J. Plant Biol. (2011) 54:294–302
DOI 10.1007/s12374-011-9166-7

ORIGINAL RESEARCH

Genetic Variation and Evolution of the *Pi9* Blast Resistance Locus in the AA Genome *Oryza* Species

Jinling Liu · Yajun Hu · Yuese Ning · Nan Jiang · Jun Wu · Jong-Seong Jeon · Yinghui Xiao · Xionglun Liu · Liangying Dai · Guo-Liang Wang

Received: 18 May 2011 / Accepted: 24 May 2011 / Published online: 13 August 2011
© The Botanical Society of Korea 2011

Abstract The rice nucleotide-binding site–leucine-rich repeat (NBS-LRR)-encoding resistance (R) gene *Pi9* confers broad-spectrum resistance to the fungal pathogen *Magnaporthe oryzae*. The *Pi9* locus comprises many NBS-LRR-like genes and is an ancient locus that is highly conserved in cultivated and wild rice species. To understand the genetic variation and molecular evolutionary mechanism of the *Pi9* alleles in different rice species, we studied five AA genome *Oryza* species including two cultivated rice species (*Oryza sativa* and *Oryza glaberrima*) and three wild rice species (*Oryza nivara, Oryza rufipogon*, and *Oryza barthii*). A 2.9-kb fragment spanning the NBS-LRR core region of the *Pi9* gene was amplified and sequenced from 40 accessions. Sequence comparison revealed that the *Pi9* alleles had an intermediate-diversified nucleotide polymorphism among the AA genome *Oryza* species. Sequence variations were more abundant in the LRR region than in the NBS region, indicating that the LRR region has played a more important role for the evolution of the *Pi9* alleles. Furthermore, positive selection was found to be the main force promoting the divergence of the *Pi9* alleles, especially in the LRR region. Our results reveal the characteristics and evolutionary dynamics of the *Pi9* alleles among the two cultivated and three wild rice species.

Keywords *Oryza* species · Rice blast · *Pi9* · Polymorphism · Evolution · Rice resistance gene · *Magnaporthe oryzae*

Resistance (R) genes have been effectively used to control crop diseases for over a century. During the past decade, more than 70 R genes have been molecularly characterized in different plant pathogen systems (Liu et al. 2007). The largest group of R genes encodes a nucleotide-binding site (NBS) and leucine-rich repeat (LRR) protein, which confer resistance to a variety of pathogens including viruses, bacteria, fungi, and oomycetes (Collier and Moffett 2009). Two sub-groups of NBS-LRR R genes have been classified and are distinguished by differences in their N-terminal, which contains either the Drosophila Toll and mammalian interleukin-1 receptor homology region (TIR) or the coiled-coil (CC) motif (Meyers et al. 1999). Multiple evolutionary mechanisms involving genetic recombination, gene duplication, and gene conversion affect the generation of NBS-LRR genes (Liu et al. 2007). Studies of R gene allelic polymorphism revealed that two types of evolution (fast and slow) maintain resistance specificity to different sets of pathogen populations (Shen et al. 2006; Yang et al. 2008). A high level of polymorphism generally promotes a rapid

Jinling Liu and Yajun Hu contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s12374-011-9166-7) contains supplementary material, which is available to authorized users.

J. Liu · Y. Hu · Y. Ning · N. Jiang · J. Wu · Y. Xiao · X. Liu (✉) · L. Dai · G.-L. Wang
Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University,
Changsha,
Hunan 410128, China
e-mail: xionglun@yahoo.com

G.-L. Wang (✉)
Department of Plant Pathology, the Ohio State University,
Columbus, OH 43210, USA
e-mail: wang.620@osu.edu

J.-S. Jeon
Crop Biotech Institute & Graduate School of Biotechnology,
Kyung Hee University,
Yongin 446–701, South Korea

evolution of R genes that enables the host to adapt to a rapid change in pathogen populations (Bergelson et al. 2001). This relationship between high polymorphism and evolution rate has been supported by observations of the R gene clusters *Rpp8* (McDowell et al. 1998) and *Rpp1* (Botella et al. 1998) of *Arabidopsis*, *Cf4/9* of tomato (Parniske et al. 1997), and *RGC2* of lettuce (Kuang et al. 2004), and also by observations of the single gene locus *Rpp13* of *Arabidopsis* (Bittner-Eddy et al. 2000; Rose et al. 2004; Allen et al. 2004; Ding et al. 2007). In contrast, a low-level polymorphism was observed in *Rps4* and *RPW8* of *Arabidopsis* (Gassmann et al. 1999; Orgil et al. 2007), *Pita* of rice (Jia et al. 2003), type II of *RGC2* paralogs of lettuce (Kuang et al. 2004), and *Pto* of tomato (Rose et al. 2007).

Balancing selection and positive selection are two major evolutionary forces that maintain R gene allelic polymorphism. In *Arabidopsis*, for example, balancing selection plays a unique role in the evolution of several R loci such as *Rpm1, Rps5*, and *Rpp13*, in which multiple alleles coexist in natural populations (Tian et al. 2002; Rose et al. 2004, 2007; Shen et al. 2006). Conversely, positive selection is the evolutionary force that maintains polymorphism and resistance specificity of many R genes in the arms race between R genes and avirulence (Avr) genes (Sun et al. 2006; Zhou et al. 2007; Seeholzer et al. 2010). It is well documented that the LRR domain in NBS-LRR R genes is the main target for positive selection (Bent and Mackey 2007). For example, Mondragón-Palomino et al. (2002) showed that 4% to 20% of the amino acid sites in different phylogenetic groups of the *Arabidopsis* NBS-LRR genes are subjected to positive selection and that 70% of the positive-selection sites are located in the LRR domain. In the same study, however, a whole-genome polymorphism survey of the LRR regions in 27 R genes of 96 *Arabidopsis* accessions indicated a lack of convincing evidence for positive selection. Only a few alleles are candidates for the rapid increases in frequency expected under positive selection. Mondragón-Palomino et al. (2002) concluded that balancing selection is more important than positive selection in the evolution of most of the R loci in the *Arabidopsis* genome.

Rice blast, caused by the fungal pathogen *Magnaporthe oryzae*, is one of the most destructive diseases of rice. This disease is most effectively controlled by use of resistant rice cultivars. More than 70 R genes in the rice genome have been molecularly mapped. Among the 13 R genes that have been cloned, 11 encode NBS-LRR proteins (Liu et al. 2010), suggesting that rice may use a highly conserved and common resistance pathway to reduce rice blast infection. Therefore, elucidating the evolutionary behavior of the blast NBS-LRR R genes could provide important insights into R gene evolutionary mechanisms. Allele polymorphism studies in rice cultivars have revealed a strikingly low diversity at the *Pita* locus, suggesting a selective sweep by artificial domestication at that locus (Jia et al. 2003; Huang et al. 2008; Lee et al. 2009). In addition, an interspecies divergence of the *Pi-ta* locus was recently reported (Lee et al. 2009). For example, balancing selection was observed in the two cultivated rice species *Oryza sativa* and *Oryza glaberrima* and in the wild species *Oryza barthii*. In contrast, directional selection was found in the wild rice species *O. rufipogon* (Lee et al. 2009).

The rice R gene *Pi9* confers broad-spectrum resistance to *M. oryzae* and is a member of an NBS-LRR gene cluster (Qu et al. 2006). At this locus, at least six resistance alleles with different specificities to rice blast have been identified (Liu et al. 2010). To date, three of the alleles (*Pi9*, *Pi2*, and *Piz-t*) have been cloned (Qu et al. 2006; Zhou et al. 2006). In their study of five cultivars and four wild rice species, Zhou et al. (2007) and Dai et al. (2010) found that genome structure is highly conserved at this locus and that strong positive selection in the LRR region occurs in the NBS-LRR members of the *Pi9* locus. To increase our understanding of the genetic polymorphism and molecular evolution mechanisms of the *Pi9* alleles, we analyzed a 2.9-kb region of the *Pi9* gene in 40 accessions of cultivated rice and wild AA genome species. Our results indicated that the *Pi9* alleles have an intermediate-diversified nucleotide diversity among the tested AA genome *Oryza* species. The LRR region has greater sequence variations than the NBS region, suggesting an important role of the LRR domain for the evolution of the *Pi9* alleles. Positive selection has occurred in the African cultivated rice *O. glaberrima* and its ancestor *O. barthii*, but balancing selection tends to be a stabilizing force for the evolution of the *Pi9* alleles in the Asian cultivated rice *O. sativa*. Therefore, our data suggest that complex selection forces maintain the genetic variation and thereby mediate the evolution of *Pi9* alleles in two cultivated rice species and their corresponding wild ancestor species.

Materials and Methods

Plant Materials and DNA Isolation

A total of 40 accessions belonging to two cultivated rice species and three wild species were used (Table 1). The detailed information concerning the 40 accessions is provided in Table S1. Wild rice accessions were kindly provided by the National Institute of Crop Science, Korea. Genomic DNA of the cultivated accessions was extracted from leaf tissue using the CTAB DNA extraction method (Saghai-Maroof et al. 1994). For wild rice, the genomic

Table 1 The *Oryza* species used in this study

Group	Species	No. of accessions	Genome
Cultivated rice	*Oryza sativa*	17	AA
	Oryza glaberrima	10	AA
Wild rice	*Oryza nivara*	4	AA
	Oryza rufipogon	5	AA
	Oryza barthii	4	AA
	Total	40	AA

DNA was extracted using the protocol described by Chen and Ronald (1999).

PCR Amplification, Cloning, and Sequencing

For PCR amplification, a pair of primers was designed based on the conserved regions in the three cloned blast R genes, *Pi9*, *Pi2*, and *Piz-t* (Fig. 1). The sequence of the forward primer, named NBSF, was 5′-CAT GGA TTC CTA TGC AGA AGA C-3′ (nucleotides 4279–4290; GenBank accession No. DQ285630.1), and the sequence of the reverse primer, named PIR, was 5′-AAT ATT TAA TTA AGC CTC ATT GAT CAT-3′ (nucleotides 7177–7203) (Fig. S1). PCR amplification was performed using the high fidelity Taq enzyme *ExTaq* (Takara) with the following PCR conditions: 95°C for 4 min followed by 28 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 3 min. Amplified DNA products were then cloned into the pGEM-Teasy vector from Promega following the manufacturer's instructions. At least five independent DNA clones for each PCR product were analyzed. DNA sequencing (both strands) was performed by Shanghai Yingjun Biotechnology Company (Shanghai, China).

Sequence Editing, Alignment, and Phylogenetic Analysis

The DNA sequences were edited with MEGA 4.0 (Tamura et al. 2007), and the edited sequences were aligned with the Clustal W 2.0 (Larkin et al. 2007) and manually corrected by MEGA 4.0. Gene coding regions were predicted with GeneScan (http://genes.mit.edu/GENSCAN.html) and Fegenesh (http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind) using the *Pi9* sequence as the reference. The NBS and LRR domain regions were determined using the online tools Pfam (http://pfam.wustl.edu/) and SMART (http://smart.embl-heidelberg.de/) with the Pi9 protein as the reference. Phylogenetic analysis was performed with MEGA 4.0 using the deduced protein sequences of the full-length fragment, NBS regions, and LRR regions. A neighbor-joining tree was generated using *p*-distance.

Nucleotide Polymorphisms Analysis

The DnaSP 5.10 program (Rozas et al. 2003) was used for the analysis of nucleotide and indel polymorphism. The aligned DNA sequences were imported into the DnaSP program to calculate S (number of polymorphic or segregating sites), π (nucleotide diversity), θ (Theta from S, Theta-W), and D (Tajima's D), and to draw the sliding window of nucleotide diversity (π). The BioEdit program (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) was used for the calculation of DNA and protein sequence pairwise identity, and the data were imported into Microsoft Excel for editing and analysis.

Positive-Selection Analysis

The online program “Selection Server” (http://selecton.tau.ac.il/) was used for the positive-selection analysis. Five calculation models were used to identify the positive-selection sites under the query of *Pi9*: 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 mechanistic empirical combination (MEC model; Doron-Faigenboim and Pupko 2006). The data were then imported into Microsoft Excel for statistical analysis and for the drawing of the sliding window.

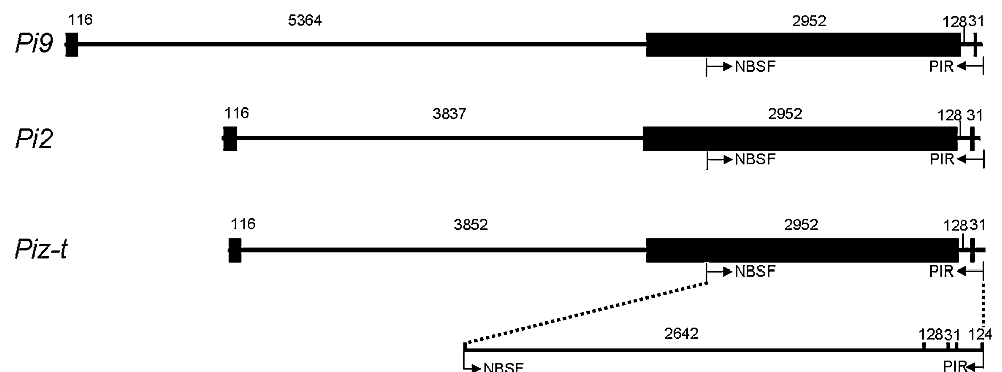

Fig. 1 The *Pi9*, *Pi2*, and *Piz-t* gene structure and the specific amplification primer position used in this study

Results

Sequence Characteristics of the *Pi9* Alleles

A total of 40 sequences of about 2.9 kb were obtained from the 40 AA genome rice accessions (Table 1 and Table S1). The pairwise alignment showed that all 40 sequences are highly homologous with the *Pi9* sequence, with an average identity of 96.7% (range 93.3% to 99.6%) at the DNA level and 93.1% (range 87.1% to 98.7%) at the protein level (Table S2). The phylogenetic analysis showed that these sequences are highly homologous and clustered into a huge clade against the cloned NBS-LRR R protein Pib as described below, suggesting that the amplified sequences are the *Pi9* alleles in their corresponding accessions.

Nucleotide Polymorphism of the *Pi9* Alleles

Of 2,927 nucleotides, 435 polymorphic sites including alignment gaps (~18.9%) were detected among the 40 sequences using the DnaSP program (Table 2). An intermediate-diversified nucleotide diversity (π=0.03348, P=0.000293; Table 2) for the *Pi9* alleles was observed based on the previously published criteria (Yang et al. 2008). The average nucleotide diversity of the LRR region (π=0.04554, P=0.00354; θ=0.04307, P=0.01263) is much higher than that of the NBS region (π=0.02098, P=0.00194; θ=0.02746, P=0.00816) (Table 2; Fig. 2). Both values are similar in cultivated rice and wild rice groups (Table 2; Fig. 2). These results suggest that the LRR domain is important in the variation of the *Pi9* alleles.

A total of 132 and 304 polymorphic sites were observed in the cultivated rice *O. sativa* and *O. glaberrima*, respectively (Table 2). The *O. sativa* group expresses a low polymorphism (π=0.01484, P=0.00124; θ=0.01342, P=0.00481) whereas *O. glaberrima* shows an intermediate -diversified nucleotide diversity (π=0.03918, P=0.00874; θ=0.03816, P=0.01546; Table 2). This result revealed that a polymorphism divergence has independently occurred in the two cultivated rice species, possibly because differences in breeding and artificial selection pressures in the field have led to different polymorphic levels of the *Pi9* alleles in the two species. Given their breeding histories, *O. sativa* has been under strong breeding selection in Asia while *O. glaberrima* has not been extensively selected in Africa (Linares 2002; Londo et al. 2006).

In the wild rice group, a total of 270 nucleotide polymorphic sites were found. The value of the nucleotide diversity (π) was 0.03476 (P=0.00277), and the value of θ from segregating sites was 0.03131 (P=0.01209), indicating an intermediate-diversified polymorphism (Table 2). Furthermore, the polymorphism of the wild rice *O. nivara* and *O. rufipogon*, which are the ancestors of Asian cultivated rice, also showed an intermediate-diversified polymorphism (Table S3; Fig. S1). The polymorphism level of the LRR region (π=0.04493, P=0.00404; θ=0.03764, P=0.01465) in the wild rice group was also much higher than that of the NBS region (π=0.02413, P=0.00205; θ=0.02395, P=0.00943) (Table 2). Interestingly, however, *O. barthii* showed an extremely low nucleotide diversity (π=0.00361, P=0.00078; θ=0.00362, P=0.00206) (Table S3). Additionally, the nucleotide diversity was lower

Table 2 Nucleotide polymorphism of the *Pi9* alleles

Group	Component	Location (nt)	S	π	P^{π}	θ	P^{θ}	Tajima's D
All	Total	1–2,927	435	0.03348*	0.00293	0.03646	0.01058	−0.61538
	NBS	1–1,290	150	0.02098*	0.00194	0.02746*	0.00816	−1.01643
	LRR	1,291–2,645	242	0.04554*	0.00354	0.04307	0.01263	−0.28403
O. sativa	Total	1–2,927	132	0.01484*	0.00124	0.01342*	0.00481	0.21518
	NBS	1–1,290	31	0.00594**	0.00083	0.00712*	0.00276	−0.88702
	LRR	1,291–2,645	93	0.02501*	0.00207	0.02036*	0.00738	0.70395
O. glaberrima	Total	1–2,927	304	0.03918*	0.00874	0.03816	0.01546	−0.18329
	NBS	1–1,290	94	0.02519*	0.00499	0.02576	0.01063	−0.30781
	LRR	1,291–2,645	180	0.05008	0.01152	0.04813	0.01961	−0.22200
Wild rice	Total	1–2,927	270	0.03476*	0.00277	0.03131	0.01209	0.03328
	NBS	1–1,290	93	0.02413*	0.00205	0.02395*	0.00943	−0.06335
	LRR	1,291–2,645	153	0.04493*	0.00404	0.03764	0.01465	0.62944

For Tajima's D, no value is significant at 0.01 level (P>0.01)

S number of polymorphic or segregating sites; π nucleotide diversity, the average number of nucleotide differences per site between two sequences; θ Watterson's nucleotide diversity estimator based on silent site; *Tajima's D* Tajima's D statistic (1989) based on the differences between the number of segregating sites and the average number of nucleotide differences

*P<0.01, statistical significance

Table 3 Positive-selection sites of the *Pi9* alleles as indicated by the M8, MEC, and M5 models

Group	Model	Significance level	Positive sites of NBS	Positive sites of LRR	Total positive sites
All	M8	*P*<0.001*	11	43	54
	M5	–	4	30	34
	MEC	AIC<M8a[a]	0	15	15
O. sativa	M8	*P*<0.001	3	25	28
	M5	–	3	26	29
	MEC	AIC<M8a	1	0	1
O. glaberrima	M8	*P*<0.001	9	32	41
	M5	–	1	14	15
	MEC	AIC<M8a	0	4	4
Wild rice	M8	*P*<0.001	8	36	44
	M5	–	3	27	30
	MEC	AIC<M8a	0	13	13

The positive-selection sites with a statistical significance (the lower bound of confidence interval>1) under Bayesian test was used for calculation here

*Represents a significance level at 0.001

[a] If the score of AIC under MEC model is less than the value of M8a, it is statistically significant

for the LRR region (π=0.00235, *P*=0.00064; θ=0.00243, *P*=0.00155) than for the NBS region (π=0.00478, *P*=0.00110; θ=0.00465, *P*=0.00276) in *O. barthii*, which is opposite to the pattern of nucleotide diversity of the other two wild rice species and the two cultivated rice species (Table S3; Fig. S1).

Phylogenetic Relationship of the *Pi9* Alleles

To evaluate the phylogenetic relationship among the *Pi9* alleles, we constructed neighbor-joining trees using the full-length fragment (Fig. S2a), the NBS region (Fig. S2b), and the LRR region (Fig. S2c). The analysis indicated that all these alleles are clustered into the same clade and belong to the same NBS-LRR sub-family (Fig. S2a, b, and c). Two obvious haplotypes are distinguishable in the phylogenetic tree (Fig. S2c). Group I is mainly composed of the Asian cultivated rice *O. sativa* and its two ancestors, *O. nivara*, and *O. rufipogon*. Most of the African cultivated rice *O. glaberrima* and its ancestor *O. barthii* are clustered in group II. These results suggest that different selection pressures have occurred in the two unique groups of the *Oryza* species during domestication and/or natural selection.

Selection of the *Pi9* Alleles

To test the evolutionary selection dynamics of the *Pi9* alleles in the AA genome rice species, we evaluated the neutral selection with the Tajima's *D* test. Although the Tajima's *D* test result was not statistically significant (Table 2, S3), a negative Tajima's *D* (−0.61538) was observed among all five AA species (Table 2), suggesting that the *Pi9* alleles may have experienced a positive selection; by definition, a negative value of Tajima's *D* indicates positive selection and a positive value indicates balancing selection (Tajima 1989). Therefore, we used the program Selection Server (http://selecton.tau.ac.il/) to further identify the positive-selection sites in the *Pi9* alleles, and we calculated the value of Ka/Ks (Ka=the rate of non-synonymous substitution, Ks=the rate of synonymous substitution) of each amino acid of the deduced Pi9-like protein sequences from the 40 accessions. Five models including three positive-selection models, M8 (Yang et al. 2000), M5 (Yang et al. 2000) and MEC, and two null models, M8a (Swanson et al. 2003) and M7 (Yang et al. 2000), were tested using the selection program (http://selecton.tau.ac.il/). In comparison to the null model M8a, the likelihood ratio test (LRT) of the M8 model was significant at *P*=0.001 except that the LRT of the *O. nivara* group was significant at *P*=0.01 (Table 3; S4). In the MEC model, the LRT was also significant and had a lower Akaike information content score of MEC than that of the M8a model for all groups except a nonsignificant level for *O. rufipogon* sub-species (Table 3; S4). The results of three calculating models were highly consistent (Table 3; S4; Fig. 3; S3). The sliding windows were drawn to show the distribution of the Ka/Ks values across the all 890 amino acids under the M8, M5, and MEC models (Fig. 3; S3). The results showed that the Ka/Ks value of most sites (~80%) was <1, suggesting that these sites were potentially subjected to a purifying selection. Many positive-selection sites (Ka/Ks>1) with a statistically significant Bayesian test were found in both NBS and LRR regions, and the frequency of the positive-selection sites was much higher in the LRR region than in the NBS region (Table S5; Fig. 3; S3, S4).

The M8 model was selected for further analysis of the positive-selection sites, and 54 potential positive-selection sites with statistical significance (6.07% of the 890 sites) were detected among the 890 sites making up the Pi9-like proteins (Table 3, S5; Fig. S4). Of the 54 positive-selection sites, 11 were located in the NBS region, and the other 43 were in the LRR region. In the Asian cultivated rice, only

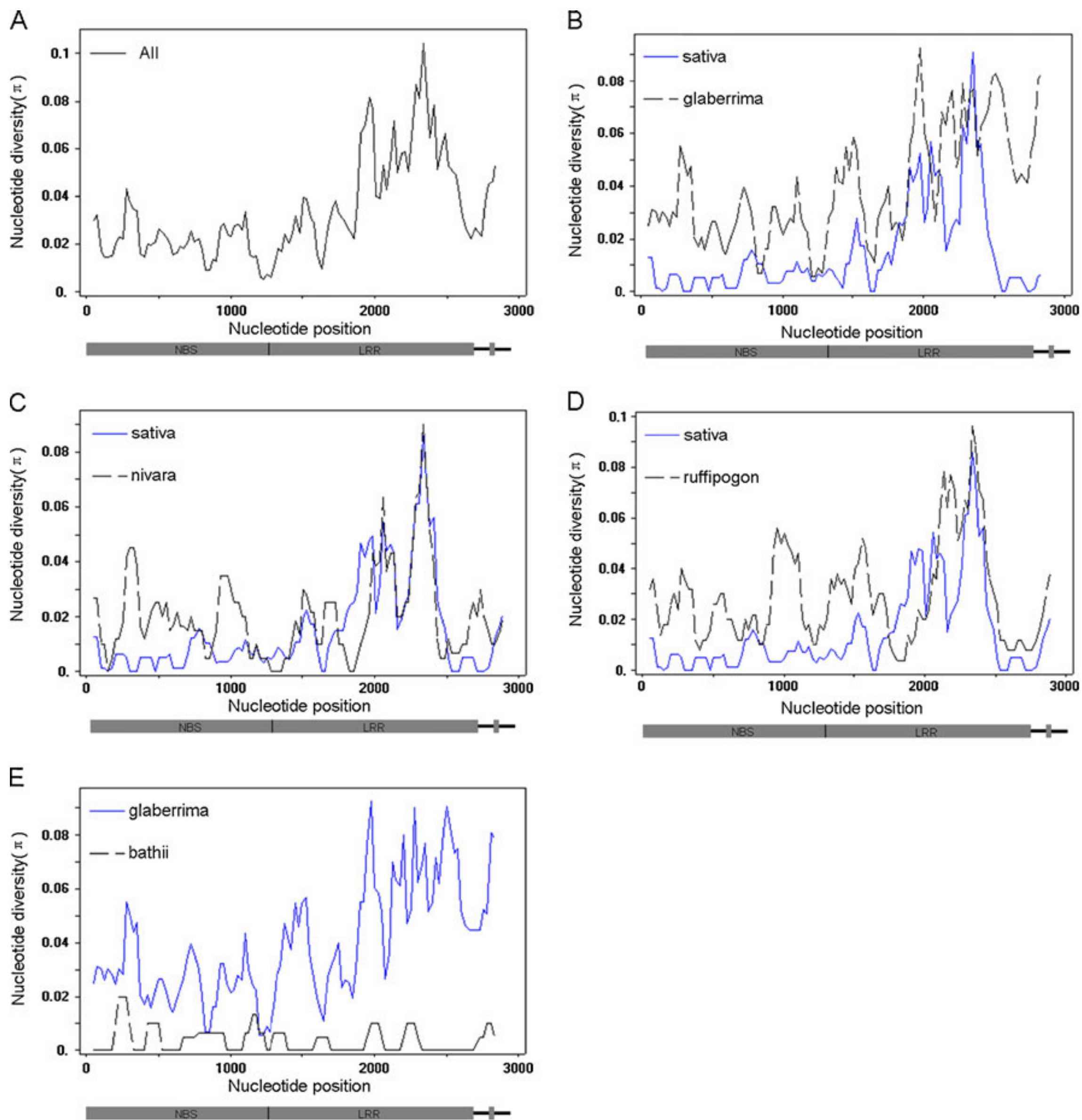

Fig. 2 Sliding window analysis of nucleotide diversity (π) of the *Pi9* Alleles. The nucleotide diversity (π; *Y*-axis) was generated by DNAsp5.0, and the *X*-axis represents the sites of nucleotides. The *map below* the sliding window is the encoding structure of the *Pi9* gene, the *shaded box* denotes the exon region, the *NBS* and *LRR* region are marked on its corresponding region, and the *bold line* represents the intron and 3′ UTR region. The nucleotide diversity of all AA genome *Oryza* species (**a**), and comparison of nucleotide diversity between *O. sativa* and *O. glaberrima* (**b**), between *O. sativa* and *O. nivara* (**c**), between *O. sativa* and *O. rufipogon* (**d**), and between *O. glaberrima* and *O. barthii* (**e**)

28 positive sites (3.15% of the total 890 sites) were found, of which three were in NBS and 25 were in LRR region. (Table 3). A total of 41 positive-selection sites (4.61% of the total sites) were found in *O. glaberrima*, suggesting that positive selection is much stronger in the African cultivated rice *O. glaberrima* than in *O. sativa*.

Of the 890 total sites in the wild rice group, 44 positive-selection sites (4.94%) were found among all three wild species (Table 3). However, many fewer positive-selection sites were observed in each species, with only 12 sites (1.35%) in *O. barthii*, 14 sites (1.57%) in *O. nivara*, and 30 sites (3.37%) in *O. rufipogon* (Table S4). These results suggest that the selection strength differs more between species than within species, which may be caused by differences in evolution pathways or in natural selection forces for the different rice species.

Discussion

The allelic variation at an R gene provides important genetic materials for the generation of new resistance

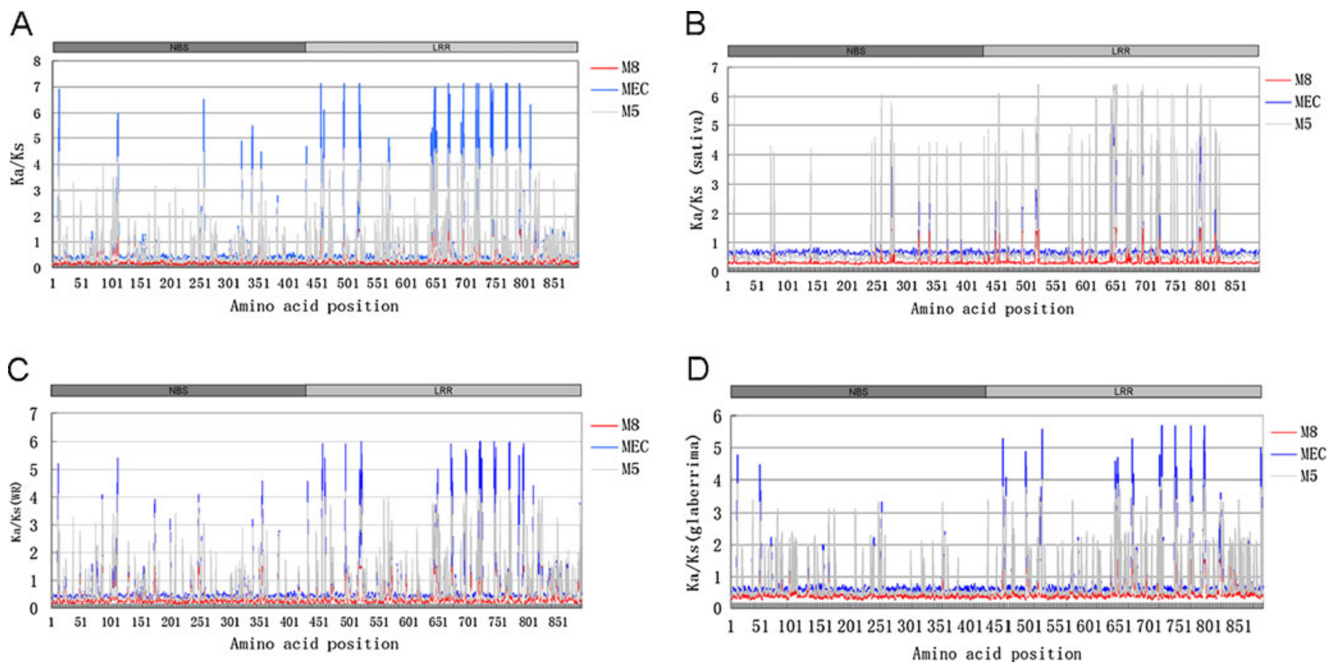

Fig. 3 Sliding window of positive-selection sites of the Pi9 alleles under M8, MEC, and M5 models. The *Y*-axis indicates the ratio of the rate of nonsynonymous substitution (Ka) to the rate of synonymous substitution (Ka) (Ka/Ks); the *X*-axis indicates the position of the Pi9 amino acids in the site. The protein structure is shown *above* the sliding window, the *shaded area* is the *NBS* region, and the *empty area* represents the *LRR* region. The sliding window of the total AA genome *Oryza* intra-species (**a**), of the *O. sativa* species (**b**), of the wild rice intra-species (**c**), and of the *O. glaberrima* species (**d**)

specificity (Shen et al. 2006). Previous research on R genes has revealed that both high and low levels of sequence diversity occur at different R gene loci (Yang et al. 2008). In this study, we investigated the polymorphism of the *Pi9* alleles in 40 accessions of five AA genome *Oryza* species (Table 1). Our results showed that the alleles share a strikingly high identity at nucleotide and protein levels among cultivated rice and wild rice species (Table S2; Fig. S5), suggesting the *Pi9* locus is extremely conserved in the *Oryza* genus. Based on a genome-wide R gene allelic diversity analysis of the rice genome, Yang et al. (2008) described four classes of diversified R genes including conserved (type I; π<0.5%), intermediate-diversified (type II; π=0.5–5%), highly diversified (type III; π>5%), and present/absent genes (type IV). Our analysis indicated that the *Pi9* alleles in the AA genome species likely belong to the intermediate-diversified class. Several studies demonstrated that a high level of polymorphism of R genes generally leads to rapid evolution. In contrast, an R locus with a low level of polymorphism usually has a relatively slow evolutionary rate (Shen et al. 2006; Yang et al. 2008). Therefore, the intermediate level of polymorphism of the *Pi9* alleles suggests that the *Pi9* alleles may have experienced an intermediate rate of evolution during the co-evolution with the rice blast pathogen (Ding et al. 2007; Yang et al. 2008). In addition, our results showed that two obvious clades were clustered in the phylogenetic tree (Fig. S2). Interestingly, these two clades corresponded to two main kinds of cultivated rice species, African and Asian, suggesting that human and natural selection after the rice species differentiated substantially influenced the evolutionary divergence of the rice blast R gene *Pi9*. The LRR region is known to be a major determinant of R gene resistance specificity and plays a role in the variation of the NBS-LRR genes (Collier and Moffett 2009). Our results also showed that, except in the wild rice *O. barthii*, sequence variation is higher for the LRR region than for the NBS region among and within *Oryza* species (Table 3; S3).

Both balancing and positive-selections have been observed for the evolution of R genes. To determine which kinds of selection forces are driving R gene evolution, researchers have developed a variety of tests including the HKA test, Tajima's *D* test, and the McDonald–Kreitman test (Hudson et al. 1987; Tajima 1989; McDonald and Kreitman 1991). This study used the Tajima's *D* test to evaluate the evolution of the *Pi9* gene among and within *Oryza* species. Although the values were not statistically significant (perhaps because the sample population was small), the negative values for Tajima's *D* test suggest that the *Pi9* gene have been subjected to positive selection (Table 2; S3). Therefore, we calculated the value of Ka/Ks of each amino acid using the online program Selection Server to evaluate which sites are under positive selection (Table S5). The analysis led us to three

major conclusions. First, about 6.07% of the sites are potentially subjected to positive selection across all five species (Table 3; S4). Second, the LRR region is the major positive-selection location; about 79.63% (43 in LRR and 11 in NBS) of the positive-selection sites are in the LRR region (Table 3; S4). Third, the strength of the selection differs among *Oryza* species, with the percentage of positive-selection sites ranging from 1.35% in *O. barthii* to 4.61% in *O. glaberrima* (Table 3; S4). In the current study, however, a divergence in selection strength was also observed among the AA genome *Oryza* species (Table 3; S4). The selection strength of *O. sativa* (3.15% of positive-selection sites) significantly differed from that of its wild ancestor *O. nivara* (1.57% of positive-selection sites) but did not significantly differ from that of *O. rufipogon* (3.37% of positive-selection sites). In addition, an obvious selection difference was found between *O. glaberrima* (4.61% of positive-selection sites) and its wild ancestor *O. barthii* (only 1.35% of positive sites). These results demonstrate that a divergent selection strength has affected the evolution of the *Pi9* locus in cultivated and wild rice species.

Acknowledgments This project was supported by the "973" Project (2006CB101904), the "948" Project (2006-G61), the Hangye Project of Ministry of Agriculture, and the National Natural Science Foundation of China (No.30571063). J.-S. J. was supported by the World Class University (WCU) program, Korean Ministry of Education, Science and Technology. We thank Dr. Bruce Jaffee for editing of the paper.

References

Allen RL, Bittner-Eddy P, Grenville-Briggs L, Meitz J, Rehmany AP, Rose LE, Beynon JL (2004) Host–parasite coevolutionary confict between *Arabidopsis* and Downy Mildew. Science 306:1957–1960

Bent AF, Mackey D (2007) Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. Annu Rev Phytopathol 45:399–436

Bergelson J, Kreitman M, Stah EA, Tian D (2001) Evolutionary dynamics of plant R-genes. Science 292:2281–2285

Bittner-Eddy PD, Crute LR, Holub EB, Beynon JL (2000) RPP13 is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. Plant J 21:177–188

Botella MA, Parker JE, Forst LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, Holub EB, Jones JDG (1998) Three genes of the Arabidopsis RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell 10:1847–1860

Chen DH, Ronald PC (1999) A rapid DNA minipreparation method suitable for AFLP and other PCR applications. Plant Mol Biol Report 17:53–57

Collier SM, Moffett P (2009) NB-LRRs work a "bait and switch" on pathogens. Trends Plant Biology 14:521–529

Dai L, Wu J, Li X, Wang X, Liu X, Jantasuriyarat C, Kudrna D, Yu Y, Wing RA, Han B, Zhou B, Wang GL (2010) Genomic structure and evolution of the *Pi2/9* locus in wild rice species. Theor Appl Genet 121:295–309

Ding J, Zhang W, Jing Z, Chen JQ, Tian D (2007) Unique pattern of R-gene variation within populations in *Arabidopsis*. Mol Genet Genomics 277:619–629

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

Gassmann W, Hinsch ME, Staskawicz BJ (1999) The Arabidopsis RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant J 20:265–277

Huang C, Hwang S, Chiang Y, Lin T (2008) Molecular evolution of the Pi-ta gene resistant to rice blast in wild rice (*Oryza rufipogon*). Genetics 179:1527–1538

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–160

Jia Y, Bryan GT, Farrall L, Valent B (2003) Natural variation at the Pita rice blast resistance locus. Phytopathology 93:1452–1459

Kuang H, Woo SS, Meyers BC, Nevo E, Michelmor RW (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell 16:2870–2894

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

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

Linares OF (2002) African rice (*Oryza glaberrima*): history and future potential. Proc Natl Acad Sci USA 99:16360–16365

Liu J, Liu X, Dai L, Wang GL (2007) Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. J Genet Genomics 34:765–776

Liu J, Wang X, Mitchel T, Hu Y, Liu X, Dai L, Wang GL (2010) Recent progress and understanding of the molecular mechanisms of the rice–*Magnaporthe oryzae* interaction. Mol Plant Biol 11:419–427

Londo JP, Chiang Y, Huang K, Chiang T, Schaal BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice *Oryza sativa*. Proc Natl Acad Sci USA 103:9578–9583

McDonald JH, Kreitman M (1991) Adaptive protein evolutionat the *Adh* locus in *Drosophila*. Nature 351:652–654

McDowell JM, Dhandaydham M, Long TA, Aarts MG, Goff S, Holub EB, Dangl JL (1998) Intragenic recombination and diversifying selection contribute to evolution of Downy Mildew resistance at *RPP8* locus of *Arabidopsis*. Plant Cell 10:1861–1874

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20:317–332

Mondragón-Palomino M, Meyers BC, Michelmore RW, Gaut BS (2002) Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. Genome Res 12:1305–1315

Orgil U, Araki H, Tangchaiburana S, Berkey R, Xiao S (2007) Intraspecific genetic variations, fitness cost and benefit of RPW8, a disease resistance locus in *Arabidopsis thaliana*. Genetics 176:2317–2333

Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BB, Jones JD (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell 91:821–832

Qu S, Liu G, Zhou B, Bellizz IM, Zeng L, Dai L, Han B, Wang GL (2006) The sroad-spectrum blast resistance gene *Pi9* encodes a

nucleotide-binding site–leucine-rich repeat protein and is a member of a multigene family in rice. Genetics 172:1901–1914

Rose LE, Bittner-Eddy PD, Langley CH, Charles H, Holub EB, Michelmore RW, Beynon JL (2004) Maintenance of extreme amino acid diversity at the disease resistance gene, *RPP13*, in *Arabidopsis thaliana*. Genetics 166:1517–1527

Rose LE, Michelmore RW, Langley CH (2007) Natural variation in the Pto disease resistance gene within species of wild tomato (*Lycopersicon*). II. Population genetics of *Pto*. Genetics 175:1307–1319

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

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1994) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Seeholzer S, Tsuchimatsu T, Jordan T, Bieri S, Pajonk S, Yang W, Jahoor A, Shimizu KK, Keller B, Schulze-Lefert P (2010) Diversity at the *Mla* powdery mildew resistance locus from cultivated barley reveals sites of positive selection. Mol Plant-Microbe Interact 23:497–509

Shen J, Araki H, Chen L, Chen JQ, Tian D (2006) Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. Genetic, s 172:1243–1250

Sun X, Cao Y, Wang S (2006) Point mutations with positive selection were a major force during the evolution of a receptor-kinase resistance gene family of rice. Plant Physiol 140:998–1008

Swanson WJ, Nielsen R, 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 123:585–596

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tian D, Araki H, Stahl E, Bergelson J, Kreitman M (2002) Signature of balancing selection in *Arabidopsis*. Proc Natl Acad Sci USA 99:11525–11530

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

Yang S, Gu T, Pan C, Feng Z, Ding J, Hang Y, Chen J, Tian D (2008) Genetic variation of NBS-LRR class resistance genes in rice lines. Theor Appl Genet 16:165–177

Zhou B, Qu S, Liu G, Dolan M, Sakai H, Lu G, Bellizzi M, Wang GL (2006) The eight amino-acid differences within three leucine-rich repeats between *Pi2* and *Piz-t* resistance proteins determine the resistance specificity to *Magnaporthe grisea*. Mol Plant Microbe Interact 19:1216–1228

Zhou B, Dolan M, Sakai H, Wang GL (2007) The genomic dynamics and evolutionary mechanism of the *Pi2/9* locus in rice. Mol Plant-Microbe Interact 20:63–71