GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Natural variation of rice blast resistance gene *Pi-d2*

J.B. Li[1]*, Y.D. Sun[2,3,4]*, H. Liu[5,6], Y.Y. Wang[2,3,4], Y.L. Jia[7] and M.H. Xu[2,3,4]

[1]Agricultural Environment and Resources Research Institute, Yunnan Academy of Agricultural Sciences, Kunming, Yunnan Province, China
[2]The Key Laboratory of Biotechnology Research of Yunnan Province, Yunnan Academy of Agricultural Sciences, Kunming, Yunnan Province, China
[3]Key Lab of Southwestern Crop Gene Resources and Germplasm Innovation of Ministry of Agriculture, Yunnan Academy of Agricultural Sciences, Kunming, Yunnan Province, China
[4]Institute of Biotechnology and Genetic Resource, Yunnan Academy of Agricultural Sciences, Kunming, Yunnan Province, China
[5]Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan Province, China
[6]Graduate University of Chinese Academy of Sciences, Beijing, China
[7]United States Department of Agriculture-Agricultural Research Service, Dale Bumpers National Rice Research Center, Stuttgart, Arkansas, USA

*These authors contributed equally to this study.
Corresponding authors: M. Xu / Y. Jia
E-mail: xuminhui@sohu.com / Yulin.Jia@ars.usda.gov

Genet. Mol. Res. 14 (1): 1235-1249 (2015)
Received November 25, 2013
Accepted March 20, 2014
Published February 13, 2015
DOI http://dx.doi.org/10.4238/2015.February.13.2

ABSTRACT. Studying natural variation in rice resistance genes of cultivated and wild rice relatives can predict resistance stability to rice blast fungus. In the present study, the protein coding regions of the rice *R* gene *Pi-d2* in 35 rice accessions, including *Oryza sativa* L. subsp. *indica* Kato (Aus), *indica* (IND), temperate *japonica* (TEJ), tropical *japonica* (TRJ), *aromatic* (ARO); subgroups of *Oryza sativa*; 6 accessions of wild rice varieties; *O. nivara*; and *O. rufipogon* were analyzed. A total of 13

nucleotide differences were found in the open reading frames (ORFs) of *Pi-d2*. Translation of these ORFs revealed 9 variants; 3 were novel *Pi-d2* variants. Variants H2 and H5 were identified in accessions of cultivated rice and *O. nivara,* H1, H3, H4, H6, and H8 were only identified in cultivated rice. H2 and H5 were the common types of IND and *O. nivara*, H8 was the common type of TRJ and AUS, H6 was the specific type of AUS, and H3 was the specific type of ARO. H7 and H9 were specific haplotypes of *O. nivara* and *O. rufipogon*, respectively. These findings demonstrate that *Pi-d2* variants are useful indicators for each subgroup, and *Pi-d2* is an ancient gene that predates speciation of rice subgroups.

Key words: Resistance gene; Natural variation; Blast disease; *Oryza sativa*; *Pi-d2*

INTRODUCTION

Rice blast, caused by the filamentous ascomycete fungus (*Magnaporthe oryzae*), is one of the most damaging diseases to rice worldwide. Genetic resistance has been defined as major and minor gene-mediated resistance. Major gene-mediated resistance (*R*) is effective in recognizing strains of the fungus that contain the corresponding avirulence (*AVR*) gene, and major *R* gene is complete and powerful; however, such resistance can be easily overcome by *M. oryzae* through mutation (Silue et al., 1992; Xing et al., 2013). Mutations of the *AVR* gene, *AVR-Pita1*, in historical and contemporary field isolates have been well documented in defeating major gene-mediated resistance (Zhou et al., 2007; Dai et al., 2010; Xing et al., 2013). Blast disease is currently managed by the use of resistant cultivars that carry both the major and minor *R* genes and the application of fungicides integrated into cultural practices. Among them, the use of resistant cultivars is the most economical and environmentally benign method of reducing crop loss due to blast. Searching and stacking *R* genes with different resistance specificities has been a priority for rice breeders and pathologists worldwide (Hao et al., 2009). To date, over hundreds of major and minor rice blast *R* genes have been identified, 19 of which (including *Pi-d2*) have been cloned (Ballini et al., 2008; Sharma et al., 2012).

Sequence analysis of natural alleles of cloned rice blast *R* genes in the rice germplasm is useful for understanding the evolution of *R* genes and predicting their origins (Okuyama et al., 2011). To date, extensive evolutionary analyses of *Pi-ta*, *Pik*, and *Pi2/9* have been well documented (Wang et al., 2008; Lee et al., 2009, 2011). As a result, DNA markers were also developed from portions of genes for marker-assisted selection (MAS), such as *Pi-ta* (Jia 2003; Jia et al., 2002, 2003, 2004), *Pi-b* (Fjellstrom et al., 2004), and *Pi-km* (Costanzo and Jia, 2010). These DNA markers have been effectively used for genetic studies and MAS worldwide (Hittalmani et al., 2000).

The *Pi-d2* gene found in the indica cultivar Digu confers resistance to *M. oryzae* strain ZB15 (Chen et al., 2010a,b). Digu-containing *Pi-d2* is particularly important for breeding resistance against rice blast disease in Southern China (Chen et al., 2004). Digu A, a carrier of *Pi-d2* was used as a cytoplast male sterile donor in 1995 (Lei et al., 1998), which resulted in the release of Fuyi A, thus forming the basis for the release of 16 hybrid rice cultivars in the 1990s. Ever since their release, they have been grown in Fujian, Hubei, Sichuan, Guangxi, and Hunan Provinces in China, and the estimated cropland use for these cultivars is 130,000 ha annually. Meanwhile, Fuyi A has been used as an important cytoplasmic male sterile parent

for improving rice blast resistance in 3 lines of hybrids in Southern China (Lei et al., 2004).

Pi-d2, a single copy gene within the rice genome, was cloned from a region near the centromere of chromosome 6 and encodes a putative serine-threonine receptor-like kinase (RLK) membrane-spanning protein with 825 amino acids (Chen et al., 2006). Because of its novel extracellular domain, *Pi-d2* represents a new class of plant *R* genes (Chen et al., 2006). A single amino acid at position 441 of the Pi-d2 protein distinguishes resistant and susceptible alleles. The amino acid methionine (M) at 441 was found in the susceptible *pi-d2* allele, and isoleucine (I) at 441 was found in the resistant allele.

The objectives of the present study were to determine variation in DNA and protein sequences of Pi-d2 and analyze the natural variation of *Pi-d2* in 41 cultivated and wild rice relatives worldwide.

MATERIAL AND METHODS

Plant materials

A total of 50 accessions of cultivated and wild rice were previously resequenced (Xu et al., 2011). From these 50 accessions, a total of 41 accessions were selected, which included 35 Asian cultivated rice and 6 wild rice accessions based on the quality of DNA sequences and their representations of the rice blast *R* genes (Table 1). The 35 Asian cultivated rice accessions included 10 tropical japonica (TRJ), 8 temperate japonica (TEJ), 5 aromatic (ARO), 4 aus (AUS), and 8 indica (IND). Among them, IR36 was the only cultivar developed by the International Rice Research Institute (IRRI, Philippines). The remaining germplasms were landraces, including 6 wild rice accessions (including 1 accession of *O. rufipogon*) and 5 accessions of *O. nivara*. The sequenced variety Nipponbare was used as a reference. Germplasm accessions for pathogenicity assays were generously provided by Harold Bockelman of the United States Department of Agriculture (USDA), Agricultural Research Service (ARS), National Small Grains Collection, Aberdeen, Idaho, USA.

Table 1. Origin and group of germplasm accessions used in this study[a].

Name of accessions	Plant identification	Status	Origin	Variety group	Name of accessions	Plant identification	Status	Origin	Variety group
Mehr	IRGC 12883	Landrace	Iran	AUS	Jambu	IRGC 17757	Landrace	Indonesia	TRJ
Kalamkati	IRGC 45975	Landrace	India	AUS	AZUCENA	IRGC 328	Landrace	Philippines	TRJ
Jhona 349	IRGC 6307	Landrace	India	AUS	NPE 844	IRGC 38698	Landrace	Pakistan	TRJ
DZ78	IRGC 8555	Landrace	Bangladesh	AUS	Arias	IRGC 43325	Landrace	Indonesia	TRJ
Leung Pratew	IRGC 27762	Landrace	Thailand	IND	Gotak Gatik	IRGC 43397	Landrace	Indonesia	TRJ
IR 36	IRGC 30416	Improve	Brazil	IND	Trembese	IRGC 43675	Landrace	Indonesia	TRJ
Popot 165	IRGC 43545	Landrace	Indonesia	IND	Canella De Ferro	IRGC 50448	Elite	Brazil	TRJ
Ai-Chiao-Hong	IRGC 51250	Landrace	China	IND	Lemont	IRGC 66756	Elite	USA	TRJ
Guan-Yin-Tsan	IRGC 51300	Landrace	China	IND	Davao	IRGC 8244	Landrace	Philippines	TRJ
Gie 57	IRGC 8231	Landrace	Vietnam	IND	Kitrana 508	IRGC 12793	Elite	Madagascar	ARO
TD2	IRGC 9148	Elite	Thailand	IND	Bico Branco	IRGC 38994	Elite	Brazil	ARO
JC91	IRGC 9177	Elite	India	IND	JC101	IRGC 9060	Elite	India	ARO
Ta Hung Ku	IRGC 1107	Landrace	China	TEJ	JC111	IRGC 9062	Elite	India	ARO
Haginomae Mochi	IRGC 2540	Elite	Japan	TEJ	Firooz	RA 4952	Landrace	Iran	ARO
Darmali	IRGC 27630	Landrace	Nepal	TEJ	042/87/34	IRGC 105327	Wild rice	India	Nivara
Phudugey	IRGC 32399	Landrace	Bhutan	TEJ	MV 89-80	IRGC 106105	Wild rice	India	Nivara
Norin 20	IRGC 418	Landrace	Japan	TEJ	L 89-12	IRGC 106154	Wild rice	Laos	Nivara
Chodongji	IRGC 55471	Landrace	Korea	TEJ	HK 47	IRGC 80470	Wild rice	India	Nivara
Mansaku	IRGC 8191	Landrace	Japan	TEJ	CA 97-053	IRGC 89215	Wild rice	Cambodia	Nivara
Nipponbare	NP	Elite	Japan	TEJ	PADI PADIAN	IRGC 105958	Wild rice	Indonesia	Rufipogon
Maintmolotsy	IRGC 11010	Elite	Madagascar	TRJ					

[a]These germplasm accessions were described in Xu et al. (2012).

Sequence analysis

The coding sequences of *Pi-d2* in Digu (GenBank accession No. FJ915121.1) and Nipponbare (accession No. NC008399) were obtained from http://blast.ncbi.nlm.nih.gov. *Pi-d2*^{Np} (NC008399) and *Pi-d2*^{Digu} (FJ915121.1) were 99% identical and identified from nucleotide positions 18040176-18037639 on rice chromosome 6. The target fragments of the *Pi-d2* alleles in the 41 rice accessions were obtained from the same location. Briefly, DNA sequences of chromosome 6 from the 41 rice accessions were opened by the Editplus v3.40 software (http://www.editplus.com/files.html), and nucleotide sequences of each rice accession from 18040176 to 18037639 were extracted and saved to a new file. DNA sequences of *Pi-d2* in 41 rice accessions were aligned using the DNASTAR v7.10 software (http://www.dnastar.com/). The diversity index was calculated as the frequency of haplotypes or protein types in the accession populations following Fontaine's method (Fontaine et al., 2004) by using the formula:

$$\text{Diversity index} = (1 - \sum^{n}{}_{i=1} p_i^2)$$

where p_i is the frequency of the haplotype *i* in a population. A phylogenetic tree was constructed using the DNASTAR v7.10 software by the ClustalW method. Tree view was constructed using the MEGA 4.0 software (Tamura et al., 2007).

Nucleotide polymorphism analysis

Nucleotide polymorphisms were calculated according to Rozas's methods by the DnaSP v5.10.01 software (Rozas et al., 2003). The aligned DNA sequences were imported into the DnaSP v5.10.01 program to calculate *S* (number of polymorphic or segregating sites), π (nucleotide diversity), θ (Theta from S, Theta-W), and *D* (Tajima's *D*) (Tajima et al., 1989); then, we drew the sliding window of nucleotide diversity (π).

Positive selection analysis

The positive selection analysis was performed using the Selection Server program downloaded from the website http://selecton.tau.ac.il/. Five calculation models were used to identify the positive selection sites under the query of Pid2 as follows: M8 (positive selection enabled, beta + w > = 1) (Yang et al., 2000), M8a (beta + w = 1, null model) (Swanson et al., 2003), M7 (beta, null model) (Yang et al., 2000), M5 (positive selection enabled, gamma) (Yang et al., 2000), and the mechanistic empirical combination (MEC) model (Doron-Faigenboim et al., 2006). The data were then imported into Microsoft Excel for statistical analysis and to draw the sliding window.

Pathogenicity assay

Pathogenicity assays were performed on 25 available rice accessions at the USDA ARS Dale Bumpers National Rice Research Center, Stuttgart, Arkansas, USA. *M. oryzae* isolates, one avirulent (AVR) race IB1 (ARB146), one virulent (VIR) race, IEIK (TM2) for rice cultivars carrying the rice blast *R* gene *Pi-ta* (Xing et al., 2013) were used for pathogenicity tests. There were 4 replicates for each germplasm. The reactions to rice blast were determined

using a modified standard pathogenicity assay as previously described by Jia et al. (2003). Specifically, rice seedlings at the 3rd-4th leaf stages in a plastic bag were spray-inoculated with a rice blast spore suspension at 1-5 x 10^5 spores/mL. After inoculation, the plastic bags were sealed to maintain high humidity for 24 h; subsequently, the plants were removed from the bags. Then, the plants were maintained in a greenhouse for an additional 6 days to allow for development of notable disease symptoms. The reactions were rated as 0-2 for resistant and 3-5 for susceptible based on the visual number of lesions at the second youngest leaf.

RESULTS

Natural variation in the *Pi-d2* allele

The length of the open reading frame (ORF) of *Pi-d2* was found to be 2478 bp in all rice accessions. Nucleotide variations were found at positions 962, 1323, 1997, 1998, 2172, and 2463 between the resistant *Pi-d2*^{Digu} and susceptible *pi-d2*^{Np}, and a total of 13 nucleotide variations were found among the 41 rice accessions (Table 2). The variation frequencies at positions 1997, 1998, and 2172 in the 41 accessions were higher than those of the other regions. Nucleotides at these positions were different than those of *Pi-d2*^{Digu} in most accessions. Among the 13 different positions, the same nucleotides at 495, 1998, and 2172 were detected in the cultivated rice and 2 species of wild rice, whereas 347 and 1997 were detected in the cultivated rice and *O. nivara*. However, the same nucleotide at position 1497 was detected only in *O. rufipogon*. These findings suggest that there are significant differences in the sequences of *Pi-d2* between *O. rufipogon* and the cultivated rice.

Table 2. Sequence variation loci of *Pi-d2* ORF in different rice groups[b].

Variant locus	No. of sequenced accessions	Bases in *Pi-d2* gene	Base in Nipponbare	Substituted bases	No. of accessions	Frequency (%)	Number						
							ARO	AUS	TEJ	TRJ	IND	Nivara	Rufipogon
20	41	G	-	A	1	2.4	0	0	0	0	0	1	0
66	41	A	-	T	5	12.2	5	0	0	0	0	0	0
347	41	T	-	A	3	7.3	0	0	0	0	2	1	0
495	41	G	-	A	10	24.4	0	0	1	2	3	2	1
962	41	C	T	T	15	36.6	0	2	8	5	0	0	0
1107	41	G	-	A	5	12.2	5	0	0	0	0	0	0
1323	41	A	G	G	15	36.6	0	2	8	5	0	0	0
1497	41	G	-	A	1	2.4	0	0	0	0	0	0	1
1719	41	C	-	T	1	2.4	0	0	0	1	0	0	0
1997	41	A	G	G	41	100	5	4	8	10	8	5	0
1998	41	T	C	C	33	80.5	5	4	8	10	3	2	1
2172	41	T	C	C	33	80.5	5	4	8	10	3	2	1
2463	41	C	T	T	15	36.6	0	2	8	5	0	0	0

[b]GenBank accession numbers for these germplasm were KB738450-738490 that will be provided upon acceptance of the manuscript.

Nucleotide variation at position 20 was only detected in one accession of *O. nivara*, and the same nucleotide at positions 66, 962, 1107, 1323, and 1719 was detected only in the cultivated rice. There were 1-7 nucleotide substitutions among the 35 cultivated rice accessions (Table 3). Among them, the same nucleotide was identified only at position 1719 in one accession of TRJ, whereas the same nucleotide at position 66 and 1107 was only identified in one

accession of ARO (Table 2). These results suggest that DNA sequence variation for *Pi-d2* can be used to distinguish between cultivated and wild rice, and among cultivated rice accessions.

Table 3. Haplotypes of *Pi-d2* ORF in different accessions.

Haplotypes	Varied locus in ORF of *Pi-d2*													No. of variant bases	Accessions
	20	66	347	495	962	1107	1323	1497	1719	1997	1998	2172	2463		
Pi-d2^{Digu}	G	A	T	G	C	G	A	G	C	A	T	T	C		Digu
Pi-d2^{Np}	-[b]	-	-	-	T	-	G	-	-	G	C	C	T	6	Nipponbare
H1	-	-	-	-	T	-	G	-	-	G	C	C	T	6	IRGC12883, IRGC8555, IRGC1107, IRGC2540, IRGC27630, IRGC32399, IRGC418, IRGC55471, IRGC8191, NP, IRGC11010, IRGC38698, IRGC66756, IRGC8244
H2	-	-	-	-	-	-	-	-	-	G	-	-	-	1	IRGC27762, IRGC43545, IRGC51300, IRGC9148, IRGC9177, IRGC105327, IRGC106154, IRGC89215
H3	-	T	-	-	-	A	-	-	-	G	C	C	-	5	IRGC12793, IRGC38994, IRGC9060, IRGC9062, RA 4952
H4	-	-	-	A	-	-	-	-	-	G	C	C	-	4	IRGC51250, IRGC17757, IRGC43325, IRGC43675, IRGC50448
H5	-	-	A	A	-	-	-	-	-	G	C	C	-	5	IRGC30416, IRGC8231, IRGC80470
H6	-	-	-	-	T	-	G	-	T	G	C	C	T	7	IRGC328
H7	A	-	-	A	-	-	-	-	-	G	C	C	-	5	IRGC106105
H8	-	-	-	-	-	-	-	-	-	G	C	C	-	3	IRGC43397, IRGC45975, IRGC6307
H9	-	-	-	A	-	-	-	A	-	G	C	C	-	5	IRGC105958

[b]Indicates the same nucleotide as that of the *Pi-d2* gene from Digu.

The nucleotide diversity of *Pi-d2* ORFs was examined among the 41 sequence assemblies using the DnaSP program (Table 4). A total of 10 and 6 polymorphic sites were observed in the cultivated and wild rice accessions, respectively (Table 4). In the cultivated rice, the value of the nucleotide diversity (π) was 0.00122, which was higher than that of the wild rice group (π = 0.00113), suggesting higher diversity in the cultivated rice group (Table 4). Furthermore, the wild rice *O. nivara* showed much higher diversity than that of the cultivated rice subgroups (i.e., AUS, TEJ, TRJ, IND, and ARO), suggesting a slightly higher diversity in *O. nivara* (Table 4). Interestingly, the nucleotide diversity was extremely low for ORFs in TEJ and ARO; the values of *S*, π, and θ were zero (Table 4). In *O. sativa*, TRJ had the lowest diversity (π = 0.00095 and θ = 0.00071) in the ORFs (Table 4), whereas similar values of π and θ for ORFs were found in AUS and IND. These results suggest that the *Pi-d2* alleles of *O. nivara* were subjected to positive selection. The *O. sativa* subgroups with low values of π and θ (Table 4) suggest that AUS, TRJ, TEJ, IND, and ARO had been subjected to purified selection.

In order to examine the evolutionary dynamics of the *Pi-d2* alleles in the cultivated and wild rice groups, we evaluated neutral selection with Tajima's *D* test regardless of demographic consideration. Whereas, Tajima's *D* in all rice accessions was not statistically significant, a positive Tajima's *D* (0.40276) was observed among all accessions (Table 4), suggesting that the *Pi-d2* alleles may have experienced purified selection. The Selection Server was used to identify the positive-selection sites in the *Pi-d2* alleles, and we calculated the value of Ka/Ks (Ka = the rate of nonsynonymous substitutions, Ks = the rate of synonymous substitutions) for each amino acid of the deduced Pi-d2-like protein sequences from the 41 accessions. Five models, including 3 positive-selection models, M8 (Yang et al., 2000), M5 (Yang et al., 2000), and MEC; and 2 null models, M7 (Yang et al., 2000), and M8a (Swanson et al., 2003), were

tested using the Selection Server program (http://selecton.tau.ac.il/). In comparison to the null model M8a, the likelihood ratio test (LRT) of the M8 model was not significant in all rice groups (Tables 5 and 6). The results of 3 calculated positive-selection models were inconsistent (Tables 5 and 6). The sliding windows were drawn to show the distribution of the Ka/Ks values across all 825 amino acids under the M5, M8, and MEC models (Figure 1). The results showed that the Ka/Ks values of most sites (~90%) were <1, suggesting that these sites were potentially subjected to purifying selection. Few positive-selection sites (Ka/Ks > 1) with statistically significant results were found in the ORFs.

Table 4. Nucleotide polymorphism of the *Pi-d2* alleles[f].

Group	Subgroup	Component	Location (nt)	*S*	π	θ	Tajima's *D*[j]
All		Total	1-2,478	12	0.00128	0.00113	0.40276
Cultivated rice		Total	1-2,478	10	0.00122	0.00098	0.74480
	AUS	Total	1-2,478	3	0.00081	0.00066	2.01187
	TEJ	Total	1-2,478	0	0	0	0
	TRJ	Total	1-2,478	5	0.00095	0.00071	1.43514
	IND	Total	1-2,478	4	0.00082	0.00062	1.39420
	ARO	Total	1-2,478	0	0	0	0
Wild rice		Total	1-2,478	6	0.00113	0.00106	0.37522
	O. nivara	Total	1-2,478	5	0.00105	0.00097	0.56199

[f]S number of polymorphic or segregating sites; π nucleotide diversity; θWatterson's nucleotide diversity estimator based on silent site. [j]Tajima's *D* statistic (Tajima et al., 1989) based on the differences between the number of segregating sites and the average number of nucleotide differences.

Table 5. Positive-selection sites of the *Pi-d2* alleles in different rice groups as indicated by the M8, MEC, and M5 models.

Group	Model	Significance level	Total positive sites
All	M8	NON[g]	0
	M5	-	0
	MEC	AIC > M8a[h]	4
Cultivated rice	M8	NON	0
	M5	-	0
	MEC	AIC > M8a	4
Wild rice	M8	NON	0
	M5	-	0
	MEC	AIC < M8a	2

The positive-selection sites with a statistical significance (the lower bound of confidence interval > 1) under Bayesian test was used for calculation here. [g]Indicates likelihood ratio test between the two models (M8 and M8a) shows non-significant. [h]If the score of AIC under MEC model is less than the value of M8a, it is statistically significant.

Distribution of *Pi-d2* haplotypes in rice accession

Half of the accessions in AUS were H1 and H8; 62.5, 25.0, and 12.5% of accessions in IND were H2, H5, and H4, respectively. All of the accessions of ARO and TEJ were H3, and 40.0% of the accessions in TRJ were H1 and H4. Four haplotypes were detected in 10 accessions of TRJ. The diversity index of haplotypes was calculated as 0.66; TRJ possessed the highest diversity. Similarly, the diversity indices were 0.56, 0.53, and 0.50, for *O. nivara*, IND, and AUS, respectively. The lowest diversity index was 0.00, which was found in TEJ, ARO, and *O. rufipogon* (Table 7). In summary, the diversity indices of *Pi-d2* can be ordered as follows: TRJ > *O. nivara* > IND > AUS > *O. rufipogon*, ARO, and TEJ.

Table 6. Nucleotide polymorphism of the *Pi-d2* alleles in different rice subgroups[g].

Group	Subgroup	Model	Significance level	Total positive sites
Cultivated rice	AUS	M8	NON[g]	2
		M5	-	2
		MEC	AIC > M8a[h]	0
	TEJ	M8	NON	0
		M5	-	0
		MEC	AIC > M8a	825
	TRJ	M8	non	0
		M5	-	0
		MEC	AIC < M8a	2
	IND	M8	NON	0
		M5	-	0
		MEC	AIC < M8a	1
	ARO	M8	NON	0
		M5	-	0
		MEC	AIC > M8a	825
Wild rice	*O. nivara*	M8	NON	0
		M5	-	0
		MEC	AIC < M8a	2

The positive-selection sites with a statistical significance (the lower bound of confidence interval >1) under Bayesian test was used for calculation here. [g] Indicates likelihood ratio test between the two models (M8 and M8a). [h]If the score of AIC under MEC model is less than the value of M8a, it is statistically significant.

Functional amino acid of the Pi-d2 protein

Alignment of DNA sequence assemblies revealed 9 *Pi-d2* haplotypes (H) (Table 3). A previous study by others demonstrated that the reference cultivar Nipponbare, belonging to H1, was the susceptible *pi-d2* haplotype (Chen et al., 2006), whereas the remaining 8 *pi-d2* haplotypes were different than those of *Pi-d2*^{Digu} and *Pi-d2*^{Np}. H1 was found in 14 rice accessions (Table 7), with the highest frequency of haplotype variation (34.1%) among the 9 haplotypes (Table 7). There were 6 nucleotide substitutions from *Pi-d2*^{Np} compared to those for *Pi-d2*^{Digu}. The second highest frequency in nucleotide variation (19.5%) was H2, which was found in 8 rice accessions. There was only one nucleotide variation when *Pi-d2*^{Np} was compared to *Pi-d2*^{Digu}. The third highest frequency of nucleotide variation (12.2%) was found for H3 and H4 in 5 rice accessions. There were 5 and 4 nucleotide substitutions in *Pi-d2* H3 and *Pi-d2* H4, respectively, when compared to that of *Pi-d2*^{Digu}. The frequencies of nucleotide variation in the remaining haplotypes were <10.0%, with 3-5 nucleotide substitutions. Among the 9 haplotypes, H2 and H5 were found in rice cultivated and *O. nivara*; the H1, H3, H4, H6, and H8 haplotypes were found only in cultivated rice; H2 and H5 were the common types of IND and *O. nivara*. H6 was the specific type of TRJ, H3 was the specific type of ARO, H4 was the common type of IND and TRJ, and H8 was the common type of AUS and TRJ. H7 and H9 were specific haplotypes of *O. nivara* and *O. rufipogon*, respectively (Table 7). These results demonstrate that some rice cultivated and *O. nivara* share a few common *Pi-d2* haplotypes, whereas the rice cultivated and *O. nivara* share no common haplotypes with *O. rufipogon*. These findings suggest that genetic differentiation of *Pi-d2* between rice cultivated and *O. nivara* was lower in comparison to that of *O. rufipogon*.

Translation of the ORFs of *Pi-d2* haplotypes in 41 accessions revealed similar proteins with a putative serine-threonine RLK membrane-spanning protein of 825 amino acids. Pi-d2 also contained a predicted extracellular bulb-type mannose-specific lectin (B-lectin) binding domain, which has not been reported in other plant R proteins. The N-terminus of Pi-d2, amino

acids 1-32, contains a hydrophobic region with a predicted transit peptide function (Wasano et al., 2003). Additionally, amino acids 337-418 are predicted to encode a weak PAN domain (smart e-02) that binds proteins or carbohydrates (Chen et al., 2006). The core of the PAN domain in the region of amino acids 337-403 is possibly connected to the formation of 3 disulphide bridges (Chen et al., 2006). Moreover, only the 441st amino acid of Pi-d2 was reported as the functional locus for resistance or susceptibility. Among the amino acids of the Pi-d2 protein in the 41 accessions, there were 5 variant amino acids (i.e., R7H, F116Y, S321L, I441M, and H666R) when all Pi-d2 histidine (H) were compared to those at the same position in Pi-d2^{Digu}. These variant amino acids were located in the functional domain of the N-terminal hydrophobic region, B-lectin of the extracellular domain, and serine-threonine kinase catalytic domain.

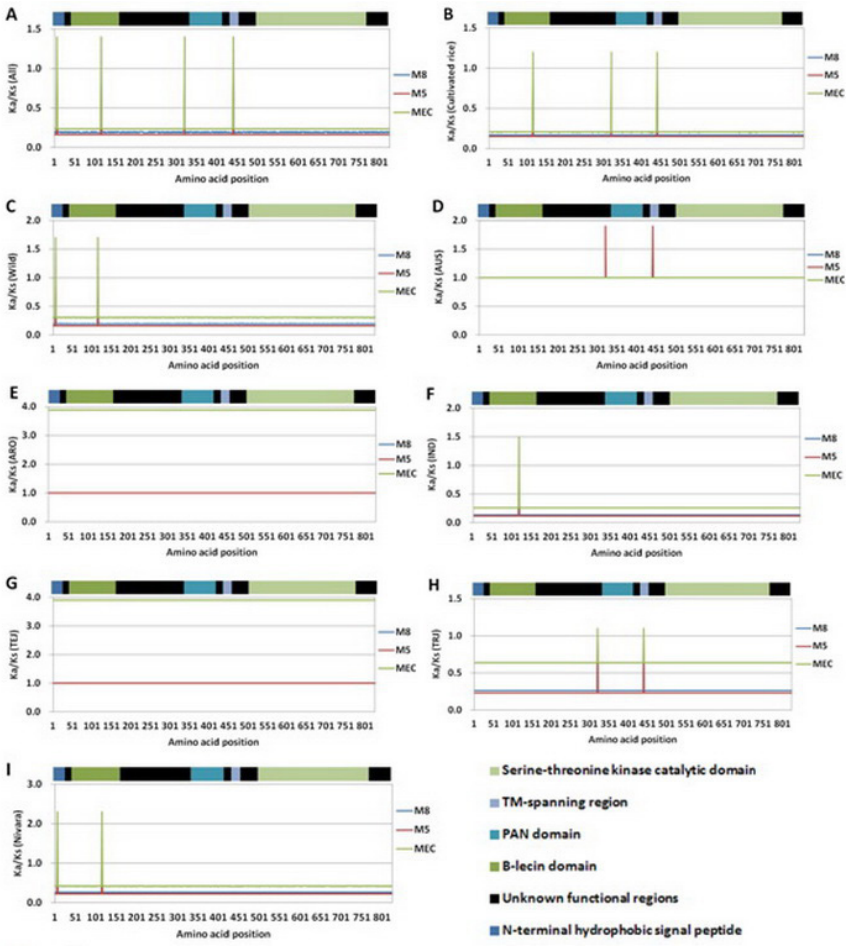

Figure 1. Sliding window of positive-selection sites of the *Pi-d2* alleles under M8, M5, and MEC models. The Y-axis indicates the ratio of the rate of nonsynonmous substitution (Ka) to the rate of synonymous substitution (Ks) (Ka/Ks); the X-axis indicates the position of the Pi-d2 amino acids in the site. The graphic description of protein with predicted domains was above each sliding window. Different predicted domains were shaded with different color at the right end of figure. The sliding window showing the total rice accessions (**A**), cultivated rice (**B**), wild rice species (**C**), AUS (**D**), ARO (E), IND (**F**), TEJ (**G**), TRJ (**H**) and wild rice *O. nivara* (**I**).

Table 7. Distribution *Pi-d2* haplotype in different rice groups.

Haplotypes	No. of accessions	Percent (%)	Cultivated rice and ecological type and frequency (%)					Wild rice and frequency (%)	
			AUS	IND	TEJ	TRJ	ARO	*O. nivara*	*O. rufipogon*
H1	14	34.1	2 (50.0)[c]	0	8 (100)	4 (40.0)	0	0	0
H2	8	19.5	0	5 (62.5)	0	0	0	3 (60.0%)	0
H3	5	12.2	0	0	0	0	5 (100)	0	0
H4	5	12.2	0	1 (12.5)	0	4 (40.0)	0	0	0
H5	3	7.3	0	2 (25.0)	0	0	0	1 (20.0%)	0
H6	1	2.4	0	0	0	1 (10.0)	0	0	0
H7	1	2.4	0	0	0	0	0	1 (20.0%)	0
H8	3	7.3	2 (50.0)	0	0	1 (10.0)	0	0	0
H9	1	2.4	0	0	0	0	0	0	1 (100%)
Total	41	100.0	4	8	8	10	5	5	1
No. of haplotypes			2	3	1	4	1	3	1
Index of diversity[d]			0.50	0.53	0.00	0.66	0.00	0.56	0.00

[c]Number and frequency (in bracket) of rice accessions of each haplotype. [d]Diversity index was calculated as the frequency of haplotypes types in the accessions population following Fontaine's method (Fontaine et al., 2004). Diversity index = $(1-\sum^{n}_{i=1}p_i^2)$ (where pi is the frequency of the haplotype i in a population).

Three amino acids, S321L, I441M, and H666R, of the Pi-d2 protein differed between *Pi-d2*^{Digu} and *Pi-d2*^{Np}. A single amino acid I at position 441 of the Pi-d2 protein determines resistant specificity, and M was found in the susceptible *pi-d2* allele in Nipponbare (Chen et al., 2006). The amino acids at this position in 9 haplotypes are summarized in Table 8. Amino acid M at position 441 was found in H1 and H6, suggesting that H1 and H6 are susceptible alleles; whereas, the remaining 7 haplotypes were identified as I at position 441, suggesting they are resistance alleles. The functions of other amino acids among all Pi-d2 haplotypes are unknown. Interestingly, the amino acids of all the tested materials at position 666 were R (arginine), including Nipponbare. The I at position 441 was found in all wild rice accessions, suggesting that the resistance *Pi-d2* allele was domesticated from wild rice relatives.

Table 8. Variation of the Pi-d2 protein in different accessions.

Haplotypes	Variation locus[e]				
	7	116	321	441	666
Digu	R	F	S	I	H
Nipponbare	-	-	L	M	R
H1	-	-	L	M	R
H2	-	-	L	-	R
H3	-	-	-	-	R
H4	-	-	-	-	R
H5	-	Y	-	-	R
H6	-	-	L	M	R
H7	H	-	-	-	R
H8	-	-	-	-	R
H9	-	-	-	-	R

[e]Amino acid as that of *Pi-d2* in Digu. R, H, F, Y, S, L, I, and M = Arginine, Histidine, Phenylalanine, Tyrosine, Serine, Leucine, Isoleucine, and Methionine, respectively.

Pathogenicity assays

H1 and H6 contained M at the 441st amino acids of the Pi-d2 protein (Table 8), which were predicted to be the susceptible *pi-d2* alleles. From pathogenicity assays, rice accession

Nipponbare was determined to be susceptible to the races IB1 and IEIK of *M. oryzae,* suggesting that Nipponbare has a susceptible *pi-d2* allele (Table 9). As shown in Table 9, reactions of the resistant haplotypes H1, H3, H4, and H5 did not correlate with the functional nucleotide of their respective structural variations.

Table 9. Summary of disease reaction of available rice accessions with selected *Pi-d2* haplotypes[X].

Name of accession	Pi-d2 haplotype	Disease reaction	
		IB1 (ARB146)[Y]	IEIK (TM2)
IRGC30416(IR36)	H5	R 0	R 0
Nipponbare	H1	S 4	S 5
IRGC 1107	H1	S 5	S 5
IRGC 418	H1	S 3	S 5
IRGC 8244	H1	S 5	S 5
IRGC 328	H6	S 5	S 5
IRGC 8555	H1	S 5	S 5
IRGC 6307	H8	R 1	S 3
IRGC 27630	H1	S 3	S 5
IRGC 32399	H1	R 0	S 4
IRGC 38698	H1	S 5	S 5
RA 4952	H3	R 0	S 3
IRGC 43325	H4	S 4	S 4
IRGC 43397	H8	R 0	S 4
IRGC 43675	H4	S 5	R 2
IRGC 45975	H8	R 2	R 0
IRGC 50448	H4	R 2	R 1
IRGC 51250	H4	R 1	R 0
IRGC 51300	H2	R 1	S 3
IRGC 11010	H1	R 0	S 3
IRGC 38994	H3	R 0	S 5
IRGC 17757	H4	R 0	S 3
IRGC 12793	H3	R 0	S 3
IRGC 8191	H1	S 4	R 0
IRGC 12883	H1	R 0	S 4

[X]Only those accessions with seeds available from USDA were used for disease evaluation. Disease was evaluated using a 0-5 scale where 0-2 indicates resistance (R) and 3-5 indicates susceptibility (S), respectively. [Y]Field isolate ARB146 in 2009 from Arkansas described by Xing et al. (2013).

Genetic structure and relationship of *Pi-d2* in accessions

To examine the population structure and relationships of *Pi-d2* in 41 rice accessions, we constructed a phylogenetic tree based on the nucleic acid sequences of the *Pi-d2* alleles (Figure 2). The rice accessions were divided into 2 clusters, 441I and 441M. The 441I cluster containing Digu, *O. nivara*, and *O. rufipogon* showed longer branches than that of the 441M cluster. However, the indica, japonica, *O. nivara*, and *O. rufipogon* groups were not separated, with IND being more dispersed and, thus, indicating higher diversity (Figure 2). The ARO was a specific cluster of *Pi-d2*. The tree for *Pi-d2* showed the star-like branches in IND, *O. nivara*, and *O. rufipogon*, with long branches. These results suggest that the *Pi-d2* resistance alleles may have originated from *O. nivara*.

DISCUSSION

Most cloned *R* genes belong to the nucleotide binding site (NBS) and leucine-rich repeat (LRR) class of proteins and contain cytoplasmic proteins with a predicted leucine zip-

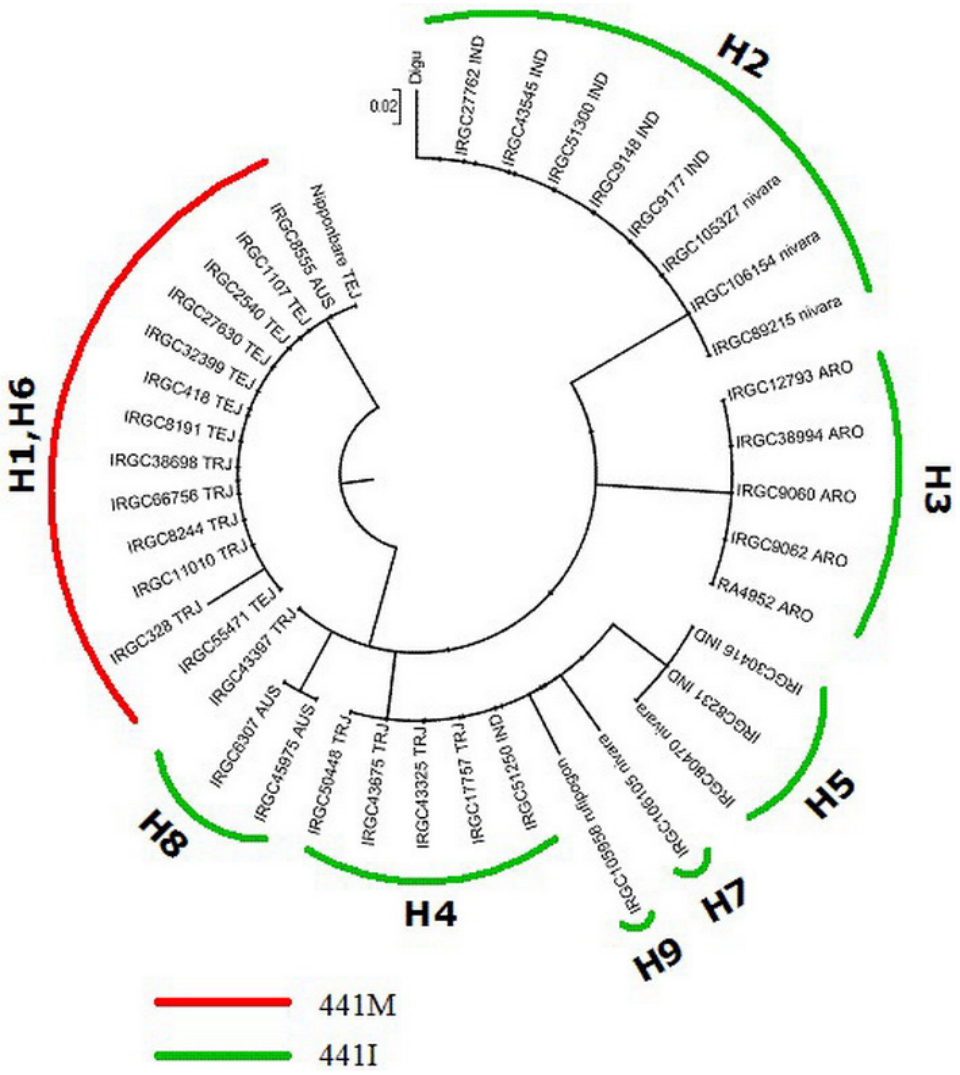

Figure 2. Phylogenetic tree constructed with the nucleotide sequences of *Pi-d2* genes from 41 rice accessions ClustalW methods of DNASTAR V7.10. Digu carries *Pi-d2* (Chen et al., 2006); 441M and 441I indicate the 441th amino acid were M and I of Pi-d2 proteins in rice accession, respectively; H1 to H9 indicate the haplotypes of Pi-d2.

per, NBS, and C-terminal LRR motifs. The *Pi-d2* gene encodes a serine-threonine-RLK membrane-spanning protein. Because of its novel extracellular domain, *Pi-d2* is a new class of rice *R* genes. Similar to the NBS-type rice blast *R* gene *Pi-ta*, *Pi-d2* is also a single-copy gene within the rice genome. *Pi-d2* encodes a predicted protein of 825 amino acids with the domain characteristics of RLK, an extracellular domain, a transmembrane domain, and an intracellular kinase domain (Chen et al., 2006).

Studying the natural variation of *Pi-d2* will help to understand the origin, evolution, and resistance stability of the rice *R* gene. In the present study, we reported structural varia-

tion of *Pi-d2* in 41 cultivated and wild rice relatives. We found 13 SNPs, 11 of which are new, in the ORFs of *Pi-d2* (Table 2). Some of the SNPs were specific to certain rice subgroups. A total of 9 haplotypes were identified based on 13 SNPs in 41 rice accessions. However, the resistant *Pi-d2* allele in Digu and the 2 susceptible *pi-d2* alleles in cultivars LTH and TP309 were identical, except for 2 base changes at nucleotide positions 963 and 1323, respectively (Chen et al., 2006). These 2 base substitutions resulted in amino acid substitutions, S321L and I441M, in the susceptible *pi-d2* alleles. Translation of 9 haplotypes of *Pi-d2* revealed 5 novel Pi-d2 variants. Based on the functional amino acid, 5 new putative resistant haplotypes from cultivated and wild rice relatives were identified. Five variant amino acids of Pi-d2, including R7H, F116Y, S321L, I441M, and H666R, were found in 41 rice accessions when compared to Digu. Among them, the R7H and F116Y variant types were not reported in a previous study by Chen et al. (2006). The amino acids of all materials tested at position 666 were R (Arginine) and consistent with that of the *Pi-d2* allele in Nipponbare; at the same position, H (Histidine) was present in Digu, suggesting that H666 is another functional amino acid for an unknown biological function that remains to be investigated.

An amino acid substitution in a critical functional domain often results in a conformational change of the protein that may impact the expected biological function. In the present study, we identified 9 amino acid variants of Pi-d2 and 5 novel haplotypes in 41 rice accessions (Table 8). Three amino acid variants, R7H, F116Y, and H666R, of the Pi-d2 protein differed from those of *Pi-d2*^{Digu}. These amino acids are located in the hydrophobic region of the N-terminus, B-lectin of the extracellular domain, and serine-threonine kinase catalytic domain. These 3 amino acid variations may affect conformation of these functional domains, suggesting that there is the potential for instability of *Pi-d2*-mediated rice blast resistance. In contrast, only one resistant haplotype, a single amino acid alanine at position 918 determining *Pi-ta* resistance specificity, was found in 159 rice accessions of *Oryza* species, including AA genome species *Oryza sativa*, *O. glaberrima, O. rufipogon, O. nivara*, and *O. barthii*, and CC genome species *O. officinalis* (Wang et al., 2008). In this study, additional amino acid variants of Pi-d2 were found compared with that of Pi-ta, suggesting that confirmation of Pi-d2 is influenced by more than just a single amino that was found in Pi-ta.

There was no correlation between the race-specific reaction and structural variation of Pi-d2, suggesting that other amino acids may influence reactions. This was not unexpected because the isolates/races IB1 (ARB14) and IE1k (TM2) of *M. oryzae* were used to determine the spectra of resistance mediated by the blast *R* gene, *Pi-ta* (Xing et al., 2013). This analysis supported a previously known theory that there are lineage exclusions in field blast populations, where *M. oryzae* strains with similar *AVR* genes have been asexually inherited. Nevertheless, these germplasm accessions may contain *Pi-ta* that may mask their predicted reactions. Analysis of *Pi-ta* haplotypes in these germplasm accessions should help resolve this prediction. It is equally possible that other *R/AVR* gene pairs in these germplasms can also mask expected reactions. Likewise, screening for the differential *M. oryzae* race of *Pi-d2* will provide additional insight into the structural and functional relationships of the Pi-d2 haplotypes.

The functional amino acid at the 441st position of Pi-d2 separated the cultivated accessions via 441M, and cultivated and wild rice relatives via 441I (Table 1; Figure 2). Despite not separating *indica*, *japonica*, *O. nivara* and *O. rufipogon*, the ARO subgroup was divided into a longer cluster, while TEJ was separated into a shorter cluster. Digu (with *Pi-d2*), IND, and nivara were grouped into one cluster, indicating that *Pi-d2* may have originated from *O.*

nivara. IND was closely related to *O. nivara*, indicating a complex evolutionary history of *Pi-d2* between IND and *O. nivara*. In fact, these rice accessions were divided into *indica*, *japonica*, *O. nivara*, and *O. rufipogon* groups based on the whole genome sequencing of rice accessions (Xu et al., 2011). These results suggest that uncoupling differential domestication processes had occurred in these germplasm accessions.

In contrast to most available genome sequencing datasets, we were able to extract 100% accurate *Pi-d2* sequences (Xu et al., 2011) and use them to predict evolutionary histories and origins. The resulting knowledge will help to understand resistance stability and develop allelic specific DNA markers from portions of genes for use in marker-assisted selection (Jia, 2003).

ACKNOWLEDGMENTS

The authors gratefully thank Professor Wang Wen of the Kunming Institute of Zoology, Chinese Academy of Sciences for providing genome sequences of rice germplasm; Harold Bockelman of the USDA-ARS National Small Grains Collection, Aberdeen, Idaho, USA for providing germplasm accessions; Zaiquan Cheng of the Institute of Biotechnology and Genetic Resources, YAAS; Yiqing Guo of Yunnan Agricultural University for useful discussions; and Michael Lin, Tracy Bianco, and Ellen McWhirter of USDA-ARS Dale Bumpers National Rice Research Center for pathogenicity assays and proofreading, respectively. Research supported by the National Natural Science Foundation of China (#31360331 and #31160355) and Natural Science Fund of Yunnan of China (#2010ZC173).

REFERENCES

Ballini E, Morel JB, Droc G, Price A, et al. (2008). A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. *Mol. Plant Microbe Interact.* 21: 859-868.

Chen D, Chen X, Lei C, Ma B, et al. (2010a). Rice blast resistance of transgenic rice plants with *Pi-d2* gene. *Rice Sci.* 17: 179-184.

Chen D, Chen X, Ma B, Wang Y, et al. (2010b). Genetic transformation of rice with *Pi-d2* gene enhances resistance to rice blast fungus *Magnaporthe oryzae. Rice Sci.* 17: 20-27.

Chen X, Li S, Xu J, Zhai W, et al. (2004). Identification of two blast resistance genes in a rice variety, Digu. *J. Phytopathol.* 152: 77-85.

Chen X, Shang J, Chen D, Lei C, et al. (2006). A B-lectin receptor kinase gene conferring rice blast resistance. *Plant J.* 46: 794-804.

Costanzo S and Jia Y (2010). Sequence variation at the rice blast resistance gene *Pi-km* locus: implications for the development of allele specific markers. *Plant Sci.* 178: 523-530.

Dai Y, Jia Y, Correll J, Wang X, et al. (2010). Diversification evolution of the avirulence gene *AVR-Pita1* in field isolates of *Magnaporthe oryzae*. *Fungal Genet. Biol.* 47: 974-980.

Doron-Faigenboim A and Pupko T (2006). A combined empirical and mechanistic codon model. *Mol. Biol. Evol.* 24: 388-397.

Fjellstrom R, Conaway-Bormans C, McClung AM, Marchetti MA, et al. (2004). Development of DNA markers suitable for marker assisted selection of three *Pi* genes conferring resistance to multiple *Pyricularia grisea* pathotypes. *Crop Sci.* 44: 1790-1798.

Fontaine C, Lovett P, Sanou H, Maley J, et al. (2004). Genetic diversity of the shea tree (*Vitellaria paradoxa* C.F. Gaertn), detected by RAPD and chloroplast microsatellite markers. *Heredity* 93: 639-648.

Jia Y (2003). Marker assisted selection for the control of rice blast disease. *Pesticide Outlook* 14: 150-152.

Jia Y, Wang Z and Singh P (2002). Development of dominant rice blast resistance *Pi-ta* gene markers. *Crop Sci.* 42: 2145-2149.

Jia Y, Valent B and Lee FN (2003). Determination of host responses to *Magnaporthe grisea* on detached rice leaves using a spot inoculation method. *Plant Dis.* 87: 129-133.

Jia Y, Redus M, Wang Z and Rutger JN (2004). Development of a SNLP marker from the *Pi-ta* blast resistance gene by tri-primer PCR. *Euphytica* 138: 97-105.

Hao Z, Wang J, Wang L and Tao R (2009). Influences of the disease resistance conferred by the individual transgenes, *Pi-d2*, *Pi-d3* and *Xa21*, on the transgenic rice plants in yield and grain quality. *African J. Biotechnol.* 8: 4845-4848.

Hittalmani S, Parco A, Mew TW, Zeigler RS, et al. (2000). Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. *Theor. Appl. Genet.* 100: 1121-1128.

Lei J, You N, Huang L, Lei S, et al. (1998) Breeding and utilization of indica rice CMS line Fuyi A. *Hybrid Rice* 13: 8-13

Lei J, You N, Hu A, Lei S, et al. (2004). Research and utilization of resistance to blast in indica rice CMS line Fuyi A. *J. Fujian Agric. Forestry Univ.* (*Natural Science Edition*) 33: 141-147.

Lee S, Costanzo S, Jia Y, Olsen K, et al. (2009). Evolutionary dynamics of the genomic region around the blast resistance gene *Pi-ta* in AA genome *Oryza* species. *Genetics* 183: 1315-1325.

Lee S, Jia Y, Jia M, Gealy D, et al. (2011). Molecular evolution of the rice blast resistance gene *Pi-ta* in invasive weedy rice in the USA. *PLoS One* 6: e26260.

Okuyama Y, Kanzaki H, Abe A, Yoshida K, et al. (2011). A multi-faceted genomics approach allows the isolation of rice *Pia* blast resistance gene consisting of two adjacent NBS-LRR protein genes. *Plant J.* 66: 467-479.

Rozas J, Sánchez-DelBarrio JC, Messeguer X and Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.

Sharma TR, Rai AK, Gupta SK, Vijayan J, et al. (2012). Rice blast management through host-plant resistance: retrospect and prospects. *Agric. Res.* 1: 37-52.

Silue D, Notteghem J and Tharreau D (1992). Evidence for a gene-for-gene relationship in the *Oryza sativa-Magnaporthe grisea* pathosystem. *Phytopathol.* 82: 577-580.

Swanson W, Nielsen R and Yang Q (2003). Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* 20: 18-20.

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.

Wang X, Jia Y, Shu Q and Wu D (2008). Haplotype diversity at the *Pi-ta* locus in cultivated rice and its wild relatives. *Phytopathology* 98: 1305-1311.

Wasano N, Ohgushi A and Ohba M (2003). Mannose-specific lectin activity of parasporal proteins from a lepidoptera-specific *Bacillus thuringiensis* strain. *Curr. Microbiol.* 46: 43-46.

Xing J, Jia Y, Correll J, Lee F, et al. (2013). Analysis of genetic and molecular identity among field isolates of rice blast fungus using an international differential system, Rep-PCR and DNA sequencing. *Plant Dis.* 97: 491-495.

Xu X, Liu X, Ge S, Jensen J, et al. (2011). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* 30: 105-111.

Yang Z, Nielsen R, Goldman N and Pedersen A (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431-449.

Zhou E, Jia Y, Singh P, Correll J, et al. (2007). Instability of the *Magnaporthe oryzae* avirulence gene *AVR-Pita* alters virulence. *Fungal Genet. Biol.* 44: 1024-1034.