Theor Appl Genet (2008) 116:165–177
DOI 10.1007/s00122-007-0656-4

ORIGINAL PAPER

Genetic variation of NBS-LRR class resistance genes in rice lines

Sihai Yang · Tingting Gu · Chunyu Pan · Zhumei Feng · Jing Ding · Yueyu Hang · Jian-Qun Chen · Dacheng Tian

Received: 6 February 2007 / Accepted: 21 September 2007 / Published online: 12 October 2007

Abstract The use of plant disease resistance (*R*) genes in breeding programs needs an understanding of their variation patterns. In our current study, we investigated the polymorphisms of 44 NBS-LRR class *R*-genes among 21 rice cultivars and 14 wild rice populations. Our data suggested that there were four basic types of variations: conserved, diversified, intermediate-diversified, and present/absent patterns. Common characteristics at a locus of conserved *R*-genes were: copy-number uniformity, clear divergence (long branches) with other paralogs, and highly identical alleles. On the other hand, copy-number variability, a nearly equal and non-zero branch lengths, and high levels of nucleotide diversity were observed at the loci of highly diversified *R*-genes. Research suggests that the ratio of diverse alleles to the total number of genes at a locus is one of the best criteria to characterize the variation pattern of an *R*-gene. Our data suggested that a significant genetic reduction was detected only in four present/absent *R*-genes, compared with the variation observed in wild rice. In general, no difference was detected between wild rice and cultivars, japonica and indica rice, or between lines from different geographic regions. Our results also suggested that *R*-genes were under strong selection, which shaped *R*-gene variation patterns.

Communicated by D. A. Lightfoot.

Sihai Yang and Tingting Gu contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0656-4) contains supplementary material, which is available to authorized users.

S. Yang · T. Gu · C. Pan · Z. Feng · J. Ding · J.-Q. Chen (✉) · D. Tian (✉)
State Key Laboratory of Pharmaceutical Biotechnology, Department of Biology, Nanjing University, Nanjing 210093, China
e-mail: chenjq@nju.edu.cn

D. Tian
e-mail: dtian@nju.edu.cn

Y. Hang
Jiangsu Province & Chinese Academy of Science, Institute of Botany, Nanjing 210014, China

Introduction

Rice (*Oryza sativa*) is a staple food consumed by nearly half the world's population (Delseny et al. 2001). Although rice production has doubled over the past few decades, due to the introduction of high-yielding varieties/hybrids and improved cultivation practices, producers are still unable to meet ever-increasing global demands. Pathogens are a major threat to the global human food supply (McDowell and Woffenden 2003). Breeding for disease resistance is, therefore, a key component of most breeding programs and has proven to be a highly efficient cultivation strategy (McDowell and Woffenden 2003). It has been well established that in crop monocultures, some *R*-genes impose strong selection on pathogen *Avr* genes resulting in mutations to virulence; consequently, plant breeders must continuously recruit new *R*-genes from wild relatives. However, detecting and screening *R*-gene variation patterns for individual disease in a large scale is a costly and time-consuming process (Pink 2002). It is necessary to have a greater understanding of the variation patterns of resistant germplasms, so that we may develop effective approaches to protect the world's food supply and facilitate breeding of resistant rice species (Jia et al. 2003; McDowell and Woffenden 2003).

Among all *R*-genes, the most prevalent one is the nucleotide-binding site plus leucine-rich repeat (NBS-LRR) genes (Meyers et al. 2003). Previous research has identified 149 and 480 NBS-LRR genes in *Arabidopsis* and rice genomes, respectively (Meyers et al. 2003; Zhou et al. 2004; Yang et al. 2007). Additionally, three blast NBS-LRR resistance disease genes in rice, *Pib* (Wang et al. 1999), *Pi-ta* (Bryan et al. 2000) and *Pi9* (Qu et al. 2006), have been cloned. Polymorphic studies have demonstrated extremely high level of diversity between different populations or even within populations in some *R*-genes, which was assumed to evolve rapidly in response to changes in pathogen populations, e.g., the *rpp5* and *rpp13* loci in *Arabidopsis* (Noël et al. 1999; Allen et al. 2004; Ding et al. 2007b), *L* in flax (Ellis et al. 1999), and Type I of *RGC2* genes in lettuce (Kuang et al. 2004). However, molecular analysis of *R*-gene orthologs and paralogs demonstrated that at least some resistance genes were evolving slowly, e.g., type II of *RGC2* paralogs in lettuce (Kuang et al. 2004), *rps4* in *Arabidopsis* (Bergelson et al. 2001) and *Pi-ta* in rice (Jia et al. 2003). A genome-wide survey of *R*-gene polymorphisms confirmed all variation patterns above in *Arabidopsis* (Bakker et al. 2006).

Recent investigations of allelic comparisons of rice *R*-genes between two fully sequenced genomes, *Oryza sativa* L. var. Nipponbare and 93-11, demonstrated variation patterns of *R*-genes in crops (Yang et al. 2006). In this study, Yang and co-workers characterized four types of NBS-LRR genes based on their allelic diversity, genomic organization, and genealogical relationship: the single-copy-gene dominant group (type I), with the lowest diversity (<0.005); the clustered-gene dominant group (type III), with a high level of diversity and the intermediate one between type I and III (type II). The remaining *R*-genes were present in one genome, but absent in another (P/A *R*-genes, type IV). Notably, this study suggested that the allelic diversity of *R*-genes was high only in approximately one-third of the genome, with the remaining two-thirds displaying very low diversity (0.34% on average). Confirmation of these variation patterns in additional rice lines would be extremely useful, not only to further elucidate *R*-gene evolution in the other crops, but also to facilitate efficient resistance breeding. In our current study, we investigated the polymorphisms of 44 NBS-LRR genes (approximately one-tenth of all the NBS-LRR *R*-genes in rice genome; Zhou et al. 2004) among 21 cultivars and 14 wild rice populations. Our data suggested that the variation pattern in each of these types was generally consistent with the patterns of two rice genomes reported by Yang and colleagues. Additionally, our results indicated that the variation pattern of most *R*-genes reflected strong natural selection pressure, and a few *R*-genes were under artificial selection.

Materials and methods

Plant material and DNA isolation

The variation patterns of *R*-genes in 21 worldwide cultivars and 14 wild rice accessions were investigated (Table S1 of Electronic supplementary material for details). Rice cultivars were obtained from USDA, the National Plant Germplasm System, USA and from Dr. Cailin Wang at the Institute of Food Crops, Jiangsu Academy of Agricultural Sciences, China. Wild rice was a gift from Dr. Dajian Pan, at the National Guangzhou Wild-Rice Conservation, China. Genomic DNA was extracted from plant leaves using the general CTAB DNA isolation procedures.

Selection of *R*-gene loci

In order to investigate a wider range of variation patterns, 44 of the 480 NBS-LRR genes (Zhou et al. 2004; Yang et al. 2006), all of which have been confirmed containing both NBS and LRR domains by using Pfam (http://pfam.wustl.edu/) and SMART (http://smart.embl-heidelberg.de/), were chosen based on the allelic diversities (π) between the two sequenced genomes (Nipponbare and 93-11). We assumed that each of the selected genes was functional at least in some alleles, as their sequences were identical (or almost identical) to the full-length cDNA and EST sequences identified in GenBank. Previous research in our laboratory (Yang et al. 2006) identified 175, 138 and 84 pairs of *R*-genes with an allelic diversity of <0.5, 0.5–5 and >5%, respectively. The remaining 105 *R*-genes were P/A *R*-genes. *R*-genes were categorized as conserved ($\pi < 0.5\%$), highly diversified ($\pi > 5\%$), intermediate-diversified ($\pi = 0.5$–5%) or present/absent (P/A) genes. Gene members were defined as a gene family if nucleotide identity >60% in a genome (Wang et al. 2005). Therefore, a multi-gene family contains $\geq$2 paralogous members (multi-copy genes), and a single-gene family contains only one paralog (single-copy gene). Based on this classification, the randomly selected genes in the current study were as follows: 12 conserved (9 single-copy and 3 multi-copy genes; Table 1), 3 intermediate-diversified single-copy genes (Table 1) and 16 P/A *R*-genes (Table 2). To further select the highly diversified *R*-genes, additional criteria, such as genomic arrangement, were applied. A total of 10 *R*-genes, within a multi-gene family in a complex-cluster, were chosen because of their particular arrangement (*AC134922* cluster in Table 1 and Table S3 of Electronic supplementary material). An additional cluster with two *Pib* homologues were selected, as previous research demonstrated that *Pib* was a functional resistance gene (Wang et al. 1999). Another highly diversified single *R*-gene was also randomly selected for this study.

Table 1 Nucleotide variation of the sampled conserved, diversified and intermediate-diversified *R*-genes among rice lines

Gene loci	Gene name	Type	Cultivar (wild rice)	Copy no.[a]	Sequenced length	Diversity between CDSs in Nipponbare and 9311 (%)[b]	Diversity among all cultivars (%)		Diversity among wild rices	
1	*Os01g23380*	Conserved	17 (4)	1	975	0.12 (0.10)	0.35		0.51	
2	*Os07g29820*	Conserved	17 (4)	1	948	0.34 (0.42)	0.24		0.28	
3	*AL006613*	Conserved	17 (4)	1	621	0.12 (0.18)	0.08		0.16	
4	*Os02g25900*	Conserved	17 (4)	1	738	0.19 (0.28)	0.21		0.18	
5	*Os01g72390*	Conserved	17 (4)	1	630	0.00 (0.00)	0.26		0.31	
6	*Os01g16390*	Conserved	17 (4)	1	549	0.00 (0.00)	0.37		0.90	
7	*Os01g16400*	Conserved	17 (4)	1	528	0.03 (0.00)	0.37		0.53	
8	*Os01g33690*	Conserved	17 (4)	1	348	0.09 (0.00)	0.00		0.00	
9	*Os08g09430*	Conserved	17 (4)	1	375	0.41 (0.53)	0.32		0.31	
AP005828	*Os06g06380*	Conserved	8 (4)	1	894	0.00 (0.00)	0.00	13.70 [c]	0.18	13.70 [c]
	Os06g06390	Conserved	8 (4)	1	786	0.00 (0.00)	0.00		0.06	
	Os06g06400	Conserved	8 (4)	1	834	0.00 (0.00)	0.00		0.06	
11	*Os06g48520*	Intermediate	13 (4)	1	465	1.31 (1.01)	0.54		0.57	
12	*Os04g02110*	Intermediate	11 (4)	1	735	0.61 (0.79)	0.63		0.79	
13	*Os02g18510*	Intermediate	10 (4)	1	678	1.71 (2.22)	0.84		1.05	
14	*Os12g28250*	Diversified	12 (4)	1	529	5.26 (8.62)	2.61		4.37	
pib	*Pib homologs*	Diversified	13 (4)	1–4	854	8.73 (7.60)		6.09 [c]		4.47 [c]
AC134922	*AC134922 homologs*	Diversified	8 (4)	3–5	1034	8.53 (9.15)		13.50 [c]		14.28 [c]

[a] Copy number represents that the number of genes in a cultivar can be amplified by a PCR method

[b] Value in parentheses refers to the allelic diversity of the sequenced LRR region between Nipponbare and 93-11

[c] This is an average of the pairwise diversities among the all homologues (include paralogs and orthologs)

Table 2 Nucleotide variation of sampled P/A *R*-genes among rice lines

Gene name	Presence in genome	No. of cultivars (wild rice)	No. of present cultivars (wild rice)	Sequenced cultivars (wild rice)	Diversity among cultivars (%)	Diversity among wild rice (%)
AAAA02018013	93-11	18 (14)	9 (7)	9 (7)	0.15	0.51
AAAA02015568	93-11	18 (14)	12 (4)	8 (4)	0.09	0.13
AAAA02032841	93-11	18 (14)	5 (6)	5 (6)	0.06	0.30
AAAA02024519	93-11	18 (14)	7 (6)	7 (6)	0.29	0.26
AAAA02021826A	93-11	18 (14)	8 (8)	8 (8)	0.00	0.42
AAAA02035731A	93-11	18 (14)	8 (3)	5 (3)	0.00	0.09
AAAA02036172	93-11	18 (14)	9 (7)	9 (7)	1.00	1.72
AAAA02032153	93-11	18 (14)	14 (7)	7 (7)	5.26	2.99
AAAA02032009	93-11	18 (14)	7 (8)	5 (8)	0.00	3.00
AAAA02032842	93-11	18 (14)	6 (3)	6 (3)	0.13	1.98
Os12g09730	Nipponbare	18 (14)	15 (6)	9 (6)	0.17	2.49
Os10g04670	Nipponbare	18 (14)	11 (4)	8 (4)	0.00	1.25
Os11g45330	Nipponbare	18 (14)	12 (2)	8 (2)	0.05	0.43
Os12g31620	Nipponbare	18 (14)	7 (1)	3 (1)	0.00	–
Os06g16790	Nipponbare	18 (14)	4 (2)	4 (2)	0.00	0.00
Os11g11550	Nipponbare	18 (14)	7 (3)	4 (3)	0.07	0.28
Average			49% (37%)[a]		0.45	1.06

[a] Presence %

PCR amplification and DNA sequencing

Several studies have shown that the LRR domains of *R*-gene are the major determinants of recognition specificity for *Avr* factors (Elli et al. 2000), and that these domains have the highest nucleotide polymorphism (Jiang et al. 2007). Therefore, the LRR region was chosen to detect *R*-gene variation pattern. PCR primers were designed based on conserved sites within the *R*-gene LRR region in Nipponbare and 93-11 (Table S2 of Electronic supplementary material). For most of the selected *R*-genes, PCR products were directly sequenced. For the multi-copy genes, however, the PCR products were cloned into a PGEM-T Easy Vector (Promega), and >20 colonies for each cultivar or wild rice were then sequenced separately until no new homologue sequence could be identified. To further confirm the allelism of the three multi-copy conserved genes, additional primers for each gene were designed to amplify specific regions, including a small part of the gene as well as a sequence of about 1 kb downstream from the stop codon. PCR products were sequenced by using an ABI 3100A automated sequencer. Sequence data were deposited in EMBL and GenBank data libraries under accession numbers, EF641964 to EF642487.

Genotyping of presence/absence *R*-genes

A total of 18 cultivars and 14 wild rice accessions were genotyped by a PCR method to identify presence/absence (P/A) *R*-genes within the 16 selected P/A genes (10 in 93-11 and 6 in Nipponbare). PCR was repeated for two additional times, once at 5°C below the annealing temperature (45–50°C). If no PCR products were obtained, the PCR reaction was repeated using a newly designed primer pair, which increased our ability to exclude false PCR results due to heterogenic sequences.

Sequence analysis

Sequences were aligned first by ClustalW 1.83 (Thompson et al. 1994) and then by MEGA version 3.1(Kumar et al. 2004) for manual corrections of alignment. Nucleotide diversity (π) was calculated by DnaSP v4.0 (Rozas et al. 2003). Phylogenetic trees were constructed based on the Bootstrap neighbor-joining (NJ) method with a Kimura 2-parameter model by MEGA version 3.1 (Kumar et al. 2004). The diverse ratio (DR) was defined as the percentage of different alleles ($\pi > 0.5\%$) from all alleles at a locus.

In order to detect positive selection, we estimated the ratio of non-synonymous (*Ka*) to synonymous (*Ks*) nucleotide substitutions (*Ka/Ks*) on all *R*-genes loci. Pathogen recognition was assumed to mainly take place at the *R*-gene LRR region. More specifically, *R*-gene specificity appeared to reside in the solvent-exposed residues in the ××L×L×× motif that was located within the LRR region. Significant clustering of amino acid replacements in the solvent-exposed residues in the ××L×L×× motif was tested as described in Wang et al. (1998). Accession pairs were ranked for their interallelic *Ka:Ks* ratios at the ××L×L×× motif and the two flanking domains. *P* values were subsequently compared with a Bonferroni-corrected threshold value based on all interallelic comparisons. Protein sequences were initially aligned by ClustalW, and the resulting amino acid sequence alignments were then used to guide the alignments of nucleotide coding sequences (CDSs) using MEGA version 3.1. The *Ka/Ks* ratio was calculated by DnaSP v4.0 based on the equations of Nei and Gojobori (1986).

Sequence exchange was investigated utilizing GENECONV 1.81 (http://www.math.wustl.edu/~sawyer/geneconv/). Gene conversion was tested by finding identical fragments between pairs of sequences within a nucleotide alignment. Global and pairwise *P* values were calculated to assess the statistical significance of the observed fragment lengths. A default setting with 10,000 permutations was used for the analysis with $P < 0.05$ defined as statistically significant.

Results

Genetic variation of conserved *R*-genes

A total of 12 selected conserved NBS-LRR class *R*-genes contained 9 single-copy genes (locus 1–9; Table 1) with an average allelic diversity (π) of 0.14% (0.17% in the examined part of LRR regions) and 3 multiple-copy genes (locus *AP005828*), without single nucleotide polymorphism between Nipponbare and 93-11 (Table 1). Further amplification and sequencing of the same LRR region for the 9 single-copy genes among 17 rice cultivars and 4 wild rice accessions (Table 1), resulted in the amplification of only one copy from each cultivar or wild rice, with continued low allelic nucleotide diversities (0.24% on average and IQR (inter quartile range) = 0.14% in cultivars, and 0.35 on average and IQR = 0.33% in wild rice).

An additional three conserved *R*-genes resided within the 18 kb region, as tandem arrays in a cluster of Nipponbare genome and displayed a high level of nucleotide similarity (average 80.3%). A PCR-cloning strategy was utilized to sequence the tandem arrays among rice lines. A total of 36 non-redundant clones, displaying varying degrees of sequence homology to these genes, were obtained from 8 rice cultivars and 4 wild rice accessions. Interestingly, only three copies of *R*-genes were amplified

from each of these lines. No nucleotide diversity was observed for each of these genes among cultivars (locus *AP005828*; Table 1). In comparison, the pairwise divergence among three paralogs was 17.5, 19.3 and 22.3%, respectively. These data demonstrated that it was feasible to unambiguously identify the allelic relationship for each of the genes in the studied tandem arrays. Downstream sequence of each *R*-gene also confirmed the allelism of the three multi-copy conserved genes (see Materials and methods for details). Members of the *R*-gene family were highly identical among alleles and highly different among paralogs, suggesting this may be a characteristic of conserved multi-copy *R*-genes.

A neighbor-joining tree was constructed for the 225 fragments of the 12 conserved *R*-genes using MEGA 3.1 (Fig. 1a). We noted that this type of *R*-gene usually had long branches between loci, but very short branches within a clade (among alleles). In other words, low nucleotide diversity was observed within each of the clades. Almost all of the pairwise diversities within a clade were <0.5%, with the exception of one clade (*Os01g23380*), in which one allele (Fig. 1a, Minghui 63 in *Os01g23380* locus) diverged from the others at >1% nucleotide diversity. This low level (0.24%) of nucleotide diversity was comparable to the average diversity (0.32%, Yang et al. 2006) of housekeeping genes between Nipponbare and 93-11, the highly conserved genes. These data suggested that most *R*-genes were highly conserved in these loci, although rapid divergence could be found within just a few alleles.

Genetic variation of highly diversified *R*-genes

In addition to the complex-cluster *AC134922*, two other highly diversified loci, *Os12g28250* and *Pib*, were chosen due to their allelic diversity of >5% (5.26–8.73% in CDS and 7.6–8.62% in the examined part of LRR region; Table 1) between Nipponbare and 93-11. The locus *Os12g28250* contains a single-copy *R*-gene. The locus *Pib* is a two *R*-paralogs cluster within an 80 kb region (Wang et al. 1999). The two *pib* paralogs were reported to be remarkably similar (nucleotide identity >95%; Wang et al. 1999). In our study, 24 non-redundant clones were obtained from 17 rice lines, as the copy number in each cultivar varied from 1 to 4 (Table 1; Fig. 1e). PCR-cloning resulted in a minimal estimate of copy-numbers within a cultivar. Therefore, probably, there were ≥4 *pib* paralogs in Tianxiang cultivar (Fig. 1e), Four cultivars, 93-11, Calrose, Hokuriku and Tohoku IL-9, and one wild rice (W0120), had ≥2 paralogs, and only one *Pib* copy was detected in each of the other 8 cultivars and 3 wild rice accessions (Fig. 1e).

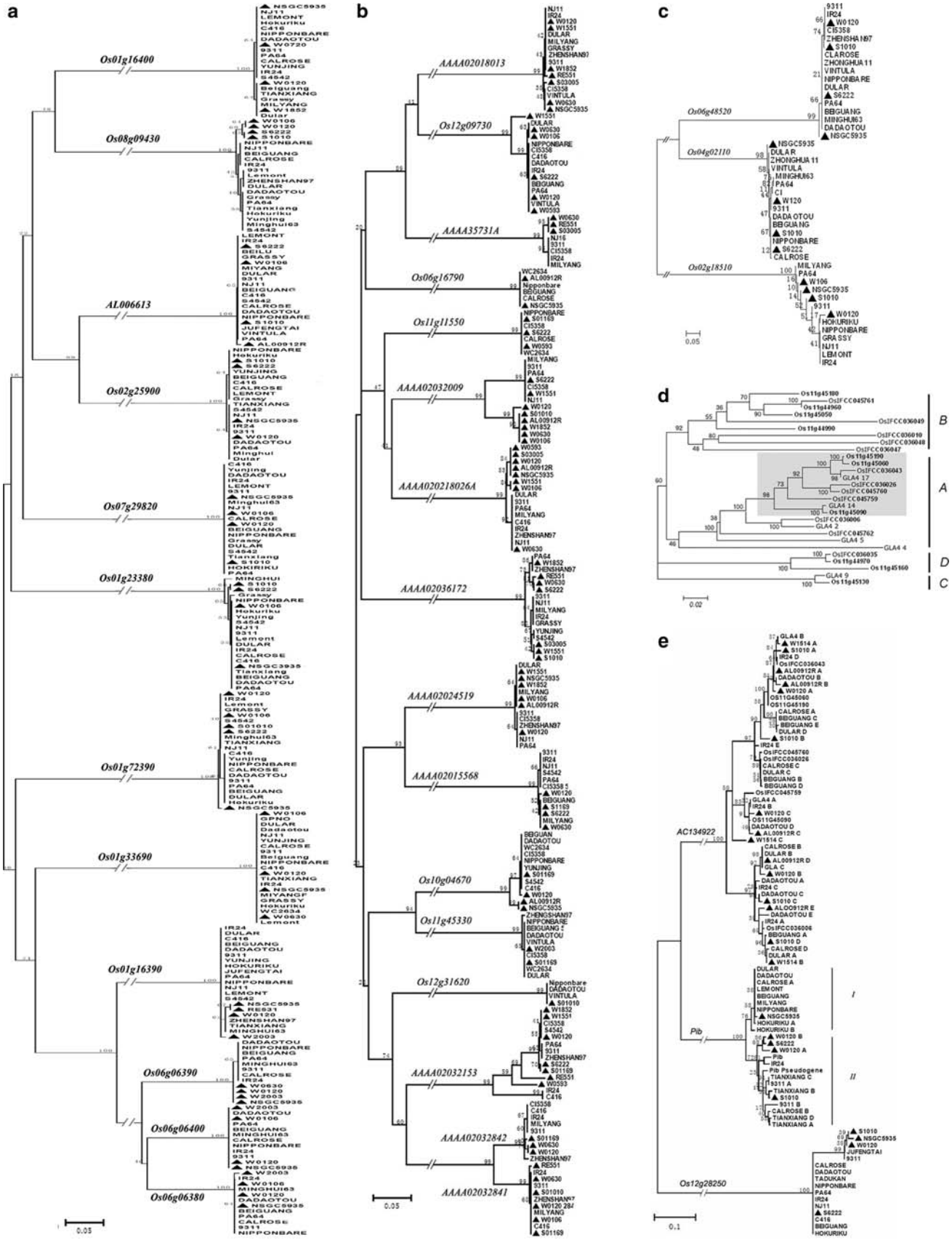
Neighbor-joining tree data suggested that most of the copy-number variation occurred in one homologous group (Group II at *Pib* locus, Fig. 1e). Copy-number varied from 0 to 4, and almost every copy was different from the other. Additionally, from branch length data we were not able to identify alleles from paralogs. Pairwise nucleotide diversity within this group was $5.28 \pm 0.43\%$. The non-zero and nearly equal branch lengths might be the characteristic of highly diversified *R*-genes. In contrast, the characteristics of another *Pib* group (Group I, Fig. 1e) were quite different, with 8.09% difference from Group II noted in nucleotide substitutions. Additionally, almost each cultivar had only one copy, and almost no differences was found between allelic sequences (0.18%), with the exception of one allele from cultivar, *Hokuriku*. Although this level of diversity was similar to the 0.24% noted in conserved *R*-genes, it was significantly less than the average diversity (5.28%) of Group II among genes from different rice lines (*t*-test, $P < 0.001$). These data suggested that different evolutionary forces affected the two *Pib*-homologous groups, resulting in the slower evolution of Group I compared to Group II genes, known to be similar with *Rpp13* in *Arabidopsis* (Ding et al. 2007a).

The highly diversified single-copy *R*-gene (*Os12g28250*) might be a true single-copy locus, where only one copy had been amplified in each of the 12 rice cultivars and 4 wild rice accessions. However, two distinct divergent groups were identified by a bootstrap value of 100% (Fig. 1e) and by a divergence of 8.51% at this locus. Although the average allelic diversity (π) was 3.93%, alleles were completely identical within one group, and slightly different within another.

Structural divergence of the *AC134922* Locus

If the non-zero and nearly equal branch lengths were used as the criteria to identify highly diversified *R*-gene family, the *AC134922* locus was one (the largest) of eight candidates identified in a genome-wide investigation between Nipponbare and 93-11 (Yang et al. 2006). This locus contained 10–12 *R*-paralogs in each of the two genomes, dispersed in approximately 130 and 420 kb DNA sequences, respectively (Fig. 2). Additionally, six *R*-homologues were also identified in a 93 kb sequence from cultivar GLA 4 (GenBank accession: BX005352). However, the total number of copies in GLA 4 is unknown as only one BAC clone was available for this locus. All Nipponbare homologues were approximately located on about 112.9 cM of chromosome 11, while the others were presumably on the same chromosome. Variable copy numbers, as well as the variable length of intergenic and intronic sequences among these three genomes demonstrated that the genomic structure was highly diversified among the studied populations.

A neighbor-joining tree (Fig. 1d), from full-length coding sequences of the three rice cultivars, further demonstrated the highly divergent nature among homologues at


a
Os01g16400
Os08g09430
AL006613
Os02g25900
Os07g29820
Os01g23380
Os01g72390
Os01g33690
Os01g16390
Os06g06390
Os06g06400
Os06g06380
0.05
b
AAAA02018013
Os12g09730
AAAA35731A
Os06g16790
Os11g11550
AAAA02032009
AAAA020218026A
AAAA02036172
AAAA02024519
AAAA02015568
Os10g04670
Os11g45330
Os12g31620
AAAA02032153
AAAA02032842
AAAA02032841
0.05
c
Os06g48520
Os04g02110
Os02g18510
0.05
d
B
A
D
C
0.02
e
AC134922
Pib
Os12g28250
I
II
0.1

◀ **Fig. 1** Phylogenetic trees of sampled *R*-genes and their alleles from different rice lines. **a** conserved *R*-genes, **b** P/A *R*-genes, **c** intermediate-diversified *R*-genes, **d** *R*-genes at AC134922 locus from three rice cultivars (the genes in *shaded area* were used to design the PCR primers), and **e** diversified *R*-genes (the genes for AC134922 are from shaded area in **d**). The *triangle* wild rice. The *Pib* gene (Accession No. AB013448) and *Pib* pseudogene (Accession No. AB013450) sequence coming from GenBank

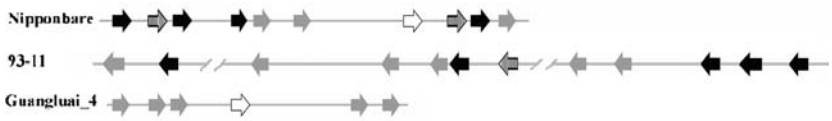


Fig. 2 Non-colinear distribution of *R*-genes from groups A, B, C and D at locus AC134922 among three rice genomes. Genes in group A (*gray arrow*), group B (*black arrow*), group C (*striped arrow*) and group D (*open arrow*). The genes of group A, B, C and D are based on their divergences in Table 3 and in Fig. 1d

this locus. Although the nucleotide similarity among the homologues was >60% (Table 3), the homologues were divided into four groups of *R*-genes (named group A, B, C and D; Fig. 1d, Table S3 of Electronic supplementary material). The nucleotide diversity among these groups was approximately >30%, and the physical location of genes within one group was dispersed in this region without colinearity. For example, homologues from the same group did not reside in corresponding positions among the three cultivars (Figs. 1d, 2). Therefore, orthologous relationships were difficult to discern among members within this locus. Analysis of pairwise nucleotide diversity showed that there was no significant difference between homologues within cultivars (paralogs) and between cultivars (*t*-test, $P = 0.310$ to 0.998; Table 3), which statistically meets the criteria of non-zero and nearly equal branch lengths.

Further analysis of nucleotide diversity between intergenic sequences showed that almost no sequences were alignable between corresponding counterparts. Taken together, these data demonstrated that the AC134922 locus was a complex gene cluster, similar to the *Rpp5* cluster in *Arabidopsis* (Noël et al. 1999) and to the *RGC2* type I in Lettuce (Kuang et al. 2004). These results suggested that the particular genome arrangement, the non-colinearity, and the non-zero and nearly equal branch lengths might contribute to the maintenance of higher genetic diversity among populations. Interestingly, many *R*-genes in this kind of cluster are members of gene families that can carry several distinct pathogen recognition specificities (Hammond-Kosack and Jones 1997).

Genetic variation of *R*-genes in the *AC134922* Locus

In order to further investigate the variation patterns in additional rice lines, we designed a pair of primers, based on conserved sites within the LRR region of group A (Fig. 1d; shaded area) to amplify the homologues at locus *AC134922*. A total of 49 non-redundant clones were obtained from eight rice cultivars and four wild rice accessions (Table 1 and Fig. 1e), as the copy number in each line varied from 3 to 5 (Fig. 1e). The nucleotide diversity varied from 0.00 to 25.2% among homologues, with the average diversity 14.4 ± 6.89%. Of the 1,176 pairwise comparisons, most (87.9%) have a nucleotide diversity of >5%, only 35 alignments <1%. The average divergence among paralogs varied from 9.7 to 16.1%, similar to the average diversity (8.5 to 14.9%) noted among homologues between cultivars (Table 4). There was no significant difference (*t*-test, $P > 0.10$) between the divergence among paralogs and among orthologs. These results, obtained from more

Table 3 Nucleotide diversity (π) of *R*-genes at locus *AC134922* among three rice genomes

Group		π between inter-cultivars				π between inter-cultivars			Average
		Nip versus 9311	9311 versus GLA4	Nip versus GLA4		Nip	9311	GLA4	
A	Range	0.0101–0.153	0.034–0.278	0.007–0.271		0.0062–0.278	0.0191–0.269	0.078–0.264	0.0062–0.278
	Average	0.085 ± 0.0406	0.131 ± 0.063	0.129 ± 0.081		0.071 ± 0.057	0.090 ± 0.057	0.160 ± 0.067	0.118 ± 0.065
	P value				0.998				
B	Range	0.0135–0.223	–	–		0.057–0.099	0.054–0.200	–	0.034–0.223
	Average	0.124 ± 0.048	–	–		0.0802 ± 0.017	0.126 ± 0.045	–	0.117 ± 0.046
	P value				0.310				
C	Pairwise	0.0857	–	–		0.034	–	–	0.0686
D		–	–	0.018		–	–	–	0.018

The *P* values represent two tailed *t*-test results' comparisons between inter-cultivars and intra-cultivars

Table 4 Nucleotide diversity (π) and *Ka/Ks* of *R*-genes at locus A of *AC134922* among rice cultivars

	Copy no.	Nipponbare		9311		Dadaotou		IR24		GLA4		Beiguang		Dular		Calrose		AL00912R	
		π	*Ka/Ks*	π	*Ka/Ks*	π	*Ka/Ks*	π	*Ka/Ks*	π	*Ka/Ks*	Π	*Ka/Ks*	π	*Ka/Ks*	π	*Ka/Ks*	π	*Ka/Ks*
Nipponbare	3	0.102	0.787		0.761		0.784		0.713		0.719		0.784		0.732		0.750		0.776
9311	4	0.088		0.097	0.750		0.726		0.575		0.669		0.749		0.657		0.678		0.735
Dadaotou	5	0.146		0.153		0.146	0.730		0.780		0.952		0.765		0.726		0.731		0.759
IR24	5	0.130		0.131		0.133		0.155	0.673		0.609		0.688		0.655		0.677		0.870
GLA4	3	0.085		0.096		0.143		0.128		0.158	0.746		0.756		0.717		0.732		0.904
Beiguang	5	0.113		0.107		0.150		0.134		0.126		0.114	0.688		0.684		0.719		0.861
Dular	4	0.142		0.140		0.137		0.136		0.148		0.121		0.161	0.712		0.701		0.794
Calrose	4	0.141		0.140		0.137		0.136		0.148		0.121		0.121		0.160	0.730		0.806
AL00912R	5	0.126		0.133		0.135		0.134		0.129		0.140		0.138		0.138		0.158	0.813

cultivars, showed a variable paralogous number among cultivars and an ambiguous relationship between paralogs and orthologs in this complex *R*-cluster, consistent with analysis results of the three genomes at this locus.

Genetic variation of intermediate-diversified *R*-genes

The three intermediate-diversified *R*-genes, selected from 138 such genes, had an average allelic diversity (π) of 1.21% (1.34% in the examined part of LRR region, Table 1). The amplification and sequencing of these genes, within the same part of LRR regions and among 14–17 rice lines, resulted in the amplification of only one copy in each of these lines, with allelic diversities still considered intermediate (0.67% on average, ranging from 0.54 to 0.84%; Table 1). This level of diversity was only 50% of that noted between two sequenced genomes, a reduction that resulted from the grouping of alleles (Fig. 1c). For example, when Nipponbare and 93-11 alleles were sampled from different allelic groups, the level of diversity was larger. However, it decreased after adding more alleles into the two same groups. Data from two or more allelic groups indicated that the average level of diversity (1.49%; Yang et al. 2006) from two genomes might have been slightly over-estimated. However, in comparison to conserved *R*-genes as a whole, these *R*-genes were real intermediate-diversified genes, due to their higher diversity, higher level of differentiation (the process of divergence), and greater number of diverse alleles (Fig. 1c; Table 1).

Genetic variation of present/absent *R*-genes

Genotyping of the 18 rice cultivars and 14 wild rice accessions confirmed that all the 16 P/A genes (10 from 93-11 and 6 from Nipponbare) were under the present/absent polymorphism in both cultivars and wild rice (Table 2). Allele-presence frequencies ranged from 4 to 15 out of 18 cultivars and from 1 to 8 out of 14 wild accessions (Table 2). Among the 16 P/A *R*-genes, *Os06g16790* had the lowest allele-presence frequency (22.2%) and *Os12g09730* the highest (83.3%) in cultivars, with the average frequency of 48.9 ± 18.0%, suggesting that a P/A *R*-gene was present in half of the cultivar population. However, the average frequency (37.0 ± 16.0%) in wild rice was significantly smaller than that in cultivars (*t*-test, $P < 0.01$).

Previous research has suggested that P/A polymorphisms could potentially result from geographic differentiation. Under this scenario, a P/A *R*-gene would be present only in a given geographic area. However, our phylogenetic data (NJ method) revealed no signs of geographic differentiation (Fig. 1b). These results suggested that the high proportion of asymmetric *R*-genes observed between 93-11 and Nipponbare was common among rice populations.

To further elucidate the variation pattern of P/A *R*-genes, 450–1,200 bp fragments were sequenced for each of the 16 P/A *R*-genes. The average nucleotide diversity was 0.45% (0∼5.26%; Table 2) in cultivars, significantly lower than the 1.06% (0∼3.00%) observed in wild rice accessions ($P < 0.05$). The nucleotide diversity was not significantly correlated between cultivars and wild accessions ($r = 0.28$, $P = 0.31$) due to the noticeable reduction in genetic diversity in cultivars at four of the 16 *R*-loci (locus *AAAA02032009, AAAA02032842, Os12g09730* and *Os10g04670* in Table 2; Fig. 1b). The average nucleotide diversity in these *R*-loci in cultivars was 0.075%, smaller than the 2.18% noted in wild rice ($P < 0.005$). Additionally, there was a positive correlation between the diversity of cultivars and wild rice at the remaining *R*-loci ($r = 0.93$, $P < 0.001$). Most (14 of 16) of the P/A loci had a nucleotide diversity of <0.50% among cultivars, suggesting they were conserved. However, this group did not share a uniform level of nucleotide diversity, as two loci displayed a nucleotide diversity >1.0%, suggesting that some of the P/A *R*-genes were under rapid differentiation, e.g., the locus *AAAA02032153* and *AAAA02036172*.

Detection of the natural selection and gene conversion

Pathogen recognition is assumed to mainly take place at the *R*-gene LRR region (Jones and Jones 1997). Based on the LRR protein structure model, solvent-exposed residues in the ××L×L×× motif often exhibit strikingly fast rates of evolution; however, other regions within the LRRs also evolve by positive selection. In the current study, six *R*-genes were detected and displayed significantly greater *Ka* versus *Ks* in the ××L×L×× motif ($P < 0.05$, Table 5). Interestingly, four out of the six *R*-genes (Table 5) had no silent nucleotide variation ($Ks = 0$) but contained segregating protein variants ($Ka > 0$) in the ××L×L×× motif of rice lines, suggesting a strong positive selection in this region. Additionally, four to six *R*-genes revealed significantly $Ka > Ks$ between alleles flanking the ××L×L×× motif ($P < 0.05$; Table 5) to the left and/or to the right, suggesting that selection might be occurring at sites outside the ××L×L×× motif.

Potential targets for selection include identical sequences or those that have significantly low nucleotide diversity (Innan and Kim 2004). *R*-genes with low nucleotide diversity among cultivars were further investigated for artificial selection by identifying the silent nucleotide diversity between *R*-genes and their nearest neighbors. The sliding window method, with a window size of 1,000 bp and a step size of 500 bp, was used to test whether the nucleotide diversity of flanking sequence of *R*-genes between Nipponbare and 93-11 was low similarly, as might be expected if there was an artificial selection in the *R*-locus. Our data suggested that the *AP005828* locus, a cluster locus including three *R*-genes (*Os06g06380*, *Os06g06390* and *Os06g06400*) with a 100% nucleotide identity among cultivars, was a potential candidate (Fig. 3). The three *R*-genes were relatively invariant among cultivars, suggesting young alleles; and all contained segregating protein variants when compared to wild rice. We analyzed a region with low polymorphism in the genome scan and found a 130 kb stretch of DNA encompassing the three cluster *R*-genes that almost completely lacked variation in the two sequenced genomes (Fig. 3). The nucleotide diversity of the 130 kb sequence (including 23 genes, from *Os06g06230* to *Os06g06460*) was significantly lower than the average diversity of the genomes (Fig. 3; International Rice Genome Sequencing Project 2005).

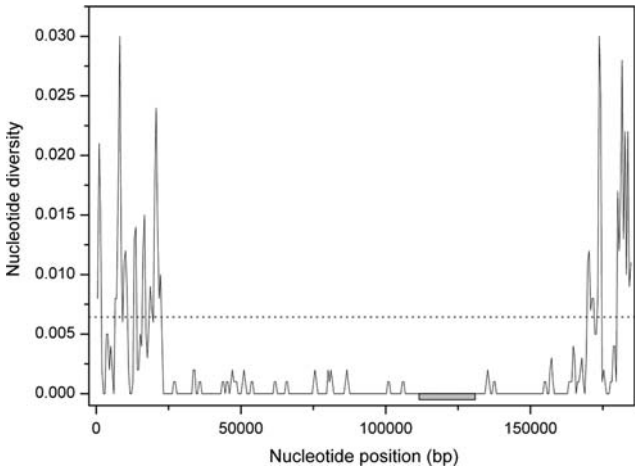


Fig. 3 Sliding window analysis of nucleotide diversity in and around the three low diversity *R*-genes. Window size is 1 kb with a 500-bp increment. The *broken line* stands for the average diversity between the two sequenced rice genomes. The *gray box* is the position of the three *R*-genes in the 190-kb sequence

We further analyzed gene conversion (and/or unequal crossing-over) at all *R*-loci utilizing the GENECONV program, trying to find out the evolutionary forces affecting *R*-genes diversification. Significant gene conversion (and/or unequal crossing-over) was identified among the 28 full-coding DNA in three rice cultivars, as well as among the 49 cloned LRR sequences in 12 different rice lines at the *AC134922* locus. In total, there were 18 and 146 significant gene-conversion events in the full coding and LRR sequences, respectively. Gene conversion fragments ranged from 31 to 787 bp, with approximately 78% of fragments larger than 100 bp. All gene conversion events among the 28 full coding *R*-genes were found in the LRR regions, with none detected in the NBS domains. These results suggested that gene conversion might be an important evolutionary force driving the diversification of the *AC134922* family

Table 5 Interallelic mean *Ka:Ks* ratios for the ××L×L×× Motif and its two flanking domain within the LRR modules

Statistical significance for *Ka* being greater than *Ks* is indicated by an asterisk ($P < 0.05$)

	Left domain			××L×L×× motif			Right domain		
	Ka	*Ks*	*Ka:Ks*	*Ka*	*Ks*	*Ka:Ks*	*Ka*	*Ks*	*Ka:Ks*
AL006613	0.004	0.003	1.333	0.004	0.000	*Ka* > *Ks**	0.000	0.002	0.000
Os08g09430	0.004	0.000	*Ka* > *Ks**	0.008	0.000	*Ka* > *Ks**	0.007	0.007	1.000
Os02g18510	0.000	0.014	0.000	0.012	0.008	1.464*	0.004	0.000	*Ka* > *Ks**
Os04g02110	0.007	0.003	2.333*	0.014	0.000	*Ka* > *Ks**	0.005	0.003	1.667*
Os06g48520	0.000	0.000	–	0.006	0.000	*Ka* > *Ks**	0.008	0.013	0.615
Os12g28250	0.0244	0.083	0.294	0.102	0.028	3.643*	0.055	0.036	1.528*

LRR domain and might be a further adaptation to produce novel resistance alleles. As noted at other loci (Michelmore and Meyers 1998), the frequent occurrence of sequence exchange between paralogs would have made it impossible to detect orthologs by comparison of two haplotypes. In the *Pib* locus, 39 significant gene conversion events were also identified. The exchanged fragments ranged from 51 to 650 bp (average 249 ± 185 bp), with 72% of fragments larger than 100 bp.

Discussion

Variation pattern of *R*-genes in rice genome

Knowledge of the variation patterns of *R*-genes is of fundamental importance for plant breeders attempting to preserve a resistant germplasm. Based on the genome-wide allelic analysis of *R*-genes between two rice cultivars, Yang and co-workers (2006) proposed to use pairwise allelic diversity, the genomic organization and the genealogical relationship with other genes as criteria to characterize variation patterns, resulting in the following categories: conserved (type I; $\pi < 0.5\%$), highly diversified (type III; $\pi > 5\%$), intermediate-diversified (type II; $\pi = 0.5$–5%) and present/absent genes (type IV). Investigation of the allelic variation among rice lines in about one-tenth of the NBS-LRR genes has resulted in a greater understanding of variation patterns. With the additional investigation of more rice lines, the sampled 44 NBS-LRR genes were still generally considered as the same type of *R*-genes classified by Yang et al. (2006), with one exception (*Os12g28250*). This conclusion was demonstrated by a significantly positive correlation between the π from the two genomes (π_{two}) and the π from additional alleles in other populations (π_{others}) ($r = 0.89$, $P < 0.001$; Table 1). This correlation indicated that the variation level of an *R*-gene was generally constant among rice lines, and that the estimation of genetic richness for an *R*-gene from two genomes by the above-mentioned criteria was generally reliable.

However, additional detailed criteria could be inferred from our polymorphic data of a large number of *R*-genes. For example, the 12 conserved NBS-LRR genes contained only one-copy in each of the cultivars, in addition to the low nucleotide diversity (<0.5% in each of these genes). Moreover, simple differentiation, which was defined as the lack of sequence exchange among the genes in the process of divergence, was found among conserved *R*-genes, i.e., *AP005828* family genes (*Os6g06380–06400* in Fig. 1a). We hypothesize that simple differentiation would lead to the long branches between the conserved *R*-genes, which might also explain why so many single-copy genes are conserved *R*-genes (57.1% in type I; Yang et al. 2006).

Our data suggested that highly diversified *R*-genes precisely displayed the contrary characters, i.e., variable copy numbers (0–4 at the group I of *pib*), complex differentiation (the differentiation which resulted in the non-zero and nearly equal branch lengths), sequence exchange among branches, and high diversity among homologues. This kind of complex differentiation could potentially create or maintain the greatest numbers of homologues that differed from each other but did not show much divergence at a locus. For instance, 13 different alleles ($\pi > 0.5\%$ for each pair of them and 5.28 ± 0.43% on average) were detected from a total of 14 genes at group II of *pib* (Fig. 1c). The diverse ratio (DR; See Materials and Methods for definition) was 92.8% at *pib* II, 93.8% at *AC134922*, 37.8% at three loci of intermediate-diversified, and ≤11% at each of the 12 conserved *R*-genes. This criterion could potentially be an ideal measure of characterization of variation patterns of an *R*-gene, because once minimal diversity was set (i.e., > 0.5%), the different alleles could recognize different molecules from pathogens. Consequently, the more diverse alleles are at a locus, the more the molecules (i.e., the products of *Avr* genes or elicitors released from pathogens) could be perceived by this locus. Interestingly, many cloned functional *R*-genes were located at the loci with a high DR. For example, there were three functional *R*-genes at the *Rpp8* locus (5.5 ± 3.1%; DR = 100% when five alleles were sampled from GenBank) and one at the *Rpp13* locus (8.9 ± 3.9%; DR = 53.3% when 60 sampled; Ding et al. 2007a) in *Arabidopsis thaliana*, four at *Cf9/4* in tomato (7.0 ± 2.3%; DR = 100% when 10 sampled; Parniske et al. 1997), three at *L* in flax (4.2 ± 1.6%; DR = 96.2% when 12 sampled; Ellis et al. 1999), and one at *pib* II (5.28 ± 0.43%; DR = 92.8%). The high proportion of functional *R*-genes revealed at the highly diversified locus suggested that the *R*-loci, with higher DR and intermediate diversity, were important resistance loci in plants. By this criterion, the proportion of truly diversified *R*-genes in rice genome should be less than the estimated (84 genes or 17.9%) by comparison of the two genomes (Yang et al. 2006). For example, the locus *Os12g28250* (Fig. 1e) was not qualified as the highly diversified *R*-genes, due to its low DR (18.7%).

The polymorphic data of 44 *R*-genes revealed another distinct type of *R*-genes, those with two-directional differentiation at a locus, similar to the locus of *Rpp13* reported to have multi-directional differentiation (Ding et al. 2007a). For example two differentiated groups of alleles were identified at the locus *Os12g28250, AAAA02032153, AAAA02036172, Os01g23380, Os01g16390, Os01g72390, Os02g25900* and two loci of intermediate-diversified *R*-genes (Fig. 1), and the nucleotide diversity was much higher between groups (3.83 ± 3.63%; ranging from 1.09 to 10.93%) than within a group (0.14 ± 0.18%; ranging

from 0.00 to 0.48%). While the alleles within groups were conserved, they were more divergent between groups. Based on our polymorphic data, there were three basic variation patterns: the conserved, the diversified and the intermediate-diversified pattern, which seemed to be determined by the patterns of differentiation: the simple, the complex and the two- or multi-directional differentiation, respectively. Therefore, the differentiation pattern resulted in alleles with low-diversity, equal-diversity, or grouped-diversity.

The *R*-gene polymorphic classes may correlate with their corresponding pathogen evolution and therefore may have distinct expression patterns. Indeed for the rapidly evolving *R*-genes, i.e., *L* in flax and *RPP13* in Arabidopsis, the highly divergent Avr alleles in pathogens (i.e., AvrL567 and ATR13) are also observed (Dodds et al. 2004; Allen et al. 2004), indicting that the polymorphic patterns in both hosts and pathogens are locked in a co-evolutionary process. In contrast, i.e., in *Bs2* of pepper plant and in *AvrBs2* of *Xanthomonas*, the highly conserved alleles are revealed in both host and pathogen (Wichmann et al., 2005). However, the limited studies on *R*-gene expression cannot distinguish the assumption that the rapidly evolving *R* members might have dramatic expression changes in different ecotypes, while those highly conserved genes might have more conserved expression pattern. Further study will increase our understanding of the correlation between polymorphism and expression patterns of *R*-genes.

Diversity of *R*-genes among wild rice, *indica* rice and *japonica* rice

Rice cultivars are the products of thousands of years of human selection. During crop improvement after domestication, growers exercised extremely strong selective pressure on target genes controlling key morphological and agronomic traits. Indeed, reduced genetic diversity was observed relative to unselected genes (Innan and Kim 2004). Artificial selection can result in a reduced level of variation at neighboring loci as a result of genetic hitchhiking (Maynard-Smith and Haigh 1974) and is believed to be the main evolutionary force acting on domesticated species since their origin 10,000 years ago (Innan and Kim 2004). These beneficial phenotypes were therefore fixed in the founder population of domesticated species in a short (probably very short) time. In cultivated crops, polymorphism is typically reduced by 60–80% (Innan and Kim 2004). One *R*-locus (*AP005828* locus) was detected as a potential artificial selective candidate including a valley of reduced heterozygosity of 130 kb by the polymorphism analysis in the neighboring sequence (Fig. 3). However, no remarkable reduction of nucleotide diversity was detected in the neighboring sequences of other *R*-genes which had a low level of diversity.

Given that the *R*-gene was one of the genes targeted by artificial selection, it would be feasible to expect there would be a different level of polymorphism between cultivar and wild rice. Indeed, a significantly lower diversity was detected at four loci of P/A *R*-genes. However in general, no remarkable difference was detected by our polymorphic data (*t*-test, $P = 0.21$). Almost all the other *R*-genes from wild rice lines (triangles, Fig. 1) were located as common members of *R*-genes at that locus, with no clearly particular branch for any of the *R*-genes from wild rice lines. Additionally, almost no difference was detected at the level of nucleotide diversity between wild rice and cultivars (Table 1). The average diversity within cultivars (0.20%) was only slightly smaller than 0.29% in the conserved and in the intermediate-diversified *R*-genes. Similarly in the diversified type of *R*-genes, *R*-genes from wild rice were not clearly divergent from those in cultivars in the same branch (Fig. 1e). These results indicated that there was no remarkable reduction of genetic diversity in most *R*-genes of domesticated rice.

Interestingly, the differentiation of *R*-genes between indica and japonica rice lines could not be detected in the conserved, intermediate-diversified and diversified *R*-genes, although the sequence divergence between these subspecies has been identified in many genes (Nasu et al. 2002; Garris et al. 2005; McNally et al. 2006). As noted in the genealogical tree of Fig. 1a, c–e, there was no grouping of indica from japonica *R*-genes. The subspecies grouping could only be detected in six loci out of the 16 P/A *R*-genes, where two (*Os10g04670* and *Os11g011550*) contained only japonica and four loci (*AAAA02035731A*, *AAAA02018013, AAAA02032009* and *AAAA02024519*) only had indica members (Fig. 1b). It is interesting to note that these indica or japonica specific P/A *R*-genes reflected those genes that were randomly lost or selectively maintained in a subspecies. Similarly, no obvious geographic relationship with the branches was observed, in spite of the fact that worldwide rice lines were used to investigate possible groupings which might arise geographically, from distant populations. Our data suggested that *R*-genes were under strong selection from similar ranges of pathogens, and that this kind of selective pressure shapes the variation patterns of *R*-genes, over-riding the possible effects from the differentiation between subspecies, the geographic distance and the difference of domesticated and wild rice.

Abundance and origin of present/absent *R*-genes

It was of note that there was a high proportion (22.2% or 105 P/A *R*-genes in two genomes; Yang et al. 2006) of present/absent *R*-genes between Nipponbare and 93-11, and, even more impressive, that the large number of P/A was not a result of sub-species differentiation. These results

suggested there was a large number of P/A *R*-genes among populations, and that these genes could be a major contributor to the genetic variation of plant resistance germplasm. If a rice line contained 10 unique P/A *R*-genes (estimating about 1/5 of P/A in Nipponbare or 93-11), there should potentially be >1,000 different P/A *R*-genes maintained in rice (assuming 100 independent lines in this species). Our genotyping and sequencing data of cultivars and wild rice revealed that no single pair of lines shared the same haplotype with respect to the 16 P/A *R*-genes. Additionally, no linkage disequilibrium between any pair of loci was detected (Chi-square test, $P > 0.10$), suggesting there was no strong interaction between each P/A locus investigated. These data supported the expectation of large numbers of P/A *R*-genes maintained among populations. However, it remains to be seen how many unique P/A *R*-genes can still be identified in each of these lines.

It is clear that the P/A *R*-genes in *Arabidopsis* were from the old duplicates and were maintained by selection (Shen et al. 2006). The long branches between any pair of the 105 P/A *R*-genes (Fig. S1 of Electronic supplementary material) suggested that rice P/A *R*-genes were from old duplicates as well. Our genotyping and sequencing among rice lines did not find more than one copy (the possible young paralog) in a genome, consistent with the expectation of old duplicates. The fact that there was no grouping of indica from japonica in the other types of *R*-genes suggested that the japonica or indica grouping in a few P/A *R*-genes could be the result of adaptive selection, due to the different spectrum of pathogens between environments used for cultivating japonica and indica rice. This is consistent with the hypothesis of selectively maintained P/A polymorphism (Shen et al. 2006).

Interestingly, the two-directional differentiation at an *R*-locus noted in the tree in Fig. 1 was very similar to the P/A tree at early stages. For example, two groups of *R*-genes were highly divergent in the tree for *Os12g28250* (Fig. 1e), *Os01g16390* or *Os01g23380* (Fig. 1a), and each branch of these trees contained only parts of alleles at either of these loci. This leads us to hypothesize that as the divergence continues, each branch may develop into a P/A *R*-locus, suggesting that P/A *R*-genes could originate from the two- or more-directional differentiation. In this scenario, an allele from a rice line should be present and absent in two closely related P/A *R*-loci. However, only 12 pairs of closely related P/A *R*-genes (Fig. S1 of Electronic supplementary material), each pair of which contained one allele from Nipponbare and one from 93-11, were identified. If these pairs resulted from two-directional differentiation, this origin of P/A could account for 22.9% (24/105) of P/A *R*-genes. Future research is, however, necessary to fully elucidate the origin of P/A *R*-genes.

Acknowledgments This work was supported by the National Natural Science Foundation of China (30570987 and 30470122), pre-program for NBRPC (2005CCA02100), 111 project and SUR grant from IBM to D.T., Y.H. or J.Q.C.

References

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

Bakker EG, Toomajian C, Keritman M, Bergelson J (2006) A genome-wide survey of R gene polymorphisms in Arabidopsis. Plant Cell 18:1803–1818

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

Bryan GT, Wu KS, Farrall L, Jia Y, Hershey HP, Mcadams SA, Faulk KN, Donaldson K, Tarchini T, Valent B (2000) A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant cell 12:2033–2046

Delseny M, Salses J, Cooke R, Sallaud C, Regad F et al (2001) Rice genomics: present and future. Plant Physiol Biochem 39:323–334

Ding J, Cheng H, Jin X, Araki H, Yang Y, Tian D (2007a) Contrasting patterns of evolution between allelic groups at a single locus in Arabidopsis. Genetica 129:235–242

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

Dodds PN, Lawrence GJ, Catanzariti A, Ayliffe MA, Ellis JG (2004) The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. Plant Cell 16:755–768

Elli J, Dodds P, Pryor T (2000) Structure function and evolution of plant disease resistance genes. Curr Opin Plant Biol 3:278–284

Ellis JG, Lawrence GJ, Luck JE, Dodds PN (1999) Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. Plant Cell 11:495–506

Garris A, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

Hammond-Kosack KE, Jones JD (1997) Plant disease resistance genes Annu Rev Plant Physiol Plant Mol Biol 48:575–607

Innan H, Kim Y (2004) Pattern of polymorphism after strong artificial selection in a domestication event. Proc Natl Acad Sci USA 101:10667–10672

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

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

Jiang H, Wang C, Ping L, Tian D, Yang S (2007) Pattern of LRR nucleotide variation in plant resistance genes. Plant Sci 173:253–261

Jones DA, Jones JDG (1997) The role of leucine-rich repeat proteins in plant defenses. Adv Bot Res 24:89–167

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

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

McDowell JM, Woffenden BJ (2003) Plant disease resistance genes: recent insights and potential applications. Trends Biotechnol 21:178–182

McNally KL, Bruskiewich R, Mackill D, Buell CR, Leach JE, Leung H (2006) Sequencing multiple and diverse rice varieties connecting whole-genome variation with phenotypes. Plant Physiol 141:26–31

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

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Maynard-Smith J, Haigh J (1974) The hitch-hiking effect of a favorable gene. Genet Res 23:23–35

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y (2002) Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP Markers. DNA Res 9:163–171

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Noël L, Moores TL, Biezen EA, Parniske M, Daniels MJ, Parker JE, Jones JDG (1999) Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of Arabidopsis. Plant Cell 11:2099–2111

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

Pink DAC (2002) Strategies using genes for non-durable resistance. Euphytica 124:227–236

Qu S, Liu G, Zhou B, Bellizzi M, Zeng L, Dai L, Han B, Wang GL (2006) Nucleotide-binding site-leucine-rich repeat protein and is a member of a multigene family in rice. Genetics 172:1901–1194

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

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*. Genetics 172:1243–1250

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Wang GL, Ruan DL, Song WY, Sideris S, Chen L, Pi LY, Zhang S, Zhang Z, Fauquet C, Gaut BS, Whalen MC, Ronald PC (1998) *Xa21D* encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell 10:756–779

Wang X, Shi X, Hao B, Ge S, Luo J (2005) Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 165:937–946

Wang ZX, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, Katayose Y, Sasaki T (1999) The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J 19:55–64

Wichmann G, Ritchie D, Kousik CS, Bergelson J (2005) Reduced genetic variation occurs among genes of the highly clonal plant pathogen *Xanthomonas axonopodis* pv. vesicatoria, including the effector gene avrBs2. Appl Environ Microbiol. 71:2418–2432

Yang S, Feng Z, Zhang X, Jiang K, Jin X, Hang Y, Chen JQ, Tian D (2006) Genome-wide investigation on the genetic variations of rice disease resistance genes. Plant Mol Biol 62:181–193

Yang S, Jiang K, Araki H, Ding J, Yang YH, Tian D (2007) A molecular isolation mechanism associated with high intra-specific diversity in rice. Gene 394:87–95

Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004) Genome-wide identification of NBS genes in rice reveals significant expansion of divergent non-TIR NBS Genes. Mol Genet Genomics 271:402–415