International Journal of Genomics and Data Mining

Research Article

Analysis of Genome Sequence Variations Among Three U.S. Rice Varieties Showing Differential Quantitative Disease Resistance to Bacterial Panicle Blight and Sheath Blight

Bishnu Shrestha[1], Dong-Ha Oh[2], Maheshi Dassanayake[2], Jong Hyun Ham[1*]

[1]Department of Plant Pathology and Crop Physiology, Louisiana State University Agricultural Center, Baton Rouge, USA

[2]Department of Biological Sciences, Louisiana State University, Baton Rouge, USA

*Corresponding author: Jong Hyun Ham, Department of Plant Pathology and Crop Physiology, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA. Tel: +12255786798; Email: jham@agcenter.lsu.edu

Abstract

Bacterial Panicle Blight (BPB) and Sheath Blight (SB) are major rice diseases in the southeastern United States, and only quantitative disease resistance is known for these diseases. We analyzed draft genome sequence data for three U.S. rice varieties showing differential disease resistance traits for BPB and SB; Trenasse (long-grain, susceptible to BPB and SB), Bengal (medium-grain, susceptible to BPB and SB), and Jupiter (medium-grain, partial resistance to BPB and SB). Comparative genome sequence analysis along with 50 rice accessions revealed that the three US varieties are genetically close and clustered together, separated from other 50 accessions. According to this analysis, the long-grain semi-dwarf variety Trenasse is a *tropical japonica* type carrying a fraction of *indica* genome, while the medium-grain varieties Bengal and Jupiter are admixtures of *tropical* and *temperate japonica* types. Consistent with the breeding history and the phenotypic trait in grain-shape (but not with the phenotype in disease resistance), more variations were found in the Jupiter/Trenasse and Bengal/Trenasse pairs compared to the Jupiter/Bengal pair. The whole genome sequence information of these US rice varieties will be a useful resource for genetic studies of disease resistance to BPB and SB as well as development of new disease-resistant lines.

Introduction

Bacterial Panicle Blight (BPB) and Sheath Blight (SB) are important chronic diseases of rice in the southeastern United States, as well as other parts of rice-growing regions around the world [1-4]. BPB is caused by two bacterial pathogens, *Burkholderia glumae* and *B. gladioli*, and the phytotoxin toxoflavin is known as the major virulence factor of the pathogens [5]. Oxolinic acid is somewhat effective to control BPB, but this chemical is not registered to use for agricultural purpose in the U.S., and resistant strains of *B. glumae* have been reported indicating the limitation of this chemical as a reliable control measure [5]. SB is caused by the fungal pathogen, *Rhizoctonia solani*, and fungicide application is the primary way to control this disease. However, recent reports of *R. solani* isolates resistant to Strobilurin-type fungicides [6,7] indicate the urgent need of developing reliable alternative

management measures for this disease, including the cultivation of disease-resistant varieties. Quantitative (or partial) disease resistance, which is usually conferred by multiple Quantitative Trait Loci (QTLs), is thought to be the primary disease resistance mechanism of rice to BPB and SB. Qualitative (or complete) disease resistance, which involves specific interactions between a resistance gene of the host and its cognate a virulence gene of the pathogen [8], has not been found in BPB or SB of rice. Major QTLs associated with the disease resistance to BPB have been identified from several rice varieties. qBPB3-1 was identified on the short arm of chromosome 3 from the resistant variety, Teqing [9]. Later, qRBS1 (renamed later as RBG1) was mapped on the short arm of chromosome 10 with the 393-kb interval from the resistance variety, Nona Bokra [10]; and RBG2 was identified on the long arm of chromosome 1 with the 502 kb interval from the resistant traditional lowland variety, Kele [11]. Genetics of SB resistance

has been studied more intensively and widely compared to that of BPB resistance. More than 50 QTLs for SB resistance have been identified, and candidate genes responsible for SB resistance have also been found within some QTL regions [12-19]. Among the QTLs identified, qSB9-2 on chromosome 9 and qSBR11-1 on chromosome 11 are known to be major QTLs identified from multiple rice varieties [17,20]. Nevertheless, our knowledge of rice disease resistance to BPB and SB is still fractional and rudimentary.

Three U.S. rice varieties, Trenasse, Jupiter and Bengal, have been used for our genetic studies of the disease resistance to BPB and SB. These rice varieties cultivated in the southeastern United States show different phenotypes in terms of major agronomic traits, including grain shape and quantitative disease resistance to BPB and SB; in that Jupiter (a medium-grain variety) shows quantitative resistance to BPB and SB [21], while Bengal (a medium-grain variety) and Trenasse (a long-grain variety) are highly susceptible to both diseases [22,23]. Nevertheless, genetic studies of the disease resistance to BPB and SB with these materials have been hindered due to the lack of polymorphic markers to be used for linkage mapping of QTLs. Whole genome sequencing of rice accessions using a High-Throughput DNA Sequencing (HTS) platform provides excellent opportunities to determine genome-wide sequence variations associated with various traits of rice, such as disease resistance, and to develop more reliable molecular markers in a high-throughput and cost-efficient way [24-26]. Especially, HTS data are very useful for identification of DNA polymorphisms between genetically close genotypes and for fine mapping. For example, comparative analyses have been conducted for the identification of DNA polymorphisms within *japonica* or *indica* rice varieties [27-31], and the whole genomes of 13 rice inbred lines derived from US varieties were analyzed for the identification of candidate genes for sheath blight resistance [32]. In this study, we sequenced and analyzed the whole genome sequences of three US rice varieties, Trenasse, Jupiter and Bengal, which represents differential phenotypes in disease resistance/susceptibility to BPB and SB, using an HTS platform (Illumina Hiseq1000) in an attempt to develop new sequence-based molecular markers and to find a genome information basis for future genomic and genetic studies of rice disease resistance.

Materials and Methods

Rice Plants and DNA Extraction

One-week-old seedlings of the rice varieties, Jupiter (medium-grain, and moderately resistant to BPB and SB), Trenasse (long-grain, and susceptible to BPB and SB) and Bengal (medium-grain, susceptible to BPB and SB) were used to extract genomic DNA for whole genome sequencing. DNeasy Plant Mini Kit (Qiagen, Valencia, CA) was used for DNA extraction following manufacturer's instructions. The genomic DNA library for

sequencing was prepared using a Nextera DNA Library Preparation Kit (Illumina Inc., San Diego, CA), and 100-bp paired-end sequencing was processed using the Illumina HiSeq1000 platform (Illumina Inc., San Diego, CA) at Virginia Bioinformatics Institute (VBI) Genomics Lab at Virginia-Tech (Blacksburg, VA).

Mapping and Identification of Variants in Genome Sequences

The quality of the sequence reads were examined using Fast QC [33], and cleaned high quality reads were aligned to the rice reference genome version 7 released by the International Rice Genome Sequence Project (IRGSP) for the *japonica* rice variety Nipponbare, using Bowtie 2 [34]. Genome-wide variants including Single Nucleotide Polymorphisms (SNPs) and small insertions and deletions (indels) between the reference genome and three rice varieties were identified and processed using SAMtools [35], and annotated using SnpEff v3.5e [36].

Population Structure Analysis

Genetic relatedness of the three rice varieties with 50 rice accessions, including *temperate* and *tropical japonica*, aromatic and *indica* types, were analyzed by using FRAPPE [37]. SNP data of the 50 rice accessions for comparison were from the study by Xu, et al. [38], and the three US varieties for this study were analyzed in terms of admixture proportions with increasing value of K (number of clusters) from 3 to 7.

Pairwise Comparison

Pairwise comparisons between two varieties (Jupiter vs. Trenasse, Jupiter vs. Bengal, and Bengal vs. Trenasse) were performed with the help of vcftools, using the vcf files obtained from SnpEff analysis [39]. The output file from each comparison was filtered for common variants present in the varieties, which were identified from the comparison with the Nipponbare reference genome. Those variants between the varieties were again annotated and classified based on their effect on various regions in the genome and their functional type, using SnpEff v3.5e [36].

Statement of Reagent and Data Availability

All the rice DNA sequence data used for this study were deposited to the NCBI SRA (accession numbers: SRX4017380, SRX4017381, and SRX4017382). DNA samples of the rice varieties, Bengal, Jupiter and Trenasse, will be sent to the researchers upon request.

Results and Discussion

High-Throughput Sequencing (HTS) Data Obtained in This Study

In this study, the genomes of three US rice varieties, Jupiter, Trenasse and Bengal, were sequenced for comparative analysis of genome sequence variations. Fifty to 84 million of 100-bp paired-

end reads were obtained from each variety, resulting in 12X to 18X coverage based on the reference genome of 'Nipponbare' (Table 1).

	Jupiter	Trenasse	Bengal
Total reads (in millions)	79	84	50
Coverage	18X	19X	12X
Mapped with chromosomal genome (%)	95.78	91.02	96.33
Mapped with organelle genome (%)	13.21	11.05	9.43

Table 1: Total sequence reads (100-bp paired end) obtained from the high-throughput sequencing in this study, and percentage of the sequence reads mapped to the reference genome of *Oryza sativa* subsp. *japonica* (Nipponbare, IRGSP pseudomolecule version 7).

An average of more than 91% were aligned with the reference genome of the International Rice Genome Sequencing Project (IRGSP) pseudomolecule version 7 (Table 1). The reads from Bengal has the highest alignment percentage with 96.33% when mapped with the chromosomal reference sequence, followed by the reads from Jupiter (95.78%) and Trenasse (91.02%). In case of the HTS data mapped to the organelle sequences, the total sequence data of Bengal contained least portion of organelle sequences (9.43%), while 11.05% and 13.21% of the total sequence data were mapped to the organelle sequences with Trenasse and Jupiter, respectively (Table 1). Combined proportions of chromosomal and organelle sequences exceeded 100% in all varieties, indicating that some portions of sequence data were mapped to both chromosomal and organelle genomes.

Population Structure Analysis Using SNPs Identified From HTS

When compared to the reference genome (Nipponbare), total 1,007,294, 817,884, and 2,139,891 SNPs were identified in Bengal, Jupiter, and Trenasse genome sequences, respectively, indicating that the tropical *japonica* medium-grain varieties, Bengal and Jupiter, are genetically more closely related to Nipponbare (a *temperate japonica* variety) than the *tropical japonica* long-grain variety, Trenasse (Table 2).

	Bengal		Jupiter		Trenasse	
	SNPs					
	# of SNPs	# of SNPs per Mb	# of SNPs	# of SNPs per Mb	# of SNPs	# of SNPs per Mb
Chromosome 1	135,180	3,124	88,362	2,042	171,060	3,953
Chromosome 2	78,248	2,177	73,313	2,040	118,974	3,311
Chromosome 3	41,286	1,134	39,733	1,091	186,792	5,130
Chromosome 4	147,305	4,149	98,462	2,773	202,335	5,699
Chromosome 5	52,613	1,756	62,823	2,097	152,115	5,078
Chromosome 6	68,284	2,185	57,508	1,840	153,493	4,912
Chromosome 7	65,303	2,199	70,094	2,360	128,665	4,333
Chromosome 8	63,626	2,237	62,114	2,184	226,527	7,964
Chromosome 9	27,344	1,188	16,854	732	167,467	7,277
Chromosome 10	130,505	5,623	135,193	5,825	165,413	7,128
Chromosome 11	150,632	5,190	68,711	2,368	283,002	9,752
Chromosome 12	46,968	1,706	44,717	1,624	184,048	6,685
	Total: 1,007,294	**Average: 2,699**	**Total: 817,884**	**Average: 2,191**	**Total: 2,139,891**	**Average: 5,733.20**
	Insertions					
	# of insertions	# of insertions per Mb	# of insertions	# of insertions per Mb	# of insertions	# of insertions per Mb
Chromosome 1	8,630	199	5,765	133	10,276	237

Chromosome 2	4,552	127	4,428	123	6,409	178
Chromosome 3	2,733	75	2,414	66	11,991	329
Chromosome 4	5,763	162	4,823	136	7,186	202
Chromosome 5	2,475	83	3,611	121	7,178	240
Chromosome 6	3,082	99	3,104	99	7,395	237
Chromosome 7	3,227	109	3,766	127	5,927	200
Chromosome 8	2,903	102	3,157	111	9,976	351
Chromosome 9	1,026	45	514	22	6,428	279
Chromosome 10	5,367	231	6,080	262	6,206	267
Chromosome 11	6,033	287	3,353	116	11,700	403
Chromosome 12	2,158	78	2,288	83	6,601	240
	Total: 47,949	**Average: 129**	**Total: 43,303**	**Average: 116**	**Total: 97,273**	**Average: 261**
Deletions						
	# of deletions	# of deletions per Mb	# of deletions	# of deletions per Mb	# of deletions	# of deletions per Mb
Chromosome 1	8,448	195	5,636	130	9,661	223
Chromosome 2	4,302	120	4,275	119	6,193	172
Chromosome 3	2,594	71	2,339	64	11,070	304
Chromosome 4	5,678	160	4,824	136	6,943	196
Chromosome 5	2,594	87	3,629	121	6,881	230
Chromosome 6	3,085	99	3,088	99	7,138	228
Chromosome 7	3,192	107	3,749	126	5,776	194
Chromosome 8	2,743	96	3,063	108	9,319	328
Chromosome 9	1,081	47	537	23	6,016	261
Chromosome 10	5,153	222	5,681	245	6,081	262
Chromosome 11	5,863	202	3,056	105	10,476	361
Chromosome 12	2,146	78	2,241	81	6,169	224
	Total: 46,879	**Average: 126**	**Total: 42,118**	**Average: 113**	**Total: 91,723**	**Average: 246**

Table 2: Number of variants on individual chromosomes identified between the reference genome, Nipponbare and the three rice cultivars.

Same patterns were also observed with the numbers of insertions and deletions (Table 2). Among the SNPs identified, 1,188,460 non-ambiguous and biallelic SNP positions shared with at least one of the 50 rice accessions analyzed in the previous study by Xu, et al. (2012) were used to reconstruct the population structure including the three varieties along with the 50 previously sequenced rice accessions, using the program FRAPPE [37]. Among the number of populations (K) from 3 to 7 to set a population structure of the 53 accessions tested, FRAPPE produced a population structure with the highest likelihood at K=7 (Figure 1).
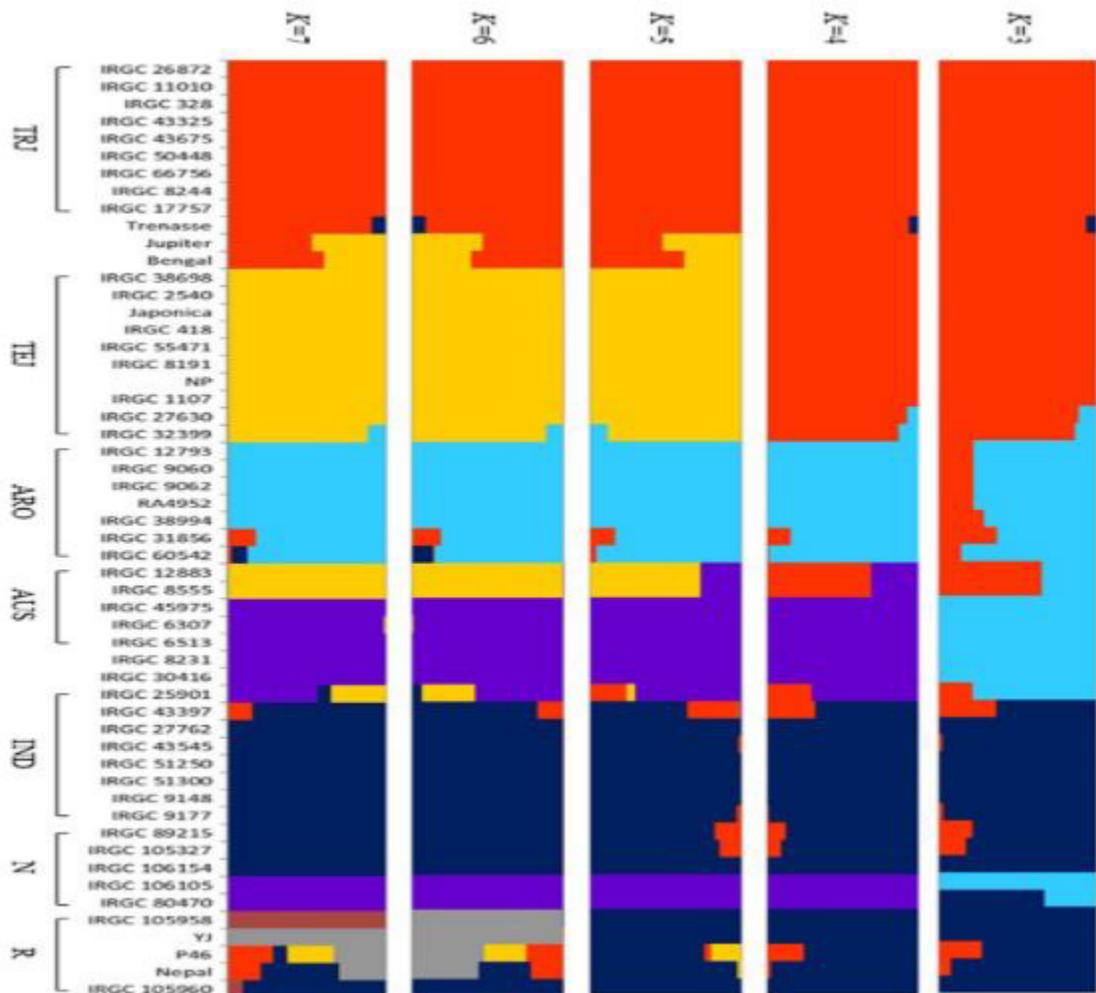
Figure 1: The genetic positions of the three rice varieties sequenced in this study (Jupiter, Trenasse, and Bengal) among the 50 rice accessions sequenced previously [38]. Total 1,188,460 non-ambiguous, biallelic SNPs were used for this population structure analysis (K = 3 to 7 with 10,000 iterations). The three rice varieties used in this study are clustered between *tropical japonica* and *temperate japonica* groups. The 50 rice accessions were labeled as in Xu et al., 2012. TRJ, *Tropical Japonica*; TEJ, *Temperate Japonica*; ARO, Aromatic rice; AUS, *aus* rice; IND, *Indica*; N, *O. nivara*; R, *O. rufipogon*. Each accession is represented vertically, and proportion of the ancestral genotypes is represented by different color segments in each vertical line.

In this analysis, it was revealed that Trenasse is mostly *tropical japonica* type with a trace of *indica* genome, while Jupiter and Bengal are admixtures of *tropical* and *temperate japonica* types (Figure 1). This result is congruent with the previous work by Lu, et al. [40], which studied the population structure and breeding patterns of 145 U.S. rice varieties based on genotypes of 169 Simple Sequence Repeat (SSR) markers. In that study, it was revealed that US varieties of *tropical japonica* medium-grain were positioned between those of *temperate japonica* and *tropical japonica* long-grain in terms of the genetic distance determined based on the SSR genotypic data [40]. The small portion of the Trenasse genome attributed to the *indica* type is seemingly derived

from the *indica* germplasms, which were utilized for introducing beneficial agronomic traits (e.g. semi-dwarf stature and high yield) to US rice varieties.

Pairwise Comparisons of Single Nucleotide Polymorphisms (SNPs) and Small Insertions and Deletions (InDels)

SNPs and InDels were first identified between the reference genome and each of three rice varieties, in which Trenasse and Jupiter showed the highest and lowest number of variations, respectively, when compared to the reference genome (Table 2). By comparing the polymorphism profiles compared to the reference genome, we also discovered variations between Jupiter

and Trenasse, Jupiter and Bengal, and Trenasse and Bengal (Tables 1-3).

	Jupiter vs Trenasse	Jupiter vs Bengal	Trenasse vs Bengal
Transitions (Ts)			
C/T	708,909	264,152	678,615
G/A	708,530	264,109	678,632
Total	1,417,439	528,261	1,357,247
Transversions (Tv)			
C/G	102,800	38,989	97,671
A/T	167,197	64,324	161,360
A/C	142,652	54,174	136,051
G/T	142,445	53,568	136,464
Total	555,094	211,055	531,546
Ts/Tv	2.55	2.50	2.55

Table 3: Number of transitions and transversions in SNPs identified from the comparisons, Jupiter vs. Trenasse, Jupiter vs. Bengal, and Trenasse vs. Bengal.

The distribution of SNPs at the chromosome level was observed by calculating the density of identified variations in 1-Mb intervals in each comparison (Figure 2A). Among the three pair-wise comparisons, the Jupiter/Trenasse and Trenasse/Bengal pairs show higher numbers of SNPs per 1Mb genome than the Jupiter/Bengal pair (Figures 2A and 2B).
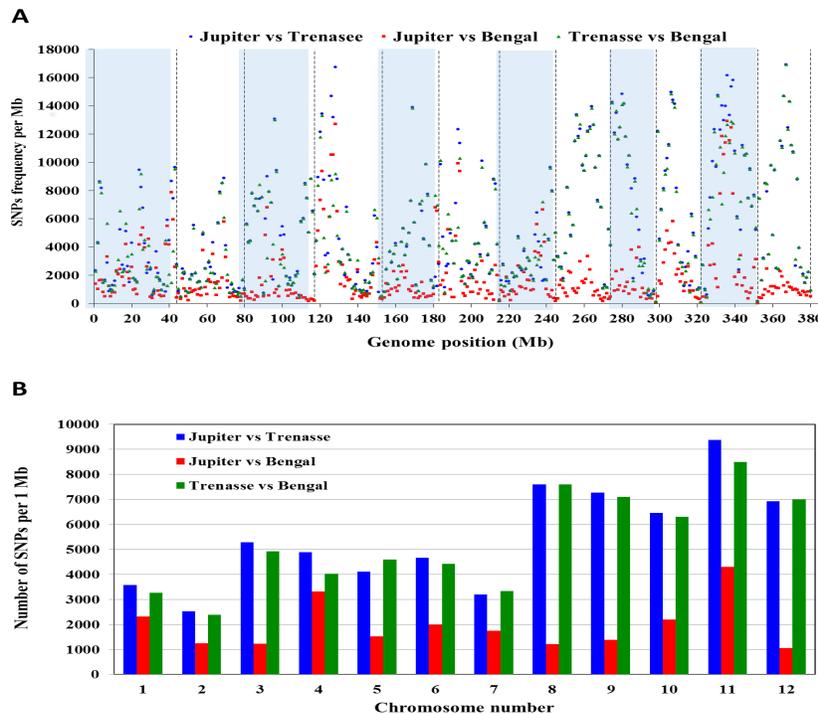

Figure 2: Frequency and distribution of SNPs in individual chromosomes when compared between Jupiter and Trenasse, Jupiter and Bengal, and

Trenasse and Bengal. A) Frequency of SNPs within 1-Mb window on individual chromosomes identified from pairwise comparisons between Jupiter and Trenasse, Jupiter and Bengal, and Trenasse and Bengal. Blue-colored circle, red-colored square, and green-colored triangle in the graph represent SNP frequency between Jupiter and Trenasse, Jupiter and Bengal, and Trenasse and Bengal, respectively. B) Average frequency of SNPs per 1 Mb in each chromosome. Bars with blue-colored, red-colored, and green-colored in the graph represent the density of SNPs between Jupiter and Trenasse, Jupiter and Bengal, and Trenasse and Bengal, respectively.

In all three pairwise comparisons, the highest density of SNPs was found on chromosome 11 (9,375, 4,312, and 8,501 per Mb, respectively), while the lowest density of SNPs was found on chromosome 2 in Jupiter/Trenasse and Trenasse/Bengal (2,525 and 2,381 per Mb, respectively), and on the chromosome 12 in Jupiter/Bengal (1,056 per Mb) (Table 1). Small insertions and deletions (InDels), ranged from 1 to 18 bp, were also analyzed with each comparison. Like SNPs, larger numbers of InDels were detected from the Jupiter/Trenasse and Bengal/Trenasse comparisons compared to the Jupiter/Bengal comparison (Figure 3A and 3B, Table 2 and 3).
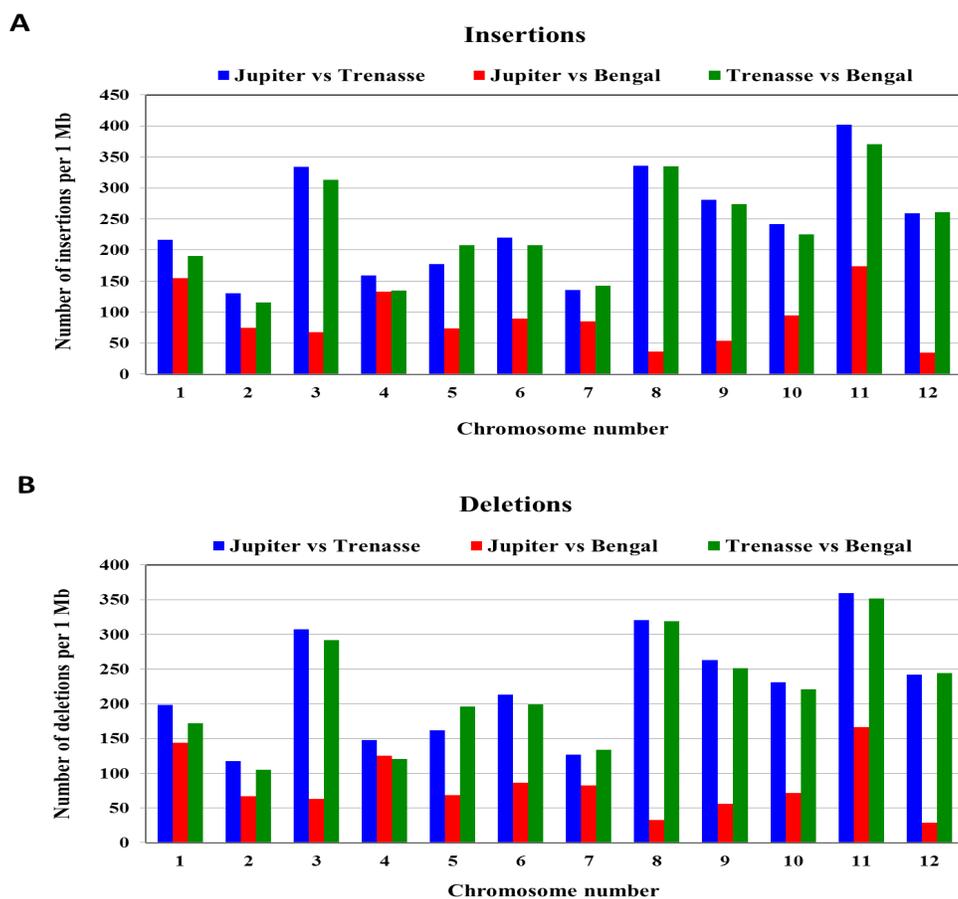

Figure 3: Average frequency of insertions (A) and deletions (B) per 1 Mb on individual chromosomes identified from pairwise comparisons between Jupiter and Trenasse, Jupiter and Bengal, and Trenasse and Bengal. Bars with blue-colored, red-colored, and green-colored in the graph represent the density of insertions or deletions between Jupiter and Trenasse, Jupiter and Bengal, and Trenasse and Bengal, respectively.

As shown in (Figure 4), number of InDels decreases exponentially in proportion to the sizes of InDels (Figure 4).
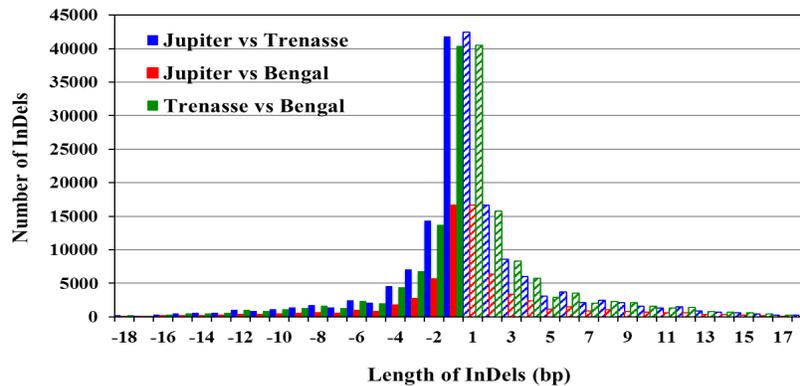

Figure 4: Distribution of insertions and deletions (InDels) variants based on their length in pairwise comparisons of three rice varieties, Jupiter, Trenasse and Bengal. The x-axis shows the length of deletions (colored-blocked bars) and insertions (colored-patterned bars). The y-axis shows the number of insertions and deletions. Blocked and patterned bars with blue-, red-, and green-color in the graph represent Jupiter vs Trenasse, Jupiter vs Bengal, and Trenasse vs Bengal, respectively.

The density of variations (SNPs and InDels) between the reference genome 'Nipponbare' and the three US varieties in this study was much higher than that between 'Nipponbare' and 'Omachi', another *temperate japonica* variety [28], which is likely due to the genomic portions of *tropical japonica* in the US varieties (Figure 1). In other studies, comparable levels of variations have been observed between 'Nipponbare' and *indica* varieties [29, 31], as well as between 'Nipponbare' and other elite *japonica* varieties including cold temperature-tolerant Hokkaido varieties [24,27,41]. In pairwise comparisons among the three US rice varieties, variants from the Jupiter/Bengal pair were much lower compared to those from the Jupiter/Trenasse or the Bengal/Trenasse pair (Figures 2 and 3, and Tables 1-3), which is also consistent with their genomic divergence revealed in this study (Figure 1).

Nucleotide Substitutions

SNPs can be classified into transitions (C/T and G/A) and transversions (C/G, T/A, A/C, and G/T). In this study, the frequency of transitions was higher than that of transversions in all three comparisons. Among transitions, little difference was found between the C/T substitution and the G/A substitution in all comparisons (Table 3).

For transversions (A/T, C/G, A/C, and G/T), however, the A/T substitution was most frequent with more than 60% higher numbers compared to the least frequent substitution, C/G, in all the comparisons (Table 3). Among the three comparisons, Jupiter/Bengal showed lower numbers of substitutions than Jupiter/

Trenasse and Trenasse/Bengal. The transitions (Ts)/Transversions (Tv) ratio was ≥2.5 in all cases (Table 3), which is higher than the previous study on rice [30,31]. The higher ratio was the result of higher transitions substitutions compared to transversions, indicating a 'transition bias.' It is suggested that transition bias occurs in natural selection because transition may conserve the protein structure better than transversions [42]. Transition bias has been previously reported in rice and chickpea [30,31,43]. Methylation causes higher frequency of C to T mutation, so higher C/T substitutions might have occurred compared to G/A [44]. Furthermore, A/T substitutions were abundance in transversions compared to other remaining substitutions C/G, A/C and G/T (Table 3), which is similar to the previous report on rice [30].

Positions of Variants in Different Regions of The Genome

Frequencies of SNPs and InDels in various genomic features, including intergenic region, upstream and downstream of gene models, UTR5' and 3', exon, and intron, were determined for all three comparisons (Table 4), using SnpEff v3.5e [36]. Those variants in the upstream of genes may have role in altering regulation of various downstream gene expression, which will ultimately alter phenotypic traits [45,46]. Variants in exons, especially those causing amino acid changes and frameshifts (i.e. non-synonymous SNPs and InDels causing frameshifts), may directly affect the functionality of the encoded protein. Regardless of the regions in the genome, highest number of variants was

detected between Jupiter and Trenasse, while lowest one was between Jupiter and Bengal (Table 4), which is congruent with other data shown in this study.

		Jupiter vs Trenasse	Jupiter vs Bengal	Trenasse vs Bengal
Upstream[a]		1,920,546	704,179	1,844,469
Intergenic[b]		1,347,790	494,130	1,284,046
Intron		393,156	143,872	373,427
UTRs[c]		62,961	21,934	61,441
Exon	(nsSNP)	231,390	92,272	221,997
	(sSNP)	208,179	86,709	205,009
	(Frameshift)	5,212	2,069	4,951
Intragenic[d]		203	68	203
Downstream[a]		1,837,883	679,579	1,770,798
[a]Upstream/Downstream: Variants in the region which is up to 5K base upstream/downstream of a gene.				
[b]Intergenic: Variants that occur in intergenic but not in upstream/downstream regions.				
[c]UTRs: Variants that hit 5′ and 3' untranslated region.				
[d]Intragenic: Variants that occur within a gene but fall outside of all transcript features.				

Table 4: Annotation of variants at various genomic regions identified after pairwise comparisons among three rice cultivars.

Synonymous and Non-Synonymous SNPs

51.6 - 52.6% of SNPs in the coding regions (CDS) were non-synonymous (nsSNPs), while 47.4 - 48.4% were synonymous SNPs (sSNPs) in all the three comparisons, resulting in the ratio of non-synonymous SNPs to synonymous SNPs (nsSNPs/sSNPs) to be around 1.06 - 1.11 (Table 5).

	Jupiter vs Trenasse			Jupiter vs Bengal			Trenasse vs Bengal		
	Number	%	nsSNPs/sSNPs	Number	%	nsSNPs/sSNPs	Number	%	nsSNPs/sSNPs
nsSNPs	231,390	52.6	1.11	92,272	51.6	1.06	221,997	52	1.08
sSNPs	208,179	47.4		86,709	48.4		205,009	48	

nsSNPs: non-synonymous single nucleotide polymorphisms
sSNPs: synonymous single nucleotide polymorphisms

Table 5: Synonymous and non-synonymous SNPs in coding sequences.

This value is similar to previous studies on *indica*, and *tropical* and *temperate japonica* rice [30,31,47], in which the nsSNPs/sSNPs ratios were around 1.2. It has been reported that the nsSNPs/sSNPs ratio tends to be lower in protein families with essential biological functions, such as cellulose synthases, but higher in the protein families with regulatory functions [38,47,48].

Conclusion

In this study, we conducted a genome-wide comparative analysis of the three US rice varieties with different quantitative resistances to BPB and SB, and detected SNP- and InDel-based polymorphisms among them. Regarding that whole genome sequence data have been considered as an excellent source for the development of reliable molecular markers [49], the information of genomewide sequence variations gained from this study is a useful resource to develop new molecular markers for future genetic studies and marker-assisted breeding of disease resistant rice using US rice varieties. Due to the close genetic relatedness among many US rice varieties, marker development relying on random screening has often been very inefficient and costly. The whole genome sequence information from this study and other similar studies with US rice accessions will greatly improve the efficiency in the development of new molecular markers, avoiding tedious screening processes with random candidate markers. In addition, this study will provide valuable information for functional and molecular studies of quantitative rice disease resistance.

References

1. Marchetti MA (1983) Potential impact of sheath blight on yield and milling quality of short-statured rice lines in the southern United-States. Plant Disease 67: 162-165.

2. Nandakumar R, Rush MC, Correa F (2007) Association of *Burkholderia glumae* and *B. gladioli* with panicle blight symptoms on rice in Panama. Plant Disease 91: 767.

3. Nandakumar R, Rush MC, Shahjahan AKM, O'Reilly KL, Groth DE (2005) Bacterial panicle blight of rice in the southern United States caused by *Burkholderia glumae* and *B. gladioli*. Phytopathology 95: S73.

4. Shahjahan AKM, Rush MC, Groth D, Clark CA (2000) Panicle blight. Rice Journal 15: 26-29.

5. Ham JH, Melanson RA, Rush MC (2011) *Burkholderia glumae*: next major pathogen of rice? Molecular Plant Pathology 12: 329-339.

6. Olaya G, Buitrago C, Pearsaul D, Sierotzki H, Tally A (2012) Detection of resistance to QoI fungicide in *Rhizoctonia solani* isolates from rice. Phytopathology 102: S488.

7. Olaya G, Sarmiento L, Edlebeck K, Buitrago C, Sierotzki H, Zaunbrecher J, et al. (2013) Azoxystrobin (QoI) resistance monitoring of *Rhizoctonia solani* isolates causing rice sheath blight in Louisiana. Phytopathology 103: S2106.

8. Jones JD and Dangl JL (2006) The plant immune system. Nature 444: 323-329.

9. Pinson SRM, Shahjahan AKM, Rush MC, Groth DE (2010) Bacterial panicle blight resistance QTLs in rice and their association with other disease resistance loci and heading date. Crop Science 50: 1287-1297.

10. Mizobuchi R, Sato H, Fukuoka S, Tanabata T, Tsushima S, et al. (2013) Mapping a quantitative trait locus for resistance to bacterial grain rot in rice. Rice 6: 13.

11. Mizobuchi R, Sato H, Fukuoka S, Tsushima S, Yano M (2015) Fine mapping of RBG2, a quantitative trait locus for resistance to *Burkholderia glumae* on rice chromosome 1. Molecular Breeding 35: 15.

12. Yadav S, Anuradha G, Kumar RR, Vemireddy LR, Sudhakar R, et al. (2015) Identification of QTLs and possible candidate genes conferring sheath blight resistance in rice (*Oryza sativa* L.). Springerplus 4: 175.

13. Wen ZH, Zeng YX, Ji ZJ, Yang CD (2015) Mapping quantitative trait loci for sheath blight disease resistance in Yangdao 4 rice. Genetics and Molecular Research 14: 1636-1649.

14. Tonnessen BW, Manosalva P, Lang JM, Baraoidan M, Bordeos A, et al. (2015) Rice phenylalanine ammonia-lyase gene OsPAL4 is associated with broad spectrum disease resistance. Plant Molecular Biology 87: 273-286.

15. Taguchi-Shiobara F, Ozaki H, Sato H, Maeda H, Kojima Y, et al. (2013) Mapping and validation of QTLs for rice sheath blight resistance. Breeding Science 63: 301-308.

16. Zuo S, Yin Y, Pan C, Chen Z, Zhang Y, et al. (2013) Fine mapping of qSB-11(LE), the QTL that confers partial resistance to rice sheath blight. Theoretical and Applied Genetics 126: 1257-1272.

17. Liu G, Jia Y, Correa-Victoria FJ, Prado GA, Yeater KM, et al. (2009) Mapping quantitative trait Loci responsible for resistance to sheath blight in rice. Phytopathology 99: 1078-1084.

18. Zuo S, Zhang L, Wang H, Yin Y, Zhang Y, et al. (2008) Prospect of the QTL-qSB-9Tq utilized in molecular breeding program of *japonica* rice against sheath blight. Journal of Genetics and Genomics 35: 499-505.

19. Li Z, Pinson SR, Marchetti MA, Stansel JW, Park WD (1995) Characterization of quantitative trait loci (QTLs) in cultivated rice contributing to field resistance to sheath blight (*Rhizoctonia solani*). Theoretical and Applied Genetics 91: 382-388.

20. Channamallikarjuna V, Sonah H, Prasad M, Rao GJN, Chand S, et al. (2010) Identification of major quantitative trait loci qSBR11-1 for sheath blight resistance in rice. Mol Breeding 25: 155-166.

21. Sha X, Linscombe SD, Groth D, Bond JA, White LM, et al. (2006) Registration of 'Jupiter' rice. Crop Science 46: 1811-1812.

22. Linscombe SD, Jodari F, Mckenzie KS, Bollich PK, White LM, et al. (1993) Registration of 'Bengal' rice. Crop Science 33: 645-646.

23. Linscombe SD, Sha X, Bond JA, Bearb K, Rush MC, et al. (2006) Registration of 'Trenasse' rice. Crop Science 46: 2318-2319.

24. Yamamoto T, Nagasaki H, Yonemaru J, Ebana K, Nakajima M, et al. (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11: 267.

25. Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology 27: 522-530.

26. Huang X, Feng Q, Qian Q, Zhao Q, Wang L, et al. (2009) High-throughput genotyping by whole-genome resequencing. Genome Research 19: 1068-1076.

27. Arai-Kichise Y, Shiwa Y, Ebana K, Shibata-Hatta M, Yoshikawa H, et al. (2014) Genome-wide DNA polymorphisms in seven rice cultivars of *temperate* and *tropical japonica* groups. PLoS ONE 9: e86312.

28. Arai-Kichise Y, Shiwa Y, Nagasaki H, Ebana K, Yoshikawa H, et al. (2011) Discovery of genome-wide DNA Polymorphisms in a Landrace cultivar of *japonica* rice by whole-genome sequencing. Plant Cell Physiology 52: 274-282.

29. Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, et al. (2004) An SNP resource for rice genetics and breeding based on subspecies *Indica* and *Japonica* genome alignments. Genome Research 14: 1812-1819.

30. Jain M, Moharana KC, Shankar R, Kumari R, Garg R (2014) Genome-ewide discovery of DNA polymorphisms in rice cultivars with contrast-

ing drought and salinity stress response and their functional relevance. Plant Biotechnology Journal 12: 253-264.

31. Subbaiyan GK, Waters DLE, Katiyar SK, Sadananda AR, Vaddadi S, et al. (2012) Genome-wide DNA polymorphisms in elite *indica* rice inbreds discovered by whole-genome sequencing. Plant Biotechnology Journal 10: 623-634.

32. Silva J, Scheffler B, Sanabria Y, De Guzman C, Galam D, et al. (2012) Identification of candidate genes in rice for resistance to sheath blight disease by whole genome sequencing. Theoretical and Applied Genetics 124: 63-74.

33. FastQC: a quality control tool for high throughput sequence data [Internet] (2010).

34. Langmead B and Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357-359.

35. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078-2079.

36. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6: 80-92.

37. Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: analytical and study design considerations. Genetic Epidemiology 28: 289-301.

38. Xu X, Liu X, Ge S, Jensen JD, Hu F, et al. (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nature Biotechnology 30: 105-111.

39. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. (2011) The variant call format and VCFtools. Bioinformatics 27: 2156-2158.

40. Lu H, Redus MA, Coburn JR, Rutger JN, McCouch SR, et al. (2005) Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. Crop Science 45: 66-76.

41. Takano S, Matsuda S, Kinoshita N, Shimoda N, Sato T, et al. (2014) Genome-wide single nucleotide polymorphisms and insertion-deletions of *Oryza sativa* L. subsp. *japonica* cultivars grown near the northern limit of rice cultivation. Molecular Breeding 34: 1007-1021.

42. Wakeley J (1996) The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. Trends in Ecology and Evolution 11: 158-162.

43. Agarwal G, Jhanwar S, Priya P, Singh VK, Saxena MS, et al. (2012) Comparative analysis of kabuli chickpea transcriptome with desi and wild chickpea provides a rich resource for development of functional markers. PLoS ONE 7: e52443.

44. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W (1978) Molecular basis of base substitution hotspots in *Escherichia coli*. Nature 274: 775-780.

45. Zhang X, Cal AJ, Borevitz JO (2011) Genetic architecture of regulatory variation in *Arabidopsis thaliana*. Genome Research 21: 725-733.

46. Thumma BR, Matheson BA, Zhang D, Meeske C, Meder R, et al. (2009) Identification of a Cis-acting regulatory polymorphism in a Eucalypt COBRA-like gene affecting cellulose content. Genetics 183: 1153-1164.

47. McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, et al. (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proceedings of the National Academy of Sciences of the United States of America 106: 12273-12278.

48. Zheng LY, Guo XS, He B, Sun LJ, Peng Y, et al. (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). Genome Biology 12: R114.

49. Imelfort M, Duran C, Batley J, Edwards D (2009) Discovering genetic polymorphisms in next-generation sequencing data. Plant Biotechnology Journal 7: 312-317.