Plant Omics Journal POJ

POJ 4(5):239-249(2011) ISSN:1836-3644

Comparative analysis of the genomic regions flanking *Xa21* locus in *indica* and *japonica* ssp. of rice (*Oryza sativa* L.)

Anirudh Kumar[#,1] , Waikhom Bimolata[#,1], Gouri Sankar Laha[2], R. Meenakshi Sundaram[3], Irfan Ahmad Ghazi[*,1]

[1]Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Gachibowli Prof. C.R. Rao Road , Hyderabad 500046, India
[2]Department of Plant Pathology, Directorate of Rice Research, Rajendranagar, Hyderabad 500030, India
[3]Department of Biotechnology, Directorate of Rice Research, Rajendranagar, Hyderabad 500030, India

[#]These authors contributed equally to the work.

***Corresponding author: irfan@uohyd.ernet.in**

Abstract

Comparative analysis of a 100 kb region flanking of major bacterial blight resistance gene *Xa21* (3.57 kb) was performed in the two subspecies of rice *Oryza sativa* L. ssp. *japonica* cv. Nipponbare and *Oryza sativa* L. ssp. *indica* cv. 93-11 to understand the evolution and divergence of *Xa21* locus. A total of 12 genes in *japonica* and 14 genes in *indica* were predicted and annotated in this region. Functional annotation revealed the presence of 4 genes and 8 genes in *japonica* and *indica*, respectively, which could be putatively associated with disease resistance in the 100 kb region of *Xa21* locus. The study also revealed that 50% of *japonica* genes and 42.8% of *indica* genes in the genomic region of interest were transposable elements protein coding genes. Analysis of each predicted gene in this region revealed more or less similar GC content in both the subspecies. A total of 109 SSRs have been identified in the region of interest in both *indica* and *japonica*. The numbers of repeated motifs were observed to decrease with the increased number of nucleotides. Interestingly, most of the leucine rich repeat (LRR) gene products were predicted to be localized in the plasma membrane and the transposable element related protein coding genes were localized in the nucleus. Phylogenetic tree analysis revealed that the majority of predicted genes with similar functions of both the subspecies were grouped together.

Keywords: Markers; Microsynteny; Transposon; Locus; SSRs.
Abbreviations used: IRGSP: International Rice Genome Sequencing Project, LRR: Leucine Rich Repeat, NCBI: National Center for Biotechnology Information, PAMP: Pathogen-Associated Molecular Patterns, SSR: Simple Sequence Repeat, SSRIT: Simple Sequence Repeat Identification Tool.

Introduction

Japonica and *indica* diverged as two separate rice subspecies as a result of constant genetic differentiation over long periods of cultivation and accumulated genetic diversity (Morishima and Oka, 1981; Glaszmann, 1987; Wang and Tanksly, 1989; Blair et al., 1999). It is assumed that the two rice subspecies diverged around 0.44 million years ago (Ma and Bennetzen, 2004; Tang et al., 2004; Vitte et al., 2004 and Tian et al., 2006) and they must have possibly diverged due to nucleotide rearrangement, substitution, insertion and deletion (Wang et al., 2003). Comparative analysis between these two ssp. revealed a significant variation in the genic and intergenic regions of their genome (Feng et al., 2002; Han and Xue, 2003). Comparative genomics helps us in finding new genes, non-coding functional elements and above all, it helps in establishing an evolutionary relationship between species. With the availability of complete genome sequence of *indica* and *japonica,* comparative genome analysis between different rice species and other plant species will help in the identification of regions that are highly conserved and rapidly evolving. Today rice has become a model crop for functional and comparative genomic studies of cereal crops due to availability of complete genome sequences in the public domain and smaller genome size. The major bacterial blight resistance gene, *Xa21* was identified by Khush et al.(1990) from *Oryza longistaminata* and isolated from IRBB21 by Song et al.(1995) using map - based cloning strategy. This gene belongs to multigene family and class 5 resistance (R) genes (Bent, 1996). It encodes a receptor kinase like protein carrying LRR (leucine rich repeat) in the putative extracellular domain presumably for recognition, a single pass transmembrane domain and a serine/threonine kinase intracellular domain for subsequent signal transduction. Apart from confronting disease resistance, this gene also plays an important role in plant development just like other disease resistance genes such as *ZmPto* of maize (Zou et al., 2011). *Xa21* mediated resistance increases gradually from susceptible at juvenile stage to resistance at adult stage (Wang and Leung, 1998). In the present study, a 100 kb region flanking the *Xa21* locus was selected in both *japonica* and *indica* for comparative analysis of organization of the genomic region in the vicinity of *Xa21* gene. We decided to elucidate the genes present around the *Xa21* locus, their physical position and sub-cellular localization, identification of SSR markers, microsynteny analysis Etc. keeping in mind that the comparative analysis of this region

may disclose some facts useful for functional genomics and/or molecular plant breeding.

Results and discussions

Gene prediction, annotation and classification

In the vicinity of the 100 kb region flanking *Xa21* locus, a total of 12 genes were predicted in *japonica* and 14 in *indica*. Equal numbers of genes were distributed both on the negative and the positive strand of both the subspecies. A *japonica* gene *Osjp12,* coding for retrotransposon protein was observed to possess the longest coding sequence with a cDNA length of 4849 bp followed by *Osjp10* (4829 bp) and *Osjp08* (4536 bp). *Osjp12* had the maximum number of exons (11) in *japonica* (Table 1). In case of *indica*, *Osin09* possessed the longest cDNA length with 11,939 bp followed by *Osin04* (3108 bp) and *Osin14* (2932 bp). *Osin01* possessed the maximum number of exons (12) with a cDNA length of 2736 bp (Table 2). The gene density in *indica* was one gene per 7.85 kb while in *japonica,* it was one gene per 8.62 kb. The gene density was more in this region as compared to the overall gene density in *Oryza sativa* as reported by IRGSP which was one gene per 9.9 kb (IRGSP, 2005). The higher density of gene in the vicinity of *Xa21* could be explained in the light of co-localization and clustering of R-genes as reported by Hulbert et al.,(2001) and Ghazi et al.,(2009). The study of the genomic region encompassing gall midge resistance gene *Gm4* and *Gm5* also provide evidence that putative R genes are often organize in clusters (Dubey and Chandel, 2010). Clustering of R genes is a result of tandem duplications of paralogous sequences (Meyers et al., 2003 and Richly et al., 2002). The annotation of the predicted genes revealed that *Osjp03*, *Osjp07* and *Osjp11* code for proteins possessing LRR domain while *Osjp08* code for LRR protein kinase. There were four genes *Osjp03,Osjp07,Osjp08 and Osjp11* which could be putative disease resistance genes in *japonica* in the 100 kb region due to presence of leucine rich repeat domain (Table 3), whereas, there were seven such genes in *indica*. These genes were *Osin01, Osin03, Osin04, Osin05, Osin06, Osin08* and *Osin12.* An interesting observation was found in case of *Osin14,* which encodes for putative retrotrasposon but had leucine rich repeats in its conserved domain suggesting that a recombination process might have occurred in this region. The repetitive structure of LRR coding region favors recombination within this region (Ronald, 1998). *Osin01* possessed a conserved domain for ABC transporter, an ATP binding site and a p-loop (Table 4). Hence, it may belong to class 3 families of R-genes. P-loop NTPase was found as a conserved domain in *Osin03*. So, this gene could be of NB-ARC NACHT (nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4 domain) type. *Osin06* and *Osin08* were observed to code for LRR protein kinase but interestingly, these two genes were found to be part of a single gene which had been separated by transposon related gene *Osin07* (Fig.1), probably, by the process of transposition. Comparative sequence analysis showed that a recombination event must have occurred within the coding region of *indica Xa21* allele. Song et al*.,* (1997) had reported the evidence for recombination in the intergenic regions of the *Xa21* gene family members. Overall there were six predicted genes coding for transposon protein each in *japonica* and *indica*. Sequence identity between *Xa21* gene (Os11g0559200) and *Osin08* was 100% while between *Xa21* and *Osin06* was 99% (Fig.1).

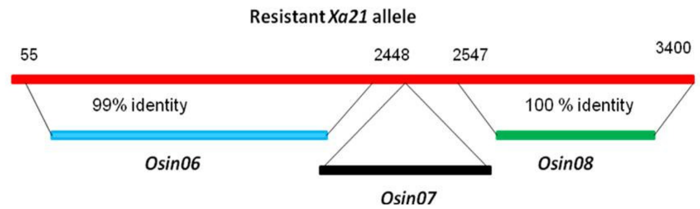

Fig 1. Alignment position of *Osin06* and *Osin08* with resistant (reference) *Xa21* allele. Position of *Osin07* indicates its insertion between *Osin06* and *Osin08* at *Xa21* locus in the long arm of chromosome 11 of *Oryza sativa* ssp. *indica.*

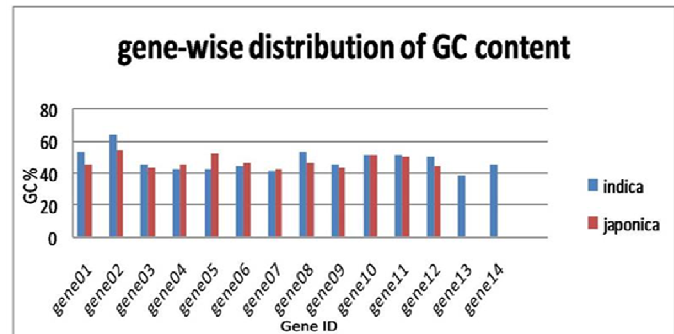

Fig 2. Gene wise distribution of GC content in *japonica* and *indica* covering 100 kb region of interest. *Indica* and *japonica* genes are represented by blue and red colored bars, respectively. X axis represents the predicted genes in order and Y-axis represents the percentage of GC content. Gene01 represents gene no. 1 of *indica* and *japonica* (*Osin01* and *Osjp01*, respectively) and gene02 represents gene no. 2 of *indica* and *japonica* (*Osin02* and *Osjp02*, respectively) and so on.

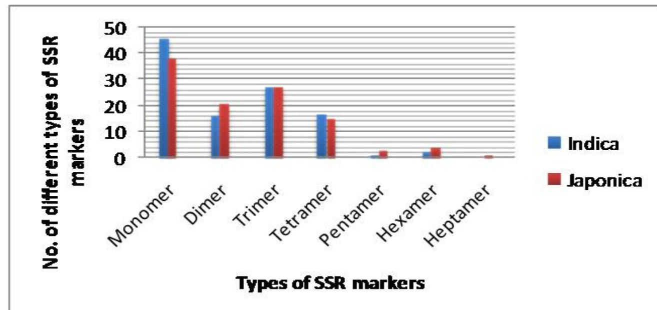

Fig 3. Distribution of different types of SSR markers in the 100 kb region of *Xa21* locus

Transposable elements

The presence of abundant transposons belonging to different super families has been well documented in rice (IRGSP, 2005). The overall transposon content in rice genome is 35% (IRGSP, 2005). According to rice genome annotation release 6, there were 16,185 transposon element (TE) related genes in the rice genome. Transposable elements have been

Table 1. List of predicted genes of *japonica* in the 100 kb region of *Xa21* locus and their annotation.

Gene	No. of exons	Start	End	cDNA length bp	Strand	Score	E-value	Function	Ortholog Hits	Paralog Hits	GC content (%)
Osjp01	5	2228	6029	832	+	42	0.78	gatD; glutamyl-tRNA(Gln) amidotransferase subunit D Pfam: Asparaginase PROSITE: ASN_GLN_ASE_1 ASN_GLN_ASE_2	835	23	45.0
Osjp02	3	8983	10883	1005	-	44	0.24	ypc00013; flavoprotein oxidoreductase protein Pfam: Flavin_Reduct PROSITE: PROKAR_LIPOPROTEIN	689	6	54.7
Osjp03	4	17196	24801	4088	+	5549	0.0	Os11g0558900; hypothetical protein Pfam: LRRNT_2 LRR_1	836	722	43.3
Osjp04	5	32976	33970	631	-	1104	0.0	Os10g0164800; hypothetical protein Pfam: p450 Peptidase_C48 PROSITE: SER_RICH ULP_PROTEASE	841	392	45.45
Osjp05	2	34155	36673	2443	-	4559	0.0	Os07g0525000; hypothetical protein Pfam: Peptidase_C48 PROSITE: ARG_RICH PRO_RICH NLS_BP ULP_PROTEASE	911	2769	52.47
Osjp06	1	39431	42736	3305	-	5727	0.0	Os07g0525900; hypothetical protein Pfam: Whi5 Transposase_21 Cps15 Peptidase_C48 Chal_sti_synt_N FAE1_CUT1_RppA ACP_syn_III Chal_sti_synt_C ACP_syn_III_C PROSITE: PRO_RICH NLS_BP ULP_PROTEASE	920	2247	46.1
Osjp07	1	43751	46516	2765	+	587	e-164	Os11g0558900; hypothetical protein Pfam: LRRNT_2 LRR_1	836	722	42.4
Osjp08	6	50055	59214	4536	+	2159	0.0	Os11g0569600; hypothetical protein Pfam: LRRNT_2 LRR_1 Myco_arth_vir_N Pkinase Pkinase_TyrAPH PROSITE: PROTEIN_KINASE_ST PROTEIN_KINASE_ATP PROTEIN_KINASE_DOM	917	1621	46.5
Osjp09	6	69359	73391	2655	+	468	e-128	Os10g0157900; hypothetical protein Pfam: Whi5 PROSITE: IG_MHC	725	78	43.3
Osjp10	4	76194	78607	4829	+	240	4e-60	Os10g0345200; hypothetical protein	0	0	51.0
Osjp11	4	84226	86528	1148	-	58	2e-05	Os11g0558900; hypothetical protein Pfam: LRRNT_2 LRR_1	836	722	50.0
Osjp12	11	91960	102669	4849	-	1235	0.0	Os11g0305400; hypothetical protein Pfam: Dimerisation Methyltransf_2 RVT_1 RnaseH zf-H2C2 PROSITE: ASN_RICH	880	178	44.2

classified into class I and class II super families based on their transposition mechanism (Kang and Kang, 2008). Class I TE transposes through RNA intermediates and high copy numbers of this transposon is present in the rice genome (Kunze and Weil, 2002). However, classes II TEs are present in low copy number and consist of terminal inverted repeats (IRGSP, 2005). Retrotransposon is an example of class I TE and CACTA family belongs to class II TE. In our study, we had identified six transposon related genes, each from *japonica* and *indica*. Five genes encoded putative Em/Spm CACTA like transposon protein and one gene encoded putative Ty3/gypsy like retrotransposon protein in *japonica* while in case of *indica,* equal numbers of class I and class II TE encoding genes were present. Transposon related genes accounted for 50% of the predicted genes in the 100 kb region in *japonica* while it was 42.8% in *indica*. The presence of higher number of TE genes in this region in the two subspecies suggest that this part of the genome might be still actively evolving. We also observed that the co-linearity of genes between the two subspecies within the 100 kb region has also been affected by the presence of transposon related elements.
En/Spm(enhancer/suppressor-mutator)like transposon which has consensus sequence CACTA in their terminal inverted repeats was dominant in *japonica*. Class II TEs were reported to be predominantly located around high gene density region (IRGSP 2005). *Xa21* locus is rich in TE and there are 14 transposon like elements in the noncoding region of *Xa21* family members. These *Xa21* associated transposon like elements may play a major role in variability among *Xa21* gene family members and its evolution (Song et al*.,* 1998). Protcomp V 8.0 (http://linux1.softberry.com/ berry.phtml) analysis for the sub cellular location of proteins showed that 80% of the predicted TEs were nuclear in localization while the remaining were cytoplasmic or mitochondrial in both the rice lines (Table 9).

GC Content in the 100 kb region

The average GC content of the predicted genes was 47.83% in *indica* and 47.03% in *japonica*. The overall GC content of the 100 kb region was 42.63% in *indica* and 43.52% in *japonica*. It was in congruence with the earlier report (Eguiarte et al*.,* 2003; Saccone and Pesole, 2003) showing 43-44% average GC content of the rice genome. While Kumar et al. (2007) reported that the average GC content was higher in case of *indica* as compared to *japonica* in the 100 kb region of *Pi-k*^{h} locus. When gene-by-gene comparison was carried out, seven genes of *indica* (excluding *Osin04, Osin 05, Osin06* and *Osin07*) had a higher GC content than that of *japonica*, whereas 4 genes of *japonica* i.e. *Osjp04, Osjp05, Osjp06* and *Osjp07* have slightly higher GC content in comparison to *indica* (Fig.2). Gene number 10 of both *indica* and *japonica* retain equal GC content (51%). The average GC content of *indica* was higher due to presence of two extra genes explicitly, viz., *Osin13* and *Osin14*. The higher GC content in monocot genes in comparison to eudicot has been reported (Carels and Bernardi, 2000). The average GC content of *Oryza sativa* (monocot) was 43-44% whereas *Arabidopsis thaliana* (dicot) contain only 36% (Eguiarte et al*.,* 2003; Saccone and Pesole, 2003). It has been shown that those gene copies which belong to subfamilies of very similar sequences (presumably undergoing gene conversion) have a higher GC content than unique gene copies (presumably not undergoing gene conversion) Galtier, (2003). This study revealed that GC content varies between the two subspecies of the same genus. High GC content

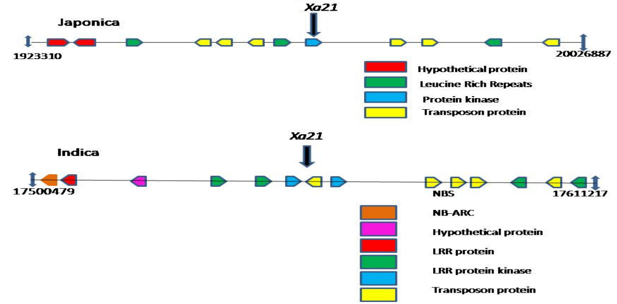

Fig 4. Physical map of *japonica* and *indica* predicted genes in the 100 kb regions of *Xa21* locus. Vertical arrow indicates the position of *Xa21* allele in *japonica* and *indica*. Position of arrow heads indicates the direction of the predicted genes.

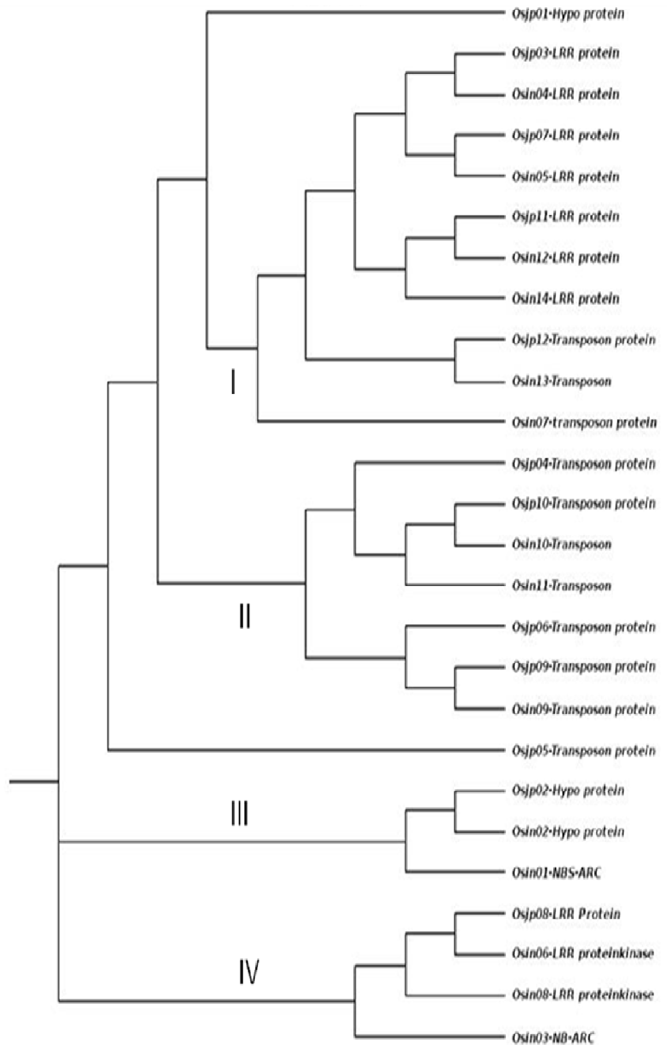

Fig 5. Phylogenetic analyses of the predicted genes of *japonica* and *indica* subspecies. All the genes of *japonica* and *indica* were clustered into two large (I and II) and two small (III and IV) clusters. Cluster I is the largest one with 10 genes followed by cluster II with 7 genes. Cluster III and IV held 3 and 4 genes, respectively.

Table 2. List of predicted genes of *indica* in the 100 kb region of *Xa21* locus and their annotation.

Gene	No. of exons	Start	End	cDNA length	Strand	Score bits	e-value	Functions	Orthologs	Paralogs	GC content %
Osin01	12	1371	6560	2736	-	1929	0.0	Os11g0416900; hypothetical protein Pfam: ABC_tran MMR_HSR1 ABC2_membrane PROSITE: ABC_TRANSPORTER_1 ABC_TRANSPORTER_2	927	223	53.28
Osin02	1	7296	7922	626	-	234	8e-59	Os04g0385600; hypothetical protein Pfam: F-box Sel1 zf-MYND PROSITE: ZF_MYND_1 ALA_RICH ZF_MYND_2	569	117	64.3
Osin03	1	19947	20870	923	-	64	2e-07	Os10g0124300; hypothetical protein Pfam: NB-ARC NACHT	915	393	45.16
Osin04	3	34136	38214	3108	+	3872	0.0	Os11g0558900; hypothetical protein Pfam: LRRNT_2 LRR_1	836	722	42.8
Osin05	2	43747	46567	2758	+	504	e-139	Os11g0565000; hypothetical protein Pfam: LRRNT_2 LRR_1 PROSITE: PROKAR_LIPOPROTEIN	698	680	42.76
Osin06	2	50055	52531	2026	+	831	0.0	Os11g0569600; hypothetical protein Pfam: LRRNT_2 LRR_1 Myco_arth_vir_N Pkinase Pkinase_Tyr APH PROSITE: PROTEIN_KINASE_ST PROTEIN_KINASE_ATP PROTEIN_KINASE_DOM	917	1621	44.72
Osin07	1	55447	56688	1241	-	270	3e-69	Os11g0305400; hypothetical protein Pfam: Dimerisation Methyltransf_2 RVT_1 RnaseH zf-H2C2 PROSITE: ASN_RICH	880	178	41.9
Osin08	3	58647	60499	747	+	581	e-163	Os11g0569600; hypothetical protein Pfam: LRRNT_2 LRR_1 Myco_arth_vir_N Pkinase Pkinase_Tyr APH PROSITE: PROTEIN_KINASE_ST PROTEIN_KINASE_ATP PROTEIN_KINASE_DOM	917	1621	53.59
Osin09	5	79452	83481	11939	+	492	e-136	Os10g0157900; hypothetical protein Pfam: Whi5 PROSITE: IG_MHC	725	78	45.18
Osin10	3	86330	87328	792	+	226	2e-56	Os10g0345200; hypothetical protein	0	0	50.91
Osin11	4	87422	89749	986	+	58	2e-05	Os10g0157900; hypothetical protein Pfam: Whi5 PROSITE: IG_MHC	725	78	50.87
Osin12	6	95391	98017	1317	-	58	2e-05	Os11g0558900; hypothetical protein Pfam: LRRNT_2 LRR_1	836	722	50.47
Osin13	3	103413	104882	1086	-	90	5e-15	Os09g0441600; hypothetical protein Pfam: p450 Class_IIIsignal Retrotrans_gag zf-CCHC rve RVT_2 PROSITE: CYTOCHROME_P450 GLY_RICH INTEGRASE	905	775	38.7
Osin14	5	106326	110450	2932	-	1697	0.0	Pfam: Dimerisation Methyltransf_2 RVT_1 RnaseH zf-H2C2 PROSITE: ASN_RICH	880	178	45.11

genes were more stable and robust. Rice belong to grass family and grass genes are extremely GC rich compared with other angiosperms (Carels and Bernardi, 2000).

Distribution and identification of SSRs in the 100 kb Xa21 locus

SSR or simple sequence repeats are short stretches of tandem nucleotide repeats (Jarne and Lagoda, 1996). In our study, we have considered a minimum number of repeats as 7 for monomers (MNRs), 5 for dinucleotide repeats (DNRs), 4 for trincleotide repeats (TNRs) and 3 for tetranucleotide repeats (TTRs), pentanucleotide repeats (PNRs), hexanucleotide and heptanucleotide repeats (HNRs), respectively. SSRs are classified into class I and class II. Class I SSRs include the commonly used DNRs, TNRs and TTRs whereas MNRs, PNRs and HNRs are included in class II SSRs. The importance of SSR markers has been well documented. They are being used popularly as genetic markers in the study of genetic diversity (Cho et al., 2000), positional cloning (Xiao et al., 1998 and Zou et al., 2000) and marker assisted backcrossing in rice (Mackill, 2007). During the sequence analysis of the 100 kb region of *Xa21* locus, we identified a total of 109 SSRs each in both *indica* and *japonica*. Monomeric repeats were highest in number in both *japonica* and *indica* followed by trimeric and dimeric repeats. Monomeric repeats constitute for 42.2% and 34.8% of the total SSRs identified in *indica* and *japonica,* respectively. The frequency of A/T/G/C repeats was almost similar in both the subspecies. However, when we checked the presence of monomeric repeats in the exonic region, we found that only one monomer i.e. 'G' repeat was present in the *indica* gene *Osin02.* 'A' repeat was found to be the most abundant among the four MNRs. The number of tetrameric repeats was 15 in *japonica* and 17 in *indica*, PNRs and HNRs were present in very less percentage while only one heptameric repeat was present in *japonica* and completely absent in *indica* (Fig. 3). The frequency of SSRs were more in *japonica* except the monomeric type. Analysis of different repeat motifs revealed that the AT/TA repeat motif was the most common one in *japonica* accounting for 13.76% of the SSRs identified (Table 5). On the contrary AT/TA and AG/GA/CT/TC repeat motifs were equally present in *indica* representing 6.4% each of the total SSRs identified (Table 6). The abundance of AT repeat motif in different plant genomes followed by AG/CT and GT/CA has been shown by previous studies (Condit et al., 1991; Lagercrantz et al., 1993; Morgante and Olivieri, 1993; Wang et al., 1994; Powell et al., 1996). The same trend was observed in both the rice subspecies. However for trimeric repeats, GC rich motifs GGC/CCG/GCG/CGC showed the highest frequency of 17 and 21 in *japonica* and *indica,* respectively. GC rich trimeric repeats accounted for 15.5% in *japonica* and 19.26% in *indica*. In legumes such as soybean and chickpea, ATT repeat motif was abundantly found (Akkaya et al., 1992; Huttel et al., 1999), however, ATT repeat motif was completely absent in the 100 kb region of *Xa21* locus in both *indica* and *japonica*. Microsatellite studies on *Arachis hypogea* also showed less frequency of ATT repeat motifs (Cuc et al., 2008). The number of repeat motif for monomeric repeats ranged from 7-13, dimeric repeats 5-32 and for trimeric repeats, it was 4-6 in *japonica.* In *indica*, the number of repeat motif was comparatively less ranging from 7-17 for monomers, 5-14 for dimers and 4-8 for trimers. It was observed that the number of repeat motifs decreased as the number of nucleotide increased. An immense variation was observed in the number of SSRs present in the exonic and intronic regions. A total of 7 trimeric repeats were identified each from *indica* and *japonica* and one monomeric repeat from *indica* in the coding regions of the predicted genes. Other types of repeats included in our study were completely absent in the genic regions. The percentage of SSRs identified in the genic region was comparatively less than that of the intergenic region. The percentage frequency of SSRs present in the exonic region was only 7.33% and 6.42% of the SSRs identified, respectively in *indica* and *japonica*. The differences between the two subspecies may be the result of species diversification followed by individual evolution of this region; however, the functional significance of these SSRs markers is yet to be determined. An assessment of genes that are associated with microsatellites may also help in deciphering the evolutionary processes of the *Oryza* genome.

Microsynteny analysis for physical mapping of predicted genes

As a part of the comparative genomic study, microsynteny analysis between *japonica* and *indica* subspecies for the predicted genes of *Xa21* locus was carried out against the available database for chromosome 11 of *japonica* and *indica* in NCBI. The predicted genes were previously grouped according to their function. Microsynteny analysis showed 100 % homology for 11 genes and 99% homology for only one gene (i.e. *Osjp12*) with chromosome 11 database of *japonica* (Table 7). In case of *indica*, three genes namely *Osin04*, *Osin09* and *Osin14* showed 99% homology and the remaining 11 predicted genes showed 100% homology with chromosome 11 database of *indica* (Table 8). Based on the above study, physical maps of the two rice subspecies for the 100 kb region were generated which covered all of the predicted genes. From gene prediction analysis, we have identified those genes, *Osjp08* and *Osin06* of *japonica* and *indica,* respectively codes for protein receptor kinase. Alignment of these two predicted genes with the reference *Xa21* sequence also showed 100 % identity. We hypothesize that *Osjp08* and *Osin06* are the alleles of *Xa21* gene respectively for *japonica* and *indica* subspecies. Such high sequence conservation also suggests that receptor kinase of both subspecies are suitable candidates to provide resistance. At the same time, strength of disease resistance conferred by the *Xa21* gene is not only dependent on its presence but also on its expression. In both the rice subspecies, *Xa21* was flanked by LRR-encoding gene in the upstream and transposon related gene in the downstream region. In *japonica,* six predicted genes were located in the sense strand and remaining genes were in the antisense strand. Even in *indica*, equal numbers of the predicted genes (7) were present in sense and antisense strand (Fig. 4).

Sub-cellular localization of predicted gene products

Many world-wide web based servers are available to predict sub-cellular localization of proteins; for example PSORT (Nakai and kanehisa, 1991), ChoroP (Emanuelsson et al., 1999), MitoProt (Claros, 1995) etc. Subcellular location of the protein has a pivotal role in protein function. The availability of such tools that predict location from sequence is essential for complete characterization of the putatively expressed protein. Exploring the sub-cellular localization by experimentally practiced methods is time consuming, tedious and highly variable (Murphy, 2000). Recent developments in bioinformatic tools have lead to rapid growth of protein databank. To make the protein databank easily accessible,

Table 3. Protein prediction of the genes in 100 kb region of *japonica* by BLASTP tool.

Gene ID	Blast hit	Bit score	e- value	Function	Conserved domains
Osjp01	EEE52280.1	511	3e-143	Hypothetical protein OsJ_34265	RIP superfamily
Osjp02	ABA94297.1	681	0.0	Hypothetical protein LOC_Os11g35430	Nil
Osjp03	ABA94357.1	1779	0.0	Leucine Rich Repeat family protein	LRR_ribonuclease inhibitor and RT
Osjp04	ABA9453.1	455	1e-126	Transposon protein, putative, CACTA, En/Spm	Peptidase_C48
Osjp05	ABA94301.1	1686	0.0	Transposon protein, putative, CACTA, En/Spm	Nil
Osjp06	ABA94302.1	2310	0.0	Transposon protein, putative, CACTA, En/Spm	Transposase_21 superfamily
Osjp07	ABA94303.1	1881	0.0	Leucine Rich Repeat family protein	LRR
Osjp08	ABA94304.1	1711	0.0	Leucine Rich Repeat family protein	RT like superfamily, PKc_like family
Osjp09	ABA94306.1	1858	0.0	Transposon protein, putative, CACTA, En/Spm	Transposase_21
Osjp10	ABA94307.1	1504	0.0	Transposon protein, putative, CACTA, En/Spm	Nil
Osjp11	ABA94309.1	998	0.0	Leucine Rich Repeat family protein	LRRNT
Osjp12	ABA94310.1	3265	0.0	Retrotransposon protein, putative, Ty3-gypsy	Retrotransposon, NTP binding site

Table 4. Protein prediction of the genes in 100 kb region of *indica* by BLASTP tool.

Gene ID	Blast hit	Bit score	e-value	Function	Conserved functional domains
Osin01	AAX92830.1	1031	0.0	Hypothetical protein LOC_Os11g22350	2 ATP binding sites,2 walker A/P-loop,ABC transporter motif,Q-loop,P-loop NTPase
Osin02	AAX92829.1	163	9e-39	Hypothetical protein LOC_Os11g22360	Nil
Osin03	EEC68362.1	573	7e-162	Hypothetical protein OsI_36497	P -loop NTPase
Osin04	ABA94357.1	1549	0.0	Leucine Rich Repeat family protein	LRR_RI
Osin05	ABA94303.1	1774	0.0	Leucine Rich Repeat family protein	LRR
Osin06	ABA82756.1	993	0.0	Receptor kinase-like protein	ATP binding site,LRR, PTKc
Osin07	CAH65969.1	849	0.0	H0820C10.2, Ty3-gypsy	Integrase core domain
Osin08	ABA94304.1	341	4e-92	Leucine Rich Repeat family protein	PKc_like superfamily
Osin09	ABA94306.1	1023	0.0	Transposon protein, putative, CACTA, En/Spm	Transposase 2 superfamily
Osin10	ABA94307.1	502	1e-140	Transposon protein, putative, CACTA, En/Spm	Nil
Osin11	ABA94307.1	606	7e-172	Transposon protein, putative, CACTA, En/Spm	Nil
Osin12	ABA94309.1	829	0.0	Leucine Rich Repeat family protein	LRR
Osin13	ABA94310.1	521	4e-146	Retrotransposon protein, putative, Ty3-gypsy	RVT_2 superfamily
Osin14	AAV31274.1	1047	0.0	Retrotransposon protein, putative, Ty3-gypsy	LRR,RnaseH and RT

proteins are classified into twelve groups (1) chloroplast (2) cytoplasm (3) cytoskeleton (4) endoplasmic reticulum (5) extracellular (6) golgi apparatus (7) lysosome (8) mitochondria (9) nucleus (10) peroxisome (11) plasma membrane and (12) vacuole (Chou and Elrod, 1999)**.** The above classification has covered almost all the organelles and sub-cellular compartments. In this study, sub-cellular localization has been predicted through neural net program (ProtCompV 8.0) using amino acid sequence. The amino acid sequence might carry a signal which locates the sub-cellular position of the protein within the cell. In *indica*, gene *Osin09* belongs to group 2; *Osin01, Osin12* and *Osin14* belong to group 5; *Osin03, Osin10* and *Osin13* belong to group 8; *Osin02, Osin07*, *Osin08* and *Osin11* belong to group 9; *Osin04, Osin05* and *Osin06* belong to group 11 which has been predicted through neural net. For *japonica*, *Osjp01* belongs to group1, *Osjp02* and *Osjp04* belong to group 2, *Osjp11* belong to group 5, *Osjp05, osjp06*, *Osjp09, Osjp10* and *Osjp12* belong to group 9, then *Osjp03, Osjp07* and *Osjp08* belong to group 11 (Table 9). We observed that most of the LRR coding gene products were localized in the plasma membrane and the transposon proteins in the nucleus. The presence of LRR proteins in the plasma membrane is an important indication that LRR protein might be involved in recognition of pathogen's elicitors. It is in support of the already established hypothesis of LRR protein function (Song et al., 1995).

Phylogenetic analysis

A phylogenetic tree was constructed for the predicted genes of *indica* and *japonica* to illustrate the relationship of the genes (Fig.5). All the 12 genes of *japonica* and 14 genes of *indica* were observed to fall into two large clusters and two small clusters (Cluster I, II, III and IV). Cluster I was the largest one containing 10 genes. Out of 10 genes, 7 were encoding for LRR proteins and 3 for transposons related proteins. In Cluster II, all the 7 genes were transposons related. Cluster III held 3 genes and it included 2 hypothetical proteins and one NB-ARC protein whereas cluster IV was composed of 4 genes in which 3 were encoding for kinases and one was encoding for NB-ARC. The tree showed that the majority of predicted genes with similar function of both the subspecies were grouped together as per our expectation because of their common ancestry. In cluster I, all the genes encoding the putative LRR protein and transposons of both

Table 5. Frequency of dimeric and trimeric repeats or SSRs in the 100 kb *Xa21* locus (*japonica*).

Motif	Frequency	% Frequency
AT/TA	15	13.76
GC/CG	0	00.00
AG/GA/CT/TC	5	04.50
AC/CA/TG/GT	1	00.91
AAT/ATA/TAA/ATT/TTA/TAT	0	00.00
AAG/AGA/GAA/CTT/TTC/TCT	1	00.91
AAC/ACA/CAA/GTT/TTG/TGT	1	00.91
ATG/TGA/GAT/CAT/ATC/TCA	1	00.91
AGG/GGA/GAG/CCT/CTC/TCC	3	02.75
AGC/GCA/CAG/GCT/CTG/TGC	0	00.00
ACG/CGA/GAC/CGT/GTC/TCG	3	02.75
ACC/CCA/CAC/GGT/GTG/TGG	1	00.91
GGC/GCG/CGG/GCC/CCG/CGC	17	15.50

Table 6. Frequency of dimeric and trimeric repeats or SSRs in the 100 kb *Xa21* locus (*indica*)

Motif	Frequency	% Frequency
AT/TA	7	06.40
GC/CG	2	01.83
AG/GA/CT/TC	7	06.40
AC/CA/TG/GT	0	00.00
AAT/ATA/TAA/ATT/TTA/TAT	0	00.00
AAG/AGA/GAA/CTT/TTC/TCT	0	00.00
AAC/ACA/CAA/GTT/TTG/TGT	0	00.00
ATG/TGA/GAT/CAT/ATC/TCA	1	00.91
AGG/GGA/GAG/CCT/CTC/TCC	2	01.83
AGC/GCA/CAG/GCT/CTG/TGC	0	00.00
ACG/CGA/GAC/CGT/GTC/TCG	1	00.91
ACC/CCA/CAC/GGT/GTG/TGG	2	01.83
GGC/GCG/CGG/GCC/CCG/CGC	21	19.26

indica and *japonica* subspecies were observed to cluster together. However, *Osjp01* (hypothetical protein) and *Osjp05* (transposon protein) did not cluster with functionally similar protein.

Materials and methods

The present comparative study was performed on a 100 kb region flanking the *Xa21* gene (3.57 kb) on the long arm of chromosome 11 of *japonica* and *indica* ssp of rice.

Assembly of raw sequences

The genome sequences of chromosome 11 of *japonica* and *indica* were downloaded from the International rice genome sequencing project (IRGSP) database (http://rgp.dna.affrc.go.jp/irgsp) and Beijing rice genome sequence database (http://rice.genomics.org.in/rice/index2.jsp), respectively.
The IRGSP gene Os11g0559200, which encodes *Xa21*, was downloaded from gramene database (www.gramene.org). The position of *Xa21* on chromosome 11 of *japonica* and *indica*, respectively was located through BLASTN search tool (Altschul et al., 1990). The resistant allele of *Xa21* is 3.57 kb long and encodes a receptor kinase. A 100 kb sequence flanking *Xa21* locus (with ~ 50 kb upstream and ~ 50 kb downstream of the locus) was located with the help of BioEdit software (Hall, 1999). All the analysis were performed including *Xa21* locus (3.57 kb).

Gene prediction and annotation

Gene prediction from the 100 kb region flanking to *Xa21* locus of chromosome 11 in *japonica* and *indica* rice was carried out using HMM based gene structure prediction software FGENESH tool (www.softberry.com) trained for monocot plant species (Salamov and Solovyer, 2000). This software is one of the fastest and most accurate *ab initio* gene prediction program freely available (http://linux1.softberry.com/berry.phtml). To know the functions of each predicted genes, BLASTP in NCBI (www.ncbi.nim.nih.gov) and BLASTN at (www.genome.ad.jp) was carried out. The nomenclature of predicted genes in this region were given as *Osjp01, Osjp02* Etc. (*Oryza sativa* ssp. *japonica* gene number 1 and 2, respectively) and *Osin01, Osin02* Etc. (*O. sativa* ssp. *indica* gene number 1 and 2, respectively).

Determination of GC content and identification of SSR markers

The overall GC content of the 100 kb region as well as the gene-by-gene GC content was determined using online software http://tim.saraogtim.com (Tim221175/GC content). The software tool FastPCR (Kalendar et al., 2009) was used to identify monomers. SSRIT (simple sequence repeat identification tool) (Temnykh et al., 2001) available at gramene database (www.gramene.org) was used to find out other types of SSR markers present in the 100 kb region. The tool SSRIT uses pearl regular expressions to find perfect SSR repeats within a particular sequence.

Table 7. Microsynteny analysis of the predicted genes in the 100 kb region of the *japonica* cultivar group

Gene ID	Blast hit	Bit score #	E-value #	Homology	Start (bp)*	End (bp)*
Osjp01	Chr 11_japonica	499	1e-139	270/270 (100%)	19925537	19929338
Osjp02	Chr 11_japonica	1146	0.0	620/620 (100%)	19934192	19932292
Osjp03	Chr 11_japonica	5169	0.0	2799/2799 (100%)	19940553	19943312
Osjp04	Chr 11_japonica	344	4e-93	186/186 (100%)	19956470	19956285
Osjp05	Chr 11_ japonica	2630	0.0	1424/1424 (100%)	19959982	19958559
Osjp06	Chr 11_ japonica	6080	0.0	3306/3306 (100%)	19966045	19962740
Osjp07	Chr 11_ japonica	5083	0.0	2766/2766 (100%)	19967060	19969825
Osjp08	Chr 11_ japonica	4839	0.0	2620/2620 (100%)	19973364	19975983
Osjp09	Chr 11_ japonica	1971	0.0	1067/1067 (100%)	19995634	19996700
Osjp10	Chr 11_ japonica	1561	0.0	845/845 (100%)	20000828	20001672
Osjp11	Chr 11_ japonica	1138	0.0	652/652 (100%)	20008369	20007759
Osjp12	Chr 11_ japonica	2902	0.0	1574/1575(99%)	20021203	20019630

Table 8. Microsynteny analysis of the predicted genes in the 100 kb region of the *indica* cultivar group

Gene ID	Blast hit	Bit score #	E-value #	Homology	Start (bp)*	End (bp)*
Osin01	Chr 11_indica	1781	0.0	964/964 (100%)	17502813	17501850
Osin02	Chr 11_indica	1158	0.0	627/627 (100%)	17508401	17507775
Osin03	Chr 11_indica	1707	0.0	924/924 (100%)	17521349	17520426
Osin04	Chr 11_indica	3683	0.0	1996/1997 (99%)	17535468	17537464
Osin05	Chr 11_indica	3847	0.0	2083/2083 (100%)	17544226	17546308
Osin06	Chr 11_indica	3057	0.0	1655/1655 (100%)	17550534	17552188
Osin07	Chr 11_indica	2294	0.0	1242/1242 (100%)	17557167	17555926
Osin08	Chr 11_indica	702	0.0	380/380 (100%)	17560599	17560978
Osin09	Chr 11_indica	2084	0.0	1130/1131 (99%)	17582830	17583960
Osin10	Chr 11_indica	704	0.0	381/381 (100%)	17587158	17587538
Osin11	Chr 11_indica	1234	0.0	668/668 (100%)	17588269	17588936
Osin12	Chr 11_indica	848	0.0	459/459 (100 %)	17596552	17596194
Osin13	Chr 11_indica	887	0.0	480/480 (100%)	17604371	17603892
Osin14	Chr 11_indica	2809	0.0	1524/1525 (99%)	17608748	17607225

Footnote for Table 7 & 8: * Column 6 (start) and col 7 (end) indicates the starting position and last nucleotides of the predicted gene on chromosome 11. # Score describes the overall quality of an alignment. Higher number corresponds to higher similarity. Lower E-value indicates most significant score.

It can detect repeats between 2 to 10 bases in length, but eliminates mononucleotides repeats (http://www.gramene.org /db/searchers/ssrtool).

Sub-cellular location of genes

ProtcompV8.0 (www.softberry.com) software tool was used to predict the sub cellular localization of the gene products using protein sequence of the predicted genes. ProtCompv program recognizes animal /fungal and plant proteins separately. Its accuracy rate of protein localization prediction in the cell is 80-90%.

Physical positioning and classification of genes

To know the physical positions of each of the predicted genes, BLASTN analysis was carried out against *japonica* pseudomolecule chromosome 11 (build 5) and *indica* chromosome 11 as described by Kumar et al.,(2007). The genes were then classified based on their functions. A physical map of the 100 kb region of *Xa21* locus for *japonica* and *indica* genes was prepared manually.

Phylogenetic analysis

To study the phylogenetic relationship among the predicted genes of both the rice subspecies, a phylogenetic tree was drawn using clustalW program (http://www.ebi.ac.uk/Tools /msa/ clustalw2/) (Thompson et al., 1994). visualized using programs like treeviewX and Neighbor Joining (NJ) plot (Saitou and Nei,1987).

Conclusion

The present study is expected to facilitate cloning and expression analysis of candidate resistance genes present in the vicinity of *Xa21* gene and decipher their role in *Xa21* mediated resistance. One has to elucidate the co-expressed gene and its role in broad spectrum disease resistance. The disease resistance genes described in this study can be used as a probe to identify novel R gene or its homologues in rice and other cereals which can be used for developing molecular approaches in various breeding program. The predicted chromosomal location of the genes can be employed in the successive breeding through deployment of functional markers. From a previous study, we know that there are at least seven genes and one quantitative trait locus at *Xa21* locus, possibly encoding bacterial, viral and fungal resistance genes clustered within 30 cM on chromosome 11 (Song et al.,1995). Similarly, the *Pto*, *M*, *Cf9* and *N* resistance genes are all members of clustered gene families (Staskawicz et al*.,* 1995). Sequence comparison of the genes under study may give some clues regarding evolution of plant disease resistance. Through the prediction of protein sub-cellular localization, we found that 34.6% (9 genes) genes coding for transposon related elements and these genes were localized in the nucleus and 23% (6 genes) were localized in the plasma membrane and majority of those were coding for leucine rich repeats. Plasma membrane associated LRR proteins may be involved in the recognition of pathogen-

Table 9. Subcellular localization of *indica* and *japonica* genes.

Gene	Neural net prediction	Gene	Neural net prediction▲
	Location		Location
Osjp01	Chloroplast	*Osin01*	Extracellular (secreted)
Osjp02	Cytoplasmic	*Osin02*	Nuclear
Osjp03	Plasma membrane	*Osin03*	Mitochondrial
Osjp04	Cytoplasmic	*Osin04*	Plasma membrane
Osjp05	Nuclear	*Osin05*	Plasma membrane
Osjp06	Nuclear	*Osin06*	Plasma membrane
Osjp07	Plasma membrane	*Osin07*	Nuclear
Osjp08	Plasma membrane	*Osin08*	Nuclear
Osjp09	Nuclear	*Osin09*	Cytoplasmic
Osjp10	Nuclear	*Osin10*	Mitochondrial
Osjp11	Extracellular (secreted)	*Osin11*	Nuclear
Osjp12	Nuclear	*Osin12*	Extracellular
		Osin13	Mitochondrial
		Osin14	Extracellular (secreted)

Footnote:▲ Neural net is a program combined with ProtComp V software which predicts the localization of a particular protein in the cell.

associated molecular patterns (PAMPs), such as bacterial flagellin, lipopolysaccharides and fungal-oomycete cellulose-binding elicitor proteins that activate basal defense, a first line of defense against pathogens. *Xa21* belongs to receptor kinase class of disease resistance gene with great potential against bacterial blight; hence further molecular effort is required for its complete elucidation. It is a significant gene from socioeconomic point of view for sustainable management of bacterial blight and to combat hunger across the world.

Acknowledgement

We thank IRGSP and Beijing Genomics Institute for the availability of rice genome sequence in public domain. We thank Bioinformatics infrastructure facility funded by DBT and financial supports from XI plan seed money, UoH DBT-CREBB, DST-FIST, UGC-SAP-CAS DBT grant and maintenance grant from University of Hyderabad. Support from the CSIR as fellowship (Waikhom Bimolata) is gratefully acknowledged. We also thank Prof. A. R. Podile for critical reading of the manuscript.

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphism of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.

Blair MW, Panaud O, McCouch SR (1999) Inter-simple sequence repeat amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). Theor Appl Genet 98:780–792.

Bent AF (1996) Plant disease resistance genes: function meets structure. Plant Cell. 8:1757–1771.

Carels, Bernardi G (2000) The compositional organization and the expression of the *Arabidopsis* genome. FEBS Lett 472:302–306.

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

Chou KC and Elrod DW (1999) Protein sub-cellular location prediction. Protein Eng 12:107–118.

Claros, M (1995) MitoProt, a macintosh application for studying mitochondrial proteins. Comput Appl Biosci 11:441-447.

Condit R, Hubbel SP (1991) Abundance and DNA sequence of two base repeat regions in tropical tree genomes. Genome 34:66-71.

Cuc LM, Mace ES, Crouch JH, Quang VD, Long TD , Varshney RK (2008) Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). BMC Plant Biol 8:55

Dubey M, Chande 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.). Plant Omics J 3(5):140-148

Eguiarte LE, Castillo A, Souza V (2003) Molecular and genomic evolution in angiosperms. Interciencia 28:141– 147.

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8: 978-984.

Feng Zhang QY, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, Hu X, Jia P, Zhang Y, Zhao Q, Ying K, Yu S, Tang Y, Weng Q, Zhang L, Lu Y, Mu J, Lu Y, Zhang LS, Yu Z, Fan D, Liu X, Lu T, Li C, Wu Y, *et al* (2002) Sequence and analysis of rice chromosome 4. Nature 420:316–320.

Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 74:21–30.

Galtier N (2003) Gene conversion drives GC content evolution in mammalian histones**.** Trends Genet 19: 65-8.

Ghazi IA, Srivastava PS , Dalal V, Gaikwad K, Singh AK , Sharma TR , Singh NK, Mohapatra T (2009) Physical mapping, expression analysis and polymorphism survey of resistance gene analogues on chromosome 11 of rice. J Biosc 34:251-261.

Han B, Xue YB (2003) Genome-wide intraspecific DNA-sequence variations in rice. Curr Opin Plant Bio l6:134–138.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41: 95–98.

Hüttel B, Winter P, Weising K, Choumane W, Weigand F, Kahl G (1999) Sequence-tagged microsatellite-site markers for chickpea (*Cicer arietinum* L.). Genome 42:210-217.

Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: evolution and utilization. Annu Rev Phytopatho l39:285–312.

IRGSP (2005) The map-based sequence of the rice genome. Nature 436:793–800.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back review. Trends Ecol Evol, 11:424-429.

Kalendar R, Lee D, Schulman AH (2009) FastPCR Software for PCR Primer and Probe Design and Repeat Search. Genes, Genomes and Genomics, 3(1):1-14.

Kang HW, Kang KK (2008) Genomic characterization of *Oryza* species-specific CACTA-like transposon element and its application for genomic fingerprinting of rice varieties. Mol Breed 21:283–292.

Kunze R, Weil CF (2002) The hAT and CACTA superfamilies of plant transposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM Press,Washington 565–609.

Kumar SP, Dalal V, Singh NK, Sharma TR (2007) Comparative Analysis of the 100 kb Region Containing the *Pi-*k_h Locus Between *indica* and *japonica* Rice Lines. Genomics, Proteomics Bioinformatics 5:35- 44.

Khush GS, Bacalangco E and Ogawa T (1990) A new gene for resistance to bacterial blight from *O. longistaminata*. Rice Genet Newslett 7:121-122.

Largercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111-1115.

Ma JX and Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA 101:12404–12410.

Mackill DJ (2007) Molecular markers and marker-assisted selection in rice. In: R K Varshney and R Tuberosa (ed) Genomics assisted crop improvement. Genomics Applications in Crops. Springer, New York 2:147–168.

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15:809-834.

Morishima H, Oka HI (1981) Phylogenetic differentiation of cultivated rice. XXII. Numerable evaluation of the *Indica-Japonica* differentiation. Jpn J Breed 31:402–413.

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175-182.

Murphy RF, Boland MV, Velliste M (2000) Towards A systematics for protein subcellular location: quantitative description of protein localization patterns and automated analysis of fluorescence microscope images. Proc Int Conf Intell Syst Mol Biol 8:251-259.

Nakai K , Kanehisa M (1991) Expert system for predicting protein localization sites in Gram-negative bacteria. Proteins 11: 95-110.

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sc 1:215-222.

Richly E, Kurth J and Leister D (2002) Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. Mol Biol Evol 19:76-84.

Ronald PC (1998) Resistance gene evolution. Curr Opin Plant Bio 1:294-298.

Saccone C and Pesole G (2003) Handbook of comparative genomics: Principles and Methodology. J Wiley & Sons NJ USA

Salamov AA and Solovyer VV (2000) *Ab initio* gene finding in Drosophila genomic DNA. Genome Res 10:516-522.

Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol4:406–425.

Staskawicz BJ, Ausubel FM, Baker B, Ellis JG, Jones JD (1995) Molecular genetics of plant disease resistance. Science 268:661-667.

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by disease resistance gene, *Xa21*. Science 270:1804–1806.

Song WY, Pi LY, Wang GL, Gardner AJ, HoIsten BT, Ronald PC (1997) Evolution of the Rice *Xa2I* Disease Resistance Gene Family. The Plant Cell 9:1279-1287.

Song WY, Pi LY, Bureau TE, Ronald PC (1998) Identification and characterization of 14 transposon-like elements in the noncoding regions of members of the Xa21 family of disease resistance genes in rice. Mol Gen Genet 258:449-456.

Tang, J, H Xia, M Cao, X Zhan, W Zeng, S Hu, W Tong, J Wang, J Wang, J Yu, H Yang, and L Zhu (2004) A comparison of rice chloroplast genomes. Plant Physiol 135:412–420.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S and McCouch S (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.

Tian X, Zheng J, Hu S, Yu J (2006) The rice mitochondrial genomes and their variations. Plant Physiol 140:401–410.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22):4673-80.

Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Mol Gen Genomics 272:504–511.

Wang GL, Leung H (1998) Molecular biology of host-pathogen interactions in rice diseases. In: Shimamoto K (ed) Molecular Biology of Rice. Springer Verlag, Tokyo pp 201−232.

Wang J, Yu J, Wang J (2003) Systematical research of rice genome based on the whole genome sequence. World Sci-tech R&D 25:35–41(in Chinese).

Wang ZY, Tanksley SD (1989) Restriction fragment length polymorphism in *Oryza sativa* L. Genome 32: 1113–1118.

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1-6.

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD , McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909.

Zou JH, Pan XB, Chen ZX, Xu JY, Lu JF, Zhai W X , Zhu LH (2000) Mapping quantitative trait loci controlling sheath blight resistance in two rice cultivars (*Oryza sativa* L.). Theor Appl Genet 101:569–573.

Zou H, Zhongjian S, Wu Z, Zhang X, Liu H, Ma G, Huang C (2011) Isolation and analysis of ZmPto from maize, a homologue to Pto. Plant Omics J 4(2):53-59