. PRINTS prepares for the new millennium. Nucleic Acids Res., 27 (1): 220-225.

Bateman, A., L. Coin, R. Durbin, R.D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E.L.L. Sonnhammer, D.J. Studholme, C. Yeats and S.R. Eddy, 2004. The Pfam protein families database. Nucleic Acids Res., 32 (suppl_1): 138-141.

Boeke, J.D. and V.G. Corces, 1989. Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol., 43 (1): 403-434.

Burge, C. and S. Karlin, 1997. Prediction of complete gene structures in human Genomic DNA. J. Mol. Biol., 268 (1): 78-94.

Corpet, F., J. Gouzy and D. Kahn, 1999. Recent improvements of the prodom database of protein domain families. Nucleic Acids Res., 27 (1): 263-267.

Harushima, Y., M. Yano, A. Shomura, M. Sato, T. Shimano, Y. KuboKi, T. Yamamoto, S.Y. Lin, B.A. Antonio and A. Parco, 1998. A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics, 148 (1): 479-494.

Henikoff, J.G., S. Henikoff and S. Pietrokovski, 1999. New features of the blocks database servers. Nucleic Acids Res., 27 (1): 226-228.

Higo, K., Y. Ugawa, M. Iwamoto and T. Korenaga, 1999. Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res., 27 (1): 297-300.

Hofmann, K., P. Bucher, L. Falquet and A. Bairoch, 1999. The PROSITE database, its status in 1999. Nucleic Acids Res., 27 (1): 215-219.

Lentini, Z., P. Reyes, C. Martinez and W.M. Roca, 1995. Androgenesis of highly recalcitrant rice genotypes with maltose and silver nitrate. Plant Sci., 110 (1): 127-138.

Lowe, T.M. and S.R. Eddy, 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res., 25 (5): 955-964.

Lukaskin, A.V. and M. Borodovsky, 1998. Genemarkhmm: New solutions for gene finding. Nucleic Acids Res., 26 (4): 1107-1115.

Sasaki, T. and B. Burr, 2000. International Rice Genome Sequencing Project: The effort to completely sequence the rice genome. Curr. Opin. Plant Biol., 3 (2): 138-141.

Sasaki, T., 2001. The Rice Genome Project in Japan. In: Dealing with Genetically Modified Crops, Wilson, R.F., C.T. Hou and D.F. Hildebrand (Eds.). AOCS Press, Champaign Illinois, pp: 102-109.

Schuler, G.D., 1997. Sequence mapping by electronic PCR. Genome Res., 7 (5): 541-550.

Siangliw, M., T. Toojinda, S. Tragoonrung and A. Vanavichit, 2003. Thai jasmine rice carrying QTL ch9 *Sub*QTL submergence tolerance. Ann. Bot. (London), 91 (2): 255-261.

Toojinda, T., M. Siangliw, S. Tragoonrung and A. Vanavichit, 2003. Molecular genetics of submergence tolerance in rice: Quantitative traits loci (QTLs) analysis of key traits. Ann. Bot. (London), 91 (2): 243-253.

Uberbacher, E.C. and R.J. Mural, 1991. Locating protein-coding regions in human DNA sequences by a multiple sensor neural network approach. Proc. Natl. Acad. Sci. USA, 88 (24): 11261-11265.

Ware, D., P. Jaiswal, J. Ni, X. Pan, K. Chang, K. Clark, L. Teytelman, S. Schmidt, W. Zhao and S. Cartinhour, 2002. Gramene: A resource for comparative grass genomics. Nucleic Acids Res., 30 (1): 103-105.

Yamamoto, K. and T. Sasaki, 1997. Large-scale EST sequencing in rice. Plant Mol. Biol., 35 (1): 135-144.

Yu, J., J. Wang, W. Lin, S. Li and H. Li *et al*., 2005. The genomes of *Oryza sativa*: A history of duplications. PLoS. Biol., 3 (2): e38.