Theor Appl Genet (2010) 121:295–309
DOI 10.1007/s00122-010-1310-0

ORIGINAL PAPER

Genomic structure and evolution of the *Pi2/9* locus in wild rice species

Liangying Dai · Jun Wu · Xunbo Li · Xuejun Wang · Xionglun Liu · Chatchawan Jantasuriyarat · Dave Kudrna · Yeisoo Yu · Rod A. Wing · Bin Han · Bo Zhou · Guo-Liang Wang

Received: 14 October 2009 / Accepted: 21 February 2010 / Published online: 14 March 2010

Abstract Rice blast, caused by the fungal pathogen *Magnaporthe oryzae*, is a devastating disease of rice worldwide. Among the 85 mapped resistance (R) genes against blast, 13 have been cloned and characterized. However, how these genes originated and how they evolved in the *Oryza* genus remains unclear. We previously cloned the rice blast R-genes *Pi2, Pi9*, and *Piz-t,* and analyzed their genomic structure and evolution in cultivated rice. In this study, we determined the genomic sequences of the *Pi2/9* locus in four wild *Oryza* species representing three genomes (AA, BB and CC). The number of *Pi2/9* family members in the four wild species ranges from two copies to 12 copies. Although these genes are conserved in structure and categorized into the same subfamily, sequence duplications and subsequent inversions or uneven crossing overs were observed, suggesting that the locus in different wild species has undergone dynamic changes. Positive selection was found in the leucine-rich repeat region of most members, especially in the largest clade where *Pi9* is included. We also provide evidence that the *Pi9* gene is more related to its homologues in the recurrent line and other rice cultivars than to those in its alleged donor species *O. minuta*, indicating a possible origin of the *Pi9* gene from *O. sativa.* Comparative sequence analysis between the four wild *Oryza* species and the previously established reference sequences in cultivated rice species at the *Pi2/9* locus has provided extensive and unique information on the genomic structure and evolution of a complex R-gene cluster in the *Oryza* genus.

Introduction

In the last two decades, enormous efforts have been devoted to the mapping and cloning of resistance (R) genes in plants. To date, about 70 R-genes from plants and 30 *Avr* genes from pathogens have been cloned and characterized (Liu et al. 2007). Based on their protein structures, cloned R-genes can be divided into several classes. Most of the characterized R-genes encode proteins with conserved nucleotide-binding sites (NBS) and leucine-rich repeats (LRRs). The NBS-LRR family can be further classified into two categories depending on whether the R-genes contain a toll interleukin 1 receptor (TIR) domain that is homologous

---

Communicated by E. Guiderdoni.

L. Dai and J. Wu contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1310-0) contains supplementary material, which is available to authorized users.

L. Dai · J. Wu · X. Wang · X. Liu · G.-L. Wang
Hunan Provincial Key Laboratory of Crop Germplasm
Innovation and Utilization, Hunan Agricultural University,
Changsha 410128, China

J. Wu · X. Li · B. Han · B. Zhou
National Center For Gene Research,
Chinese Academy of Sciences, Shanghai 200233, China

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

B. Zhou (✉)
Institute of Biotechnology, Zhejiang University,
Hangzhou 310029, China
e-mail: bzhzd@zju.edu.cn

D. Kudrna · Y. Yu · R. A. Wing
Arizona Genomics Institute, School of Plant Sciences
and BIO5 Institute for Collaborative Research,
University of Arizona, Tucson, AZ 85721, USA

to the intracellular domain of the *Drosophila* toll and mammalian interleukin-1 receptors in their N-termini (Meyers et al. 1998a, b; Martin et al. 2003). The non-TIR group is referred to as the coiled-coil (CC) group, because they typically have a CC domain at their N-termini. Dicotyledonous plants apparently have both types but monocotyledonous plants do not encode a typical TIR domain at the N-termini (Bai et al. 2002; Meyers et al. 2003; Monosi et al. 2004). NBS-LRR genes in plants tend to cluster at the same locus in the genomes (Hulbert et al. 2001; McHale et al. 2006). Whole genome analysis revealed that all the annotated 149 NBS-LRR genes in Arabidopsis are distributed as 40 singletons and 43 clusters and that 73.2% of them are located in the clusters (Meyers et al. 2003). Similarly, both R-genes and QTLs (quantitative trait loci) are clustered in the rice genome (Wisser et al. 2005 and Ballini et al. 2008). Many known R-genes are members of a complex NBS-LRR cluster such as the P locus in flax (Dodds et al. 2001), the *Cf-4/9* locus in tomato (Parniske and Jones 1999), the *Xa21* locus in rice (Song et al. 1997), the *DM3* locus in lettuce (Meyers et al. 1998a), and the *B4* R-gene cluster in common beans (Geffroy et al. 2009). Ameline-Torregrosa et al. (2008) identified 333 non-redundant NBS-LRRs in the current *Medicago truncatula* draft genome. The unique evolutionary features of these NBS-LRR genes include a high degree of clustering, ectopic translocations from clusters to other parts of the genome, and numerous truncations and fusions leading to novel domain compositions. In addition to its functional and agronomic importance, the NBS-LRR gene family also has a structural role in the genome (Ameline-Torregrosa et al. 2008).

Several genetic and evolutionary mechanisms of R-genes in plants have been revealed from sequence analysis of complex R-gene clusters in the last decade. First, genetic recombination via the unequal recombination event and tandem duplication have been determined to be important for the generation of the R-gene variants and for the evolution of novel resistance specificities (Ellis et al. 1999; Hulbert et al. 2001). Second, both sequence divergence at the intergenic region (Parniske and Jones 1999) and composition of the transposable elements (TEs) (Wei et al. 2002) were proposed to mediate the sequence diversification. Finally, the ectopic and segmental duplication events, resulting in the duplicates to unlinked regions from the original one, are also proposed to be important in the maintenance of the sequence polymorphism at the R-gene loci (Meyers et al. 2003; Ameline-Torregrosa et al. 2008).

Rice blast has became a model pathosystem for understanding the molecular basis of plant–fungal interactions because of the availability of both host and pathogen genome sequences and the genetic resources, and the feasibility of molecular manipulation of both species (Valent 1990; Caracuel-Rios and Talbot 2007; Shimamoto and Kyozuka 2002). Genetic analysis of host resistance to blast was initiated in the 1960s, leading to the mapping of nearly 85 R-genes to date (Ballini et al. 2008). Among them, 13 rice blast R-genes and one QTL gene have been cloned (Fukuoka et al. 2009; Shang et al. 2009). With the exception of Pi-d2 and pi21, which encode a B-lectin kinase protein and a proline containing protein, respectively, all cloned rice R-genes encode NBS-LRR proteins. Recently two studies have investigated the evolution mechanism of the *Pi-ta* locus in both cultivated and wild rice and found several insertions and deletions in the coding region of the single copy gene in the rice genome (Huang et al. 2008; Wang et al. 2008).

The *Pi2/9* locus contains at least six known *R* specificities to the fungal pathogen *Magnaporthe grisea* (Zhou et al. 2007). We cloned the three allelic blast *R-genes Pi2, Pi9*, and *Piz-t* using a map-based strategy (Qu et al. 2006; Zhou et al. 2006). The genes encode highly homologous NBS-LRR proteins and are members of a multigene family on chromosome six. To further characterize the genomic dynamics and organization of the *Pi2/9* locus, we determined and compared the genomic sequences of the *Pi2/9* locus in five rice cultivars. An obvious orthologous or allelic relationship was observed at the *Pi2/9* locus in which the gene members in corresponding positions in different haplotypes (orthologs) were more similar to each other than to the homologs within their respective haplotypes (paralogs). In addition, the 5′ regulatory sequence and N-terminal intron of the paralogs within each haplotype appear to be more sequence- and size-divergent, which might be associated with the suppression of the sequence recombination among the paralogs at the *Pi2/9* locus. These results on the genomic dynamics and organization of the *Pi2/9* locus provide new insight into the evolutionary mechanism of an NBS-LRR R-gene complex in cultivated rice.

The *Oryza* genus contains 23 species with 10 genome types. These species are a historical record of genomic changes that led to the diversification of this genus around the world as well as an untapped reservoir of agriculturally important traits (Wing et al. 2005; Ammiraju et al. 2008) The ‘Oryza Map Alignment Project’ (OMAP) constructed and aligned BAC/STC based physical maps of 11 wild and one cultivated rice species to the reference genome of Nipponbare (Ammiraju et al. 2008). To further investigate the genomic structure and evolution mechanism at the *Pi2/9* locus in *Oryza* wild species, we identified and sequenced five BAC clones from the OMAP project that encompass the *Pi2/9* locus in *O. nivara*, *O. punctata*, *O. minuta*, and *O. officinalis* using PCR and colony hybridizations. The gene content and genomic organization of the NBS-LRR-gene cluster at the *Pi2/9* locus in the five BACs were compared and analyzed. Our results provide new insights into the genomic organization and evolution of this ancient R-gene cluster in rice species.

Materials and methods

Identification of BAC clones covering the *Pi2/9* locus in four wild rice species

We used both colony hybridization screening and BAC end sequence (BES)-based homology searches to identify the BAC clones that encompass the *Pi2/9* locus in four wild *Oryza* species. A 1,986-bp *Pi9* fragment corresponding to the LRR region was amplified using a gene-specific primer pair (forward 5′-gaccctaggtt acaaccacc-3′ and reverse 5′-gggaggagatcgtcagccag-3′). The amplified fragment was used as a probe to hybridize the BAC library filters derived from three wild rice species [*O. punctata* (OP_Ba), *O. minuta* (OM_Ba), and *O. officinalis* (OO_Ba)] (http://www.omap.org/cgi-bin/status/status.cgi). All the positive BAC clones were obtained from the Arizona Genomics Institute (http://www.genome.arizona.edu/orders) and further confirmed by Southern hybridization using two probes (Table 1). The first one is the same DNA fragment used for BAC filter screens. The second probe was amplified with two gene-specific primers (forward 5′-aggaggaagcaggtcg tccc-3′ and reverse: 5′-tcaagattgtgtaggactgg-3′) of the nitrite-induced protein (NIP) gene, which is located at the 5′ side of the *Pi2/9* locus in *O. sativa* (the first gene left side of the NBS-LRR cluster, Zhou et al. 2007). The clones showing positive signals to both probes in the Southern hybridization were selected for further sequencing. The relationship of the selected clones with other positive clones identified in the BAC filter screening and Southern hybridization confirmation was further investigated using the BAC *Hin*dIII fingerprinting approach. In both *O. punctata* and *O. officinalis*, all the BAC clones identified in the BAC filter screening were found to be clustered in a single contig, indicating that both two species have a single *Pi2/9* locus. However, the relationship of the positive BAC clones appeared much complicated in *O. minuta*. We, therefore, selected two BAC clones with distinctive *Hin*dIII fingerprinting for sequencing. The BAC clones from the annual and perennial ecotypes of *O. rufipogon* were identified by BES-based homology searches. The genomic sequence of the *Pi2/9* locus in cultivated rice was used as query sequence to search against the databases of BESs of both *O. rufipogon* and *O. nivara* BAC libraries.

DNA sequencing, sequence assembly and gap filling

The five BAC clones listed in Table 1 were completely sequenced, and the other two *O. minuta* BACs (OM_Ba0177F21 and OM_Ba0165G09 illustrated in Fig. S2) were partially sequenced. The shotgun library construction and subclone sequencing of three BAC clones (OP_Ba0034K08, OO_Ba0034L17, and OM_Ba0333A15) were conducted at the Arizona Genomics Institute in Tucson. The other two BAC clones (OR_BBa0100B19 and OM_Ba0024C01) were sequenced as described previously (Zhou et al. 2006). The sequence reads were assembled with the Phred and Phrap software packages (Ewing et al. 1998; Ewing and Green 1998) and were edited with the Consed program (Gordon et al. 1998). At 8–10 time redundancies, internal low-quality regions of each contig were treated as gaps, and were bridged using the primer walking procedure as described previously (Zhou et al. 2006).

Computational analysis of the sequences

The genomic sequences were annotated using the gene prediction program Fgenesh (http://www.softberry.com/berry.phtml) and were manually edited by homology search against public databases. The genomic sequences of the *Pi2/9* locus in cultivated rice were used as the main reference to predict the *Pi2/9* gene family members in the four wild *Oryza* species. Repetitive sequences were identified by searching the BAC sequences against the database [nucleotide

Table 1 The features of the Pi2/9 locus in four different wild rice species

Species	Genome designation	Locus designation	BAC clone	Size (bp)	Region for sequence analysis (bp)[c]	Boundary genes		Accession no.
						NIP	PK	
O. nivara	AA	Pi2/9_ON	OR_BBa0100B19	155,708	35,734–155,708	Yes	Yes	GQ280265
O. punctata	BB	Pi2/9_OP	OP_Ba0034K08	196,170[a]	12,965–117,974	Yes	Yes	GQ280266
O. officinalis	CC	Pi2/9_OO	OO_Ba0034L17	208,417[a,b]	36,725–204,946	Yes	Yes	GQ280269
O. minuta	BBCC	Pi2/9_OM_BB	OM_Ba0333A15	126,386	38,983–126,386	Yes	No	GQ280267
		Pi2/9_OM_CC	OM_Ba0024C01	101,868[a]	1–101,868	No	Yes	GQ280268

[a] The total size of the BAC clone was estimated by the assumption of the size of all the ordered contigs without gaps

[b] The BAC end sequence at the SP6 side matches to the expected location, whereas the T7s one was found to match at the position 131,938 bp which situates in the middle of the sequence. However, we believed that the ordered contigs was assembled correctly with both sequence verification and the scaffold of the subclones. Moreover, the sequence surrounding this location exhibits a perfect synteny with the one in BAC clone OM_Ba0024C01, suggesting again the correct assembly of this BAC sequence

[c] The sequence fragment was retrieved from the corresponding BAC clone and used for analyses in this study

Table 2 The vertical groups of the *Pi2/9* gene family in *Oryza* genus

Genes: vertical Groups	Species					
	O. sativa	*O. nivara*	*O. punctata*	*O. officinalis*	*O. minuta*-BB	*O. minuta*-CC
VG1	Nbs1-OS-NPB[a]	Nbs1-ON	NP	Nbs1-OO	NP	Nbs1-OM-CC
VG2	NP	NP	Nbs1-OP	NP	NP	NP
VG3	Nbs2-OS-75	NP	NP	Nbs2-OO	NP	Nbs2-OM-CC
VG4	Nbs2-OS-NPB Nbs4-OS-NPB Pi9	Nbs2-ON	Nbs2-OP	Nbs3-OO	Nbs1-OM-BB	Nbs3-OM-CC Nbs5-OM-CC Nbs7-OM-CC Nbs9-OM-CC
VG5	Nbs5-OS-NPB	NP	Nbs3-OP	Nbs6-OO	Nbs2-OM-BB	Nbs12-OM-CC
VG6	Nbs6-OS-NPB	NP	NP	NP	ND	NP
VG7	Nbs7-OS-NPB	NP	NP	Nbs4-OO	ND	Nbs4-OM-CC Nbs6-OM-CC Nbs8-OM-CC Nbs10-OM-CC
VG8	NP	NP	Nbs4-OP	Nbs5-OO	ND	Nbs11-OM-CC

"*NPB*" is abbreviated for the rice cultivar Nipponbare, *NP* not present, *ND* not determined due to undetermined sequence

collection (nr/nt) at NCBI] using BLASTN. Based on the homology to known repeats, these repetitive sequences were further classified into respective groups, i.e., retrotransposon, transposon, solo-LTR, and unknown repeats, as listed in Table 2. The Matcher program (http://bioweb.pasteur.fr/seqanal/interfaces/matcher.html) was used for pairwise sequence analysis. The sequence of the NBS domain of each gene was retrieved by the BLAST2 program using the NBS sequence of the *Pi9* gene, and was used for the phylogenetic analysis by Clustalx program (Thompson et al. 1997). The phylogenetic tree was viewed either with the Njplot (Perriere and Gouy 1996) or MEGA 4 (Kumar et al. 2001) program. The predicted coding sequences (CDS) of the *Pi2/9* gene family members in different vertical groups were used for the PAML analysis to detect the sites subjected to the diversifying selection (Yang 1997; Yang et al. 2000). The insertion time of LTR-retrotransposons was calculated following a similar manner as described previously (Ma and Bennetzen 2004). The pairwsie distance of two LTRs of a particular LTR-retrotransposon was calculated using the MEGA4 program (Kumar et al. 2001) and the insertion time of each LTR-retrotransposon was then estimated based on a proposed mutation rate of 1.3×10^{-8} substitutions per site per year.

Results

Establishment of the *Pi2/9* locus in four wild rice species

To determine the genomic structure of the *Pi2/9* locus in wild rice species, we identified five BAC clones from four wild rice species, i.e., *O. nivara*, *O. punctata*, *O. officinalis*, and *O. minuta* (Table 1). Other than *O. minuta*, each species has a single representative BAC clone that spans the *Pi2/9* locus. To simplify the sequence analysis, we only sequenced the region covering the *Pi2/9* locus in each BAC clone. The clone OR_BBa0019B19 identified from *O. nivara*, the annual ecotype of *O. rufipogon*, showed almost identical *Hin*dIII digestion patterns with both OR_ CBa0092G16 and OR_CBa0094P06 from the perennial ecotype of *O. rufipogon* by reciprocal Southern hybridization using the whole BAC DNA as the probe (Fig. S1). Moreover, OR_BBa0019B19 encloses both OR_CBa0092 G16 and OR_CBa0094P06 as deduced by the mapped positions of their BESs along the reference sequence of the *Pi2/9* locus in Nipponbare (GenBank Acc. DQ454158). Therefore, we used the sequence of OR_BBa0019B19 *O. nivara* to represent the *Pi2/9* locus in the wild rice species with the AA genome constitution in this study (Table 1).

We demonstrated in our previous study that the *NIP* and *PK* genes are always tightly associated with the *Pi2/9* locus at the 5′ and 3′ boundaries, respectively (Zhou et al. 2007). In this study, we utilized these two genes as the boundary genes to ensure the *Pi2/9* locus in wild rice species are completely determined. The *Pi2/9* locus in *O. nivara*, *O. punctata*, and *O. officinalis* was completely established because the BAC sequence in each species contains both the *NIP* gene at the left border and the protein kinase (*PK)* gene at the right border (Table 1). The two BAC clones of *O. minuta*, OM_Ba0333A15 and OM_Ba0024C01, contain the *NIP* and *PK* genes, respectively, indicating that the former covers the left border and the latter covers the right border of the *Pi2/9* locus (Table 1). However, these two BAC clones do not overlap. With respect to the location of the fingerprint contigs (FPCs), OM_Ba0024C01 is in contig

1,719, whereas OM_Ba0333A15 is not mapped to any contigs (http://www.omap.org/fpc/WebAGCoL/OM__Ba/WebFPC/), further indicating that these BAC clones are not in the same physically linked contig. Sequence comparison revealed that the sequences of OM_Ba0333A15 and OM_Ba0024C01 at the *Pi2/9* locus are similar to those of OP_Ba0034K08 in *O. punctata* (BB) and OO_Ba0034L17 in *O. officinalis* (CC), respectively (see details below). Given that *O. minuta* is an allotetroploid species with the BBCC genome constitution, we, therefore, speculated that the sequences of OM_Ba0333A15 and OM_Ba0024C01 may represent the *Pi2/9* locus of the BB and CC subgenomes, respectively, in *O. minuta.*

To establish the entire *Pi2/9* locus in *O. minuta,* we used the sequences of both OM_Ba0333A15 and OM_Ba0024C01 as a query to search against the BES database of *O. minuta*. Five BAC clones (OM_Ba0049B19, OM_Ba0180N01, OM_Ba0178G06, OM_Ba0165G09, OM_Ba0177F21) whose BESs are mapped at the sequence of OM_Ba0024C01 were identified. Interestingly, these five BACs were found to be located in the same contig 1,719 (http://www.omap.org/fpc/WebAGCoL/OM__Ba/WebFPC/), indicating that they are overlapped with OM_Ba0024C01. Southern blot screening with *O. minuta* BAC filters confirmed the presence of the *Pi9* homologous sequence in these clones because they all showed positive signals on the BAC filters using the *Pi9* DNA fragment as a probe (data not shown). The relationship of these BAC clones with OM_Ba0024C01 was further determined by BES locations, BAC fingerprinting analysis, and PCR validation. As illustrated in Fig. S2, OM_Ba0177F21 covers the *NIP* side of the *Pi2/9* locus, and OM_Ba0165G09 connects both OM_Ba0177F21 and OM_Ba0024C01. Both OM_Ba0165G09 and OM_Ba0177F21 were partially sequenced to identify the NBS-LRR genes in the two BACs. Therefore, the *Pi2/9* locus in the CC subgenome of *O. minuta* was completely established. However, we were not able to identify any BAC clones extending from OM_Ba0333A15, leaving the *PK* side of the *Pi2/9* locus in the BB subgenome of *O. minuta* undetermined. Given that the sequences of the BB and CC subgenomes established in this study contain the hypothesized boundary genes of the *Pi2/9* locus and share high-sequence similarity to the ones in *O. punctata* and *O. officinalis*, respectively, we believed that they are likely syntenic to each other.

Gene content and genomic organization of the NBS-LRR gene cluster at the *Pi2/9* locus in wild rice species

We adopted a similar system of nomenclature for naming the NBS-LRR genes at the *Pi2/9* locus in the four wild rice species as previously described in cultivated rice (Zhou et al. 2007). The *Pi2/9* gene family members are designated *Nbs* followed by two suffixes separated by a hyphen. The numeric suffix before the hyphen denotes the member's genomic position within the cluster. The two-letter suffix after the hyphen identifies the abbreviation of the species designation (Table 2). For the NBS-LRR genes in the BB and CC subgenomes of *O. minuta*, additional suffixes denoting the subgenome designation are specified (Table 2). We also included the *Pi2/9* family members representing each orthologous subgroups in cultivated rice for comparative analysis, with cultivar source indicated by a suffix after the second hyphen (Table 2). Nbs3-OS-NPB was not included in the analysis because it is a highly truncated gene.

The copy number of the NBS-LRR genes differs in the wild rice species. *O. nivara* contains only two NBS-LRR genes (Fig. 1b). There are four, six, and 12 NBS-LRR genes identified in *O. punctata*, *O. officinalis*, and the CC subgenome of *O. minuta*, respectively (Fig. 1c–f). The BB subgenome of *O. minuta* contains two NBS-LRR genes at the *NIP* side (Fig. 1d). The NBS-LRR gene cluster in *O. nivara* is flanked by the *PK* gene at the right border, exhibiting the same organization as that in *O. sativa* (Fig. 1b). However, the NBS-LRR gene cluster has undergone many genomic rearrangements in *O. punctata*, *O. officinalis*, and *O. minuta* in which the *PK* gene is not the flanking gene of the NBS-LRR cluster at its right border (Fig. 1c, e, f). For example, *Nbs4-OP*, *Nbs5-OO*, *Nbs6-OO*, *Nbs11-OM-CC*, and *Nbs12-OM-CC* are located outside of the *PK* gene at the right border in their respective species (Fig. 1c, e, f). The presence of the NBS-LRR genes outside of the *PK* gene made it difficult to determine the boundary of the *Pi2/9* locus in these wild rice species. Moreover, we found that *Nbs6-OO* and *Nbs12-OM-CC* have the opposite transcriptional direction when compared with their homologues at the *Pi2/9* locus in other species (Fig. 1e, f).

The *Pi2/9* gene family members in wild rice species are conserved in structure and categorized into the same subfamily

By comparing with the NBS-LRR genes in the cultivated rice, we annotated the coding sequence of each *Pi2/9* gene family member in the four wild rice species. A feature of the *Pi2/9* gene family members observed in the cultivated rice was revealed in the wild rice species (Zhou et al. 2007). Most of the *Pi2/9* gene members contain two exons split by a phase-2 intron, where a tryptophan codon (TG|G) is completely conserved. The first exon is 116- or 119-bp long followed by the intron with different sizes, from 165 bp (*Nbs2-OO* and *Nbs2-OM-CC*) to 14,729 bp (*Nbs1-OM-BB*) (Table S1). However, *Nbs1-OP* and *Nbs4-OP* are more divergent and have a 104- and 105-bp first exon, respectively (Table S1). The difference in the structure of these two genes probably results from the sequence mutations in the junction region between the first intron and exon.

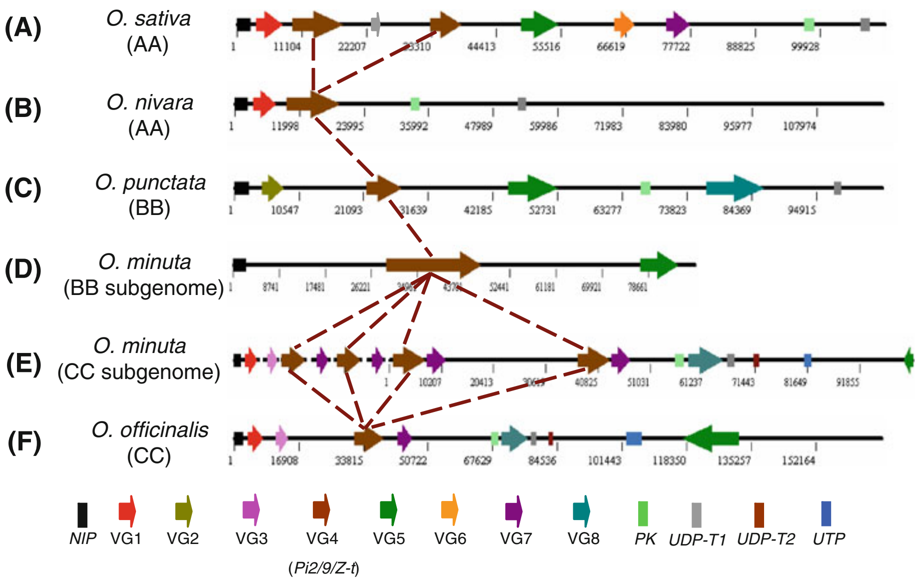

Fig. 1 Genome organization of the *Pi2/9* locus in four wild species. The *Pi2/9* gene family members are indicated by the arrows with the transcriptional direction. The family members within a same vertical group (VG) are indicated with the *same color*. The VG4 in which the three cloned resistance genes (*Pi2*, *Pi9* and *Piz-t*) are located is *highlighted* by *dashed lines* in *brown* between different species. The non-*Pi2/9* homologues are indicated with *rectangles*. *NIP* nitrate-induced protein gene, *PK* protein kinase gene, *UDP-T* UDP-glycosyltransferase-like protein, *UTP* UTP-glycosyltransferase-like protein. **a** The *Pi2/9* locus in the cultivated rice line Nipponbare. *Nbs3-OS-NPB* was indicated in a *filled gray arrow*, which shows a significant difference with other NBS-LRR genes as discussed previously (Zhou et al. 2007), **b** The *Pi2/9* locus in *O. nivara*, **c** The *Pi2/9* locus in *O. punctata*, **d** the *Pi2/9* locus in the BB subgenome of *O. minuta*, **e** The *Pi2/9* locus in the CC subgenome of *O. minuta*, **f** the *Pi2/9* locus in *O. officinalis.* The figure was drawn with the BioEdit program (Hall 1999)

The sequences of the *Pi2/9* gene family members are quite different from each other among the wild rice species. For example, the entire CDSs of *Nbs1-OP* and *Nbs2-OO* share as low as 18% identity, whereas the CDSs of *Nbs1-ON* and *Nbs1-OS-NPB* share as high as 99% identity. Further, the 5′ portion of the CDS containing the N-terminus and NBS region appears less-sequence divergent than the 3′ portion containing the LRR region. For example, *Nbs1-OP* and *Nbs2-OO* share 46 and 37% sequence identity in amino acid in their NBS and LRR regions, respectively. *Nbs1-OP* and *Nbs2-OO* were the most sequence-divergent pair of the *Pi2/9* gene family members in the wild rice species. Nevertheless, *Pi2/9* gene family members and *Pib*, the most related *Pi2/9* homologue in the rice genome, shared only 32% or less-sequence identity in amino acids. Therefore, the *Pi2/9* gene family members are more related to each other than to the closely related homologue *Pib*.

Cladistic analysis based on the NBS sequences of the Pi2/9 gene family members further revealed that the Pi2/9 gene family members are grouped into the same phylogenetic clade (Fig. 2). A similar phylogenetic tree was generated when the entire CDSs of the family members were used, further suggesting that the Pi2/9 gene family members belong to the same subfamily (Fig. S3). We could not identify obvious sequence blocks of the genes from different cladistic groups sharing significant sequence homology, indicating that the possibility of sequence recombination between different group members is low. Five genes, i.e., Nbs4-, Nbs6-, Nbs7-, Nbs9-, and Nbs12-OM-CC, were not included in the cladistic analysis because they have no NBS region (due to sequence deletion in Nbs7- and Nbs9-OM-CC) or because the NBS-encoded region was not sequenced (Nbs4-, Nbs6-, and Nbs12-OM-CC). We positioned these genes in their corresponding clades, however, based on the nucleotide sequence similarity of the remaining portion of these genes (non-NBS region) with that of other Pi2/9 gene family members (Fig. 2; Table S2). For example, both Nbs7- and Nbs9-OM-CC share overall 98% nucleotide sequence identity to Nbs3- and Nbs5-OM-CC in the non-NBS region (Table S2).

Dynamics of the *Pi2/9* gene family members

As illustrated in Fig. 2, most of the cladistic groups contain the *Pi2/9* gene family members from different *Oryza* species, suggesting that some of the genes are more sequence related to their putative orthologes in different species than to those within the same *Pi2/9* cluster. We thus classified the *Pi2/9* gene family members into eight vertical gene (VG) groups, designated VG1–VG8, based on their cladistic relationship (Fig. 2; Table 2). Among the VGs, only VG2 and VG6 contain a single gene, whereas the other six groups contain at least three genes. VG4 and VG7 contain multiple members from a particular species, suggesting that

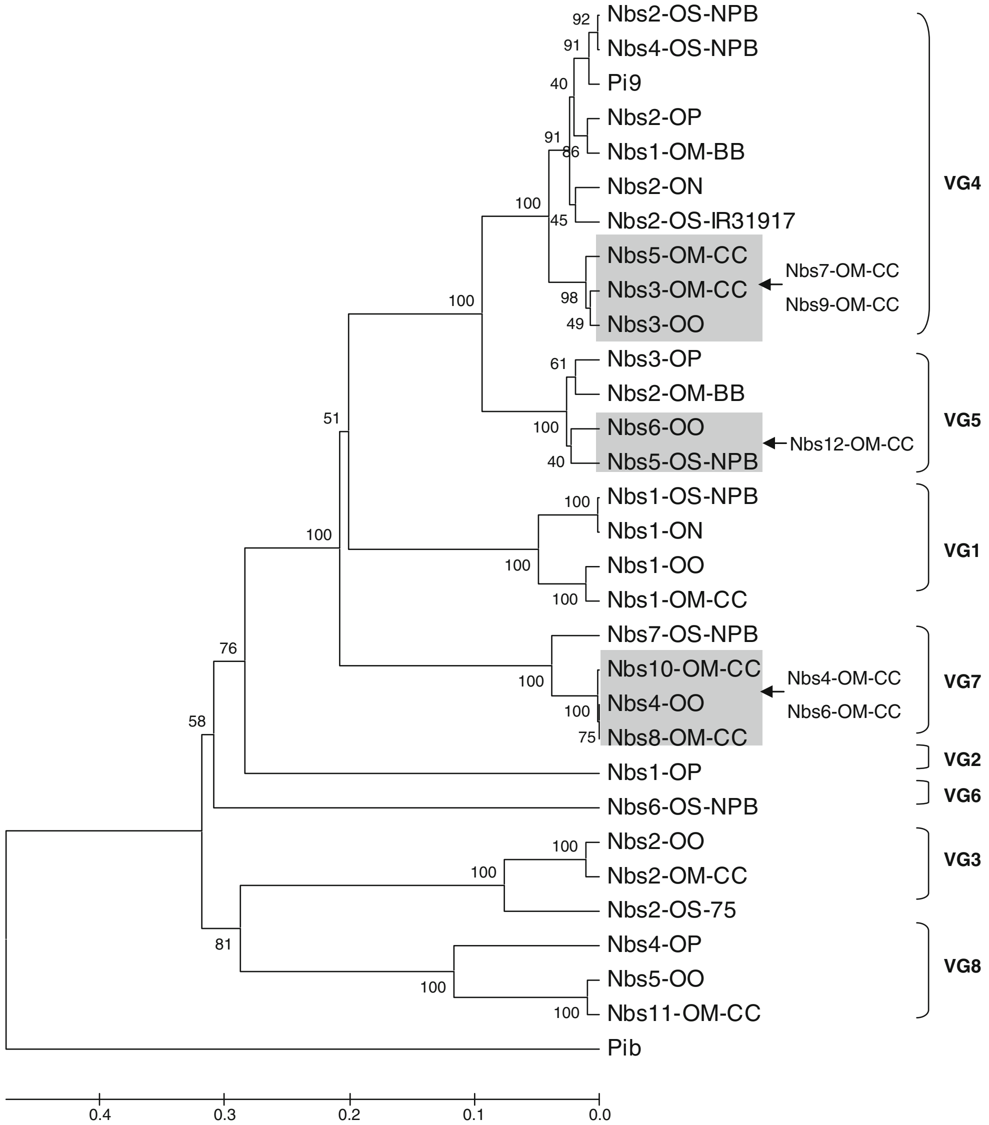

Fig. 2 Cladistic analysis of the *Pi2/9* gene family members in the *Oryza* genus. The cladistic tree was constructed based on the sequences of the NBS domains of the *Pi2/9* gene family members in the four wild rice species as well as those in cultivated rice lines. Clustalx and MEGA 4 programs were used for multiple sequence alignment and for tree viewer, respectively. The five genes without the NBS domain (with an *arrow* at their *left side*) are manually positioned into their respective clades that are *shadowed*

these genes might arise from gene duplication events within the species. It was worthwhile to note that the classification of VG3 and VG8 are not as stringent as other groups because Nbs2-OS-75 in VG3 and Nbs4-OP in VG8 are more sequence divergent to their putative orthologs in respective groups. However, significant sequence homology (over 84% sequence identity on average, Table S2) and high-bootstrap value (100 for both groups, Fig. 2) in the cladistic tree suggest that the classification should be reasonable. Interestingly, all three cloned R-genes, *Pi2, Pi9* and *Piz-t*, situated in VG4 group (*Pi2* and *Piz-t* are not included in Fig. 2 as they are highly homologous, Zhou et al. 2007). Whether other members in the group are also functional requires further investigation.

Sequence similarity in both coding and noncoding regions of the genes within each vertical group was further analyzed. Only the first intron was used as the noncoding region for sequence analysis because most the *Pi2/9* gene family members only contain this single intron. When compared with the coding sequences, the noncoding sequences, however, are less conserved in the family members within each group. For example, the sequence similarity level in VG4 is 77% in the noncoding region when compared with 90% in the coding region (Table S2). Two exceptions were found in VG4 and VG8. *Nbs4-OS-NPB* in VG4 and *Nbs4-OP* in VG8 share as low as 88 and 84% DNA identity in the coding regions, but have no significant sequence similarity in the noncoding regions to their respective group members

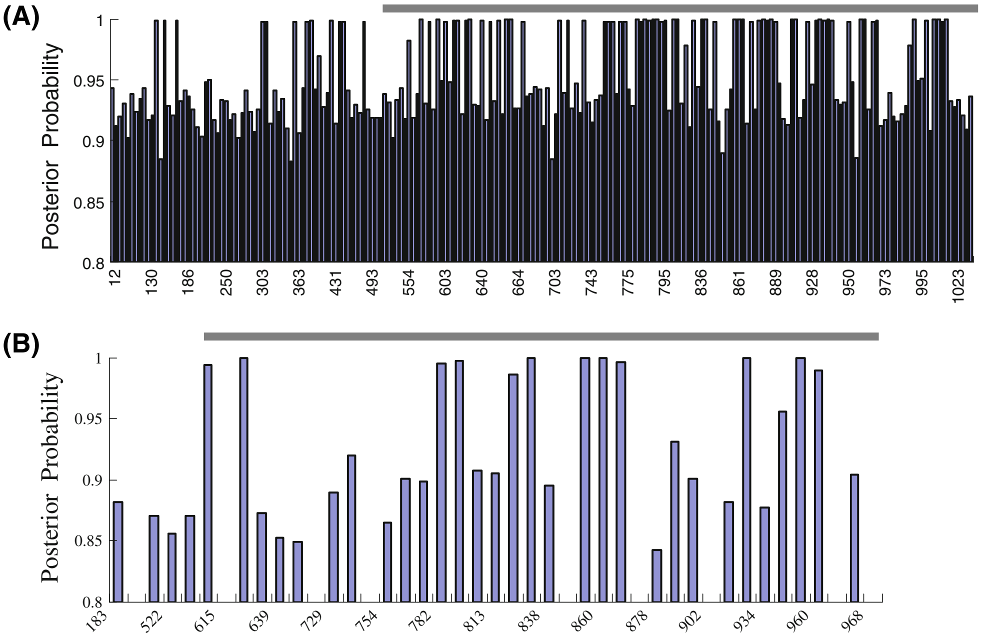

Fig. 3 Diversifying selection on the LRR region of the *Pi2/9* gene family members in both VG4 (**a**) and VG7 (**b**) groups. Posterior probabilities (*Y* axis) for nucleotide sites (*X* axis) under the selection model M3 along the *Pi2/9* gene family members are shown under a *horizontal bar*. The LRR region is indicated in *filled boxes* under the *bar*

(Table S2). These results suggest that the noncoding regions appear to undergo more sequence divergence than the coding regions in the *Pi2/9* gene family members in wild rice species. In addition, we found that the genomic location of the vertical groups at the *Pi2/9* locus is not always conserved among the cultivated and wild rice species. For example, the genomic order of the vertical groups is VG2–VG4–VG5–VG8 in *O. punctata*, VG1–VG3–VG4–VG7–VG8–VG5 in *O. officinalis,* and VG1–VG4–VG5–VG6–VG7 in *O. sativa* (Fig. 1), suggesting that sequence duplication and subsequent inversion or uneven crossing over may have occurred after speciation.

Positive selection on the *Pi2/9* gene family members

Positive selection plays a significant role in the evolution of the *Pi2/9* gene family members in cultivated rice (Zhou et al. 2007). To investigate whether positive selection is also important for evolution of the *Pi2/9* gene family members in wild rice species, we employed two pairs of maximum likelihood models of codon substitution, M3/M0 and M8/M7, in PAML programs (Yang 1997, 2000). The goal was to identify the sites subject to an elevated ratio of the rate of nonsynonmous substitution (K_a) to the rate of synonymous substitution (K_s). VG2 and VG6 were not included in the analysis because both contain only a single family member. Some gene members in both VG3 and VG8 are less conserved, making it difficult for multiple sequence alignment. Thus, the four remaining vertical groups (VG1, VG4, VG5, and VG7) were included for the sequence analysis. Based on the site number and the likelihood ratio tests (LRT), the discrete model M3 fits the data significantly better than other models (Table S3). No sites with positive selection with greater than 99% confidence were identified in both VG1 and VG5 groups (Table S3). In contrast, 68 sites in VG4 and 10 sites in VG7 have been subjected to positive selection (with greater than 99% confidence). It is noteworthy that 58 out of 68 sites in VG4 and all 10 sites in VG7 were localized in the region encoding the LRR domain (Fig. 3; Table S3).

To identify coding regions with nucleotide divergence for novel amino acids, we conducted a sliding-window analysis that showed the distribution of nonsynonymous substitutions along the aligned sequences within the four vertical groups. As shown in Fig. 4 and despite a few sites exhibiting a high level of nucleotide diversity (π), an even distribution pattern along the coding sequence of the genes in both VG1 and VG5 was observed. In contrast, the gene members in both VG4 and VG7 have more frequent nonsynonymous substitution sites at their 3′ termini where the LRR domain is located. The disproportionate number of sites in the LRR-encoded sites in both VG4 and VG7 suggests that a strong evolutionary selection acted on the gene members in these two groups.

Divergence of the *Pi2/9* locus in the AA, BB, and CC genomes of wild species

We found that the *Pi2/9* loci in different species with the same genome constitution share a similar level of sequence

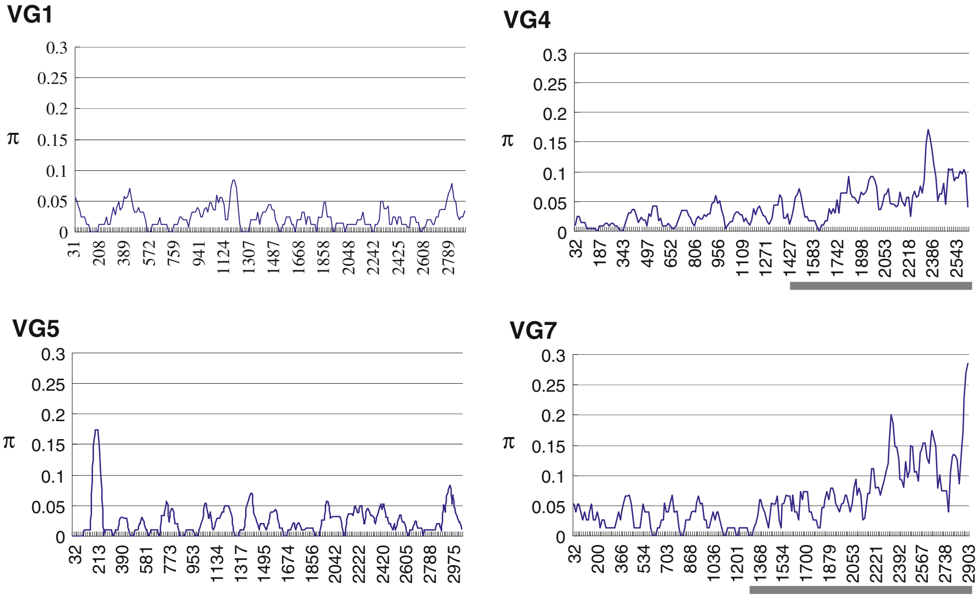

Fig. 4 Distribution of non-synonymous substitutions in the coding sequence of the *Pi2/9* gene family members in VG1, VG4, VG5, and VG7. The coding sequences of the gene members in each vertical group were used for the detection of the nonsynonymous substitutions using the DnaSP4.02 program. *Solid line* indicates the estimated nucleotide diversity (π) among the gene members within each VG. The nucleotide sites are numbered in *X* axis, and the LRR region in both VG4 and VG7 are indicated under the nucleotide sites

similarity in both genic and intergenic regions, which allowed us to compare the sequence diversity between orthologous regions in four wild *Oryza* species, and one cultivated rice line representing three different genomes (AA, BB, and CC). The results are described in the following sections.

Makeup and dynamics of TEs

Across the four wild *Oryza* species, a total of 25 TEs were identified, including 9 intact LTRs retrotransposons, four intact transposons, and 12 solo-LTRs (Table 3). Solo-LTRs are byproducts of unequal homologous recombination between paired LTRs from the same or related retrotransposons (Vitte and Panaud 2003). The designation, genomic location, size, target duplicated sequence (TDS), and features of these TEs are listed in Table 3. Except for two of the TEs (TE1-OP/TE2-OM-BB and TE2-OO/TE4-OM-CC), all the TEs are only present in one species (Table 3). Across the AA, BB, and CC genomes, we were unable to identify even a single orthologous TE. Moreover, most of these TEs have typical TDSs, the conserved signatures for de novo insertion of TEs (Table 3). These results suggest that the TEs might be newly inserted sequences in the *Oryza* genus, i.e., these sequences appeared to have evolved after the formation of different subgenomes. The insertion time of each intact LTR-retrotransposon was estimated and ranges from 0.35 to 11.2 million years ago (MYA) (Table 3). On an average, TEs constitute about 19% of the genomic sequence of the *Pi2/9* locus in *O. nivara, O. officinalis, O. punctata,* and the CC subgenome of *O. minuta*. However, 55% of the genomic sequence of the *Pi2/9* locus in the BB subgenome of *O. minuta* is composed of TEs. Pairwise comparison of the *Pi2/9* locus between *O. punctata* and *O. minuta* revealed that the genome expansion of the BB subgenome in the two wild species is mainly resulted from the insertion of the four types of transposons (Fig. S4).

Gene duplication and loss

Duplication and loss of the *Pi2/9* gene family members within each genome was observed at the *Pi2/9* locus. The *Pi2/9* locus from *O. nivara* contains only two *Pi2/9* gene family members whereas the cultivated rice Nipponbare contains seven members. Given that *Nbs5-OS-NPB* and *Nbs7-OS-NPB* have their respective orthologs in other genomes, we believed that at least these two gene members were lost in the analyzed accession of *O. nivara*. The absence of the *Nbs1-OP* ortholog in the BB subgenome of *O. minuta* is another case of gene loss at the *Pi2/9* locus. Interestingly, the remnants of *TE1-OP*, including one LTR, were identified in the orthologous region of the BB subgenome of *O. minuta* (Fig. S3), indicating that the loss of the *Nbs1-OP* ortholog might be associated with the loss of the *TE1-OP* element.

In contrast to the gene loss identified in *O. nivara* and in the BB subgenome in *O. minuta*, a gene expansion was identified in the CC subgenome of *O. minuta*. When compared with *O. officinalis*, the *Pi2/9* locus in the CC

Table 3 Feature of the transposable elements (TEs) at the *Pi2/9* locus in wild rice species

TEs	Location	Size	TDS*	Annotation	Divergence Time (MYA)	Genome constitution (%)
O. nivara						
TE1-ON	37,465–47,711	10,246	CCTAC	Ty3-gypsy subclass –LTR	0.35	19.1
TE2-ON	78,281–81,135	2,854	ACGCC	Solo-LTR		
TE3-ON	81,329–91,149	9,820	ND**	Ty3-gypsy subclass-LTR	6.49	
O. punctata						
TE1-OP	9,140–17,767	8,627	ND	Ty1-copia subclass-LTR	11.2	19.9
TE2-OP	34,481–36,539	2,059	ND	Uncharacterized repeat		
TE3-OP	39,362–41,742	2,380	CAT	Solo-LTR		
TE4-OP	68,253–73,831	5,578	GCGGG	Solo-LTR		
TE5-OP	81,960–85,051	3,091	CTTAG	Solo-LTR		
O. minuta-BB						
TE1-OM-BB	6,333–17,199	10,867	GTTGT	Ty3-gypsy subclass-LTR	7.83	55
TE2-OM-BB	17,700–18,893	1,193	ND	Solo-LTR, orthologous to the one of TE1-OP		
TE3-OM-BB	19,269–25,660	6,391	GTCAG	Ty1-copia subclass-LTR	4.69	
TE4-OM-BB	30,275–42,647	12,372	CCTAG	Unclassified retrotransposon		
TE5-OM-BB	34,726–39,781	Internal TE	ATCTC	Ty1-copia subclass-LTR	2.57	
TE6-OM-BB	49,148–52,275	3,127	CGC	CACTA, En/Spm sub-class transposon		
TE7-OM-BB	53,939–67,233	13,294	CCCGC/TCCGC	Unclassified retrotransposon	1.56	
TE8-OM-BB	74,513–75,731	1,218	ATCTG	Solo-LTR		
O. officinalis						
TE1-OO	14,492–21,754	7,262	ND	Copia-like retrotransposon	4.82	19.3
TE2-OO	48,998–61,621	12,623	TTTAC/TTTAT	Solo-LTR inserted with a retrotransposon		
TE3-OO	49,403–60,762	Internal TE	CGTAT	Ty3-gypsy subclass-LTR	0.97	
TE4-OO	83,599–87,699	ND	TTGTT	Solo-LTR, spanning the sequence gap		
TE5-OO	120,062–128,769	8,707	CTG	En/Spm Sub-class Transposon		
O. minuta-CC						
TE1-OM-CC	15,579–20,963	5,384	ATTAG	Solo-LTR		18.2
TE2-OM-CC	23,375–26,904	3,529	ND	CACTA, En/Spm sub-class transposon		
TE3-OM-CC	28,878–32,129	3,251	ND	CACTA, En/Spm sub-class transposon		
TE4-OM-CC	49,689–50,924	1,235	TTTAC/TTTAT	Solo-LTR, orthologous to TE2-OO		
TE5-OM-CC	85,131–90,324	5,193	ND	Solo-LTR		

TDS target duplicated sequence, *ND* not determined

The TEs in the *Pi2/9* locus in *O. sativa* were not included in the analysis

subgenome of *O. minuta* contains four copies of homologues of *Nbs3*- and *Nbs4-OO* (Fig. 1). *Nbs3-OO* shares overall 96% DNA similarity to *Nbs3-, Nbs5-, Nbs7-,* and *Nbs9-OM-CC,* and *Nbs4-OO* shares overall 94% DNA similarity to *Nbs4-, Nbs6-, Nbs8*, and *Nbs10-OM-CC*, which are slightly lower than the levels of respective paralogues (over 98% similarity, Table S2). Moreover, the gene order of the homologues of *Nbs3*- and *Nbs4-OO* is conserved in *O. minuta*, suggesting that this gene expansion likely arose from tandem duplication of a genomic fragment comprising *Nbs3*- and *Nbs4-OO* homologues. Comparison of the intergenic sequence between *Nbs3-OO* and *Nbs4-OO* with the sequence of the CC genome of *O. minuta* revealed that the intergenic sequence between *Nbs3*- and *Nbs4-OM-CC* is more related to the one between *Nbs3*- and *Nbs4-OO* than to those between their respective paralogues in *O. minuta*. Moreover, we identified an 870-bp deletion in the NBS-encoded region in *Nbs7*- and *Nbs9-OM-CC*, but not in *Nbs3*- and *Nbs5-OM-CC*. Taken together these results suggest that *Nbs3*- and *Nbs4-OM-CC* are likely orthologs of *Nbs3*- and *Nbs4-OO*, respectively. *Nbs5-/Nbs6-, Nbs7-/Nbs8-,* and *Nbs9-/Nbs10-OM-CC* were likely generated by tandem duplication events afterwards.

The *Pi9* gene is more related to its homologues in *O. sativa* than to those in its donor species *O. minuta*

The *Pi9* gene was introgressed from *O. minuta* (accession # 101141) into the isogenic line 75-1-127 through embryo rescue and repeated backcrosses with the recurrent cultivar IR31917-45-3-2, abbreviated as IR31917 herein (Amante-Bordeos et al. 1992). The same *O. minuta* accession was used for the construction of the *O. minuta* BAC library (http://www.genome.arizona.edu/BAC_special_projects/#Rice). Therefore, the sequence of the *Pi9* locus in both BB and CC subgenome of *O. minuta* discussed above should represent the original sequence of the *Pi9* region in the donor species. Sequence comparison between the *Pi9* gene in the introgression line and its homologues in the donor and recurrent lines should be useful for confirming the origin of the *Pi9* gene at the molecular level. Based on the sequences of the *Pi9* gene and its homologues in CO39 and Nipponbare, we amplified and sequenced the *Pi9* homologous fragment in IR31917, designated *Nbs2-OS-IR31917* (Fig. S5). Unfortunately, the 3′ terminus of *Nbs2-OS-IR31917* was not amplified, probably because of its sequence divergence from other sequenced cultivars. The cloned fragment from IR31917 included a 2,648-bp coding sequence representing the majority portion of *Nbs2-OS-IR31917*s CDS (~90%), which should be sufficiently informative for sequence comparison (Fig. S4).

As illustrated in Fig. 2, the *Pi9* gene belongs to VG4, suggesting that it likely originated from its homologues in this group. Indeed, *Nbs2-OS-IR31917*, *Nbs1-OM-BB* in the BB subgenome of *O. minuta*, and *Nbs3-, Nbs5-, Nbs7-,* and *Nbs9-OM-CC* in the CC subgenome of *O. minuta* are located in the same group. To eliminate the bias in the cladistic analysis due to sequence gaps in some sequences, i.e., a fragment deletion in both *Nbs7-OM-CC* and *Nbs9-OM-CC* and an undetermined 3′ portion of *Nbs2-OS-IR31917*, we split the sequences into four fragments, corresponding to 1–554, 555–1,429, 1,430–2,648, and 2,649–3,099 bp in the *Pi9* sequence. These four portions of the *Pi9* gene and its homologues were then designated as 5′, intermediate 1 (IM-1), intermediate 2 (IM-2), and 3′ portion as listed in Table 4. Among these four portions, the IM-2 portion does not comprise both *Nbs7*- and *Nbs9-OM-CC* and the 3′ portion does not comprise *Nbs2-OS-IR31917*. Four cladistic analyses were then conducted based on the nucleotide sequences of these four portions. Interestingly, the *Pi9* sequence in the 5′ portion is almost identical to its homologues in three different cultivars. These sequences are different from other homologues and form a single clade (Fig. 5a). In both the IM-1 and IM-2 portions, the *Pi9* sequence is more related to *Nbs2-OS-CO39* than to other homologues (Fig. 5b, c). In the 3′ portion, the *Pi9* sequence shows relatively higher sequence divergence and forms an individual clade from other homologues (Fig. 5d). However, it is still embedded within the clades with its homologues from the AA genome species. Pairwise distances of gene pairs further indicated that the *Pi9* gene has the lowest distance to its homologues in cultivated rice lines (Table 4). It seems that the most *Pi9*-related homologue is either *Nbs2-OS-NPB* or *Nbs2-OS-CO39* rather than *Nbs2-OS-IR31917* based on the pairwise distance. In addition, the noncoding sequences of the *Pi9* gene, including 5′ and 3′ UTRs and introns, are almost identical to those in the cultivated rice lines (data not shown). Moreover, we were unable to detect any similar sequence blocks in *Nbs1-OM-BB* or *Nbs3-, Nbs5-, Nbs7-,* or *Nbs9-OM-CC*. Based on these results, we conclude that the *Pi9* gene in the introgression line 75-1-127 was not transferred from *O. minuta* and might have originated from an unknown *O. sativa* cultivar.

Discussion

In the last decade, many R-gene loci have been cloned and sequenced, and their evolutionary origins have been

Table 4 The pairwise distances of the different portions of the Pi9 gene to its homologs in VG4

	5′ portion (1–554 bp)	IM-1 portion (555–1,429 bp)	IM-2 portion (1,430–2,648 bp)	3′ portion (2,649–3,099 bp)
Nbs2-OS-CO39	0.0,000	0.0023	0.0169	0.0370
Nbs2-OS-NPB	0.0,000	0.0057	0.0255	0.0195
Nbs2-OS-IR31917	0.0,000	0.0304	0.0273	NA
Nbs2-ON	0.0,470	0.0340	0.0229	0.0261
Nbs2-OP	0.0471	0.0280	0.0566	0.0639
Nbs1-OM-BB	0.0471	0.0304	0.0650	0.0754
Nbs3-OO	0.0508	0.0497	0.0984	0.0536
Nbs3-OM-CC	0.0508	0.0522	0.1025	0.0730
Nbs5-OM-CC	0.0508	0.0608	0.1025	0.0730
Nbs7-OM-CC	0.0528	NA	0.0944	0.0896
Nbs9-OM-CC	0.0528	NA	0.0963	0.0876

The distance was calculated using maximum composite likelihood model in Mega 4 program

NA not applicable due to sequence gap, *IM* intermediate

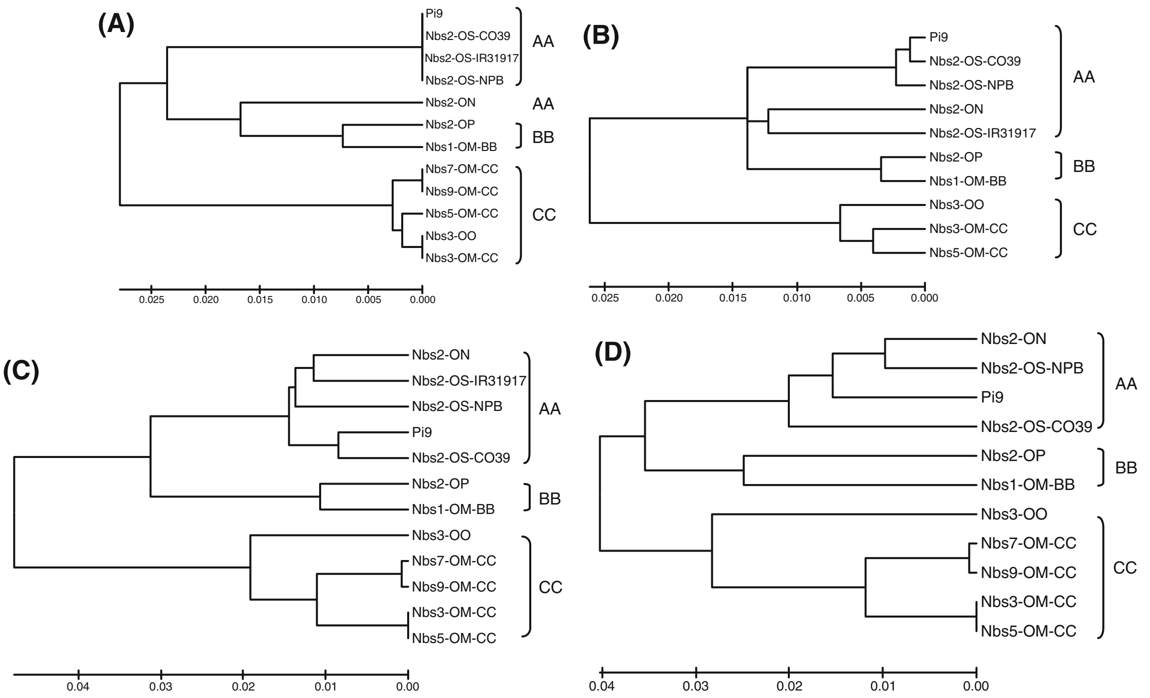

Fig. 5 Cladistic analysis of the *Pi9* gene and its homologs in different wild rice species and cultivated rice lines. The cladistic trees were constructed based on the four portions of the coding sequence of the *Pi9* gene and its homologs: **a** the 5′ portion (1–554 bp), **b** the intermediate portion 1 (555–1,429 bp), **c** the intermediate portion 2 (1,430–2,648 bp), and **d** the 3′ portion (2,649–3,099 bp). The sequence position is calculated based on the *Pi9* sequence. The genome constitution of the genes is listed at the *right*. Clustalx and MEGA 4 programs were used for multiple sequence alignment and for tree viewer, respectively

analyzed (Hulbert et al. 2001; Martin et al. 2003). Only a few of loci, however, have been analyzed across different species. Comparative analyses of the complete genomic sequences of the *Cf0*, *Cf4*, *Cf9*, and *9CD* cluster in different tomato species revealed that inter- and intragenic sequence exchange has a major impact on the evolution of the *Cf4/9* complex locus (Parniske and Jones 1999; Kruijt et al. 2004). The *RPW8* locus is another R-complex locus completely sequenced in *Arabidopsis thaliana*, *A. lyrata*, and *Brassica rapa* (Xiao et al. 2004). Comparison of the genomic sequences of the *RPW8* loci in these three species revealed that *RPW8* has evolved from recent gene duplication and subsequent functional diversification favored by diversifying selection. Recently, comparison of the *Pm3* complex loci among three different wheat species indicated an extremely dynamic evolution of the plant R-complex loci (Wicker et al. 2007); multiple sequence rearrangements in the *Pm3*-like genes and their up- and downstream regions facilitate the generation of a complex mosaic of conserved and unique sequences at the *Pm3* locus in wheat (Wicker et al. 2007). The authors further compared the loci harboring the *Pm3*-like gene cluster in both *indica* and *japonica* subspecies of *O. sativa*. They found that the *Pm3*-like gene cluster on chromosome 3 is highly conserved over a large distance between two subspecies, displaying contrasting rates of evolution of rice *Pm3*-like locus from its counterpart in wheat (Wicker et al. 2007).

Although 13 rice blast R-genes have been cloned to date, how these genes originated and evolved in the *Oryza* genus remains unclear. Huang et al. (2008) recently analyzed the molecular evolution of the *Pi-ta* locus in 36 accessions of wild rice *Oryza rufipogon* (Huang et al. 2008). The LRR domain has a high nonsynonymous/synonymous ratio and the amino acid Ala-918 in the LRR domain has a close relationship with the resistant phenotype. Wang et al. (2008) also investigated the haplotype diversity at the *Pi-ta* locus in 51 accessions of cultivated rice and six wild relatives obtained similar results (Wang et al. 2008). Yoshida and Miyashita (2009) studied the intra- and interspecific DNA variations in the blast resistance gene *Pita* in wild rice (*O. rufipogon*), cultivated rice (*O. sativa*), and two other related wild rice species (*O. meridionalis* and *O. officinalis*). They found that the resistance *Pita* allele had lower levels of variation than the susceptibility *pita* allele. Recently, Lee et al. (2009) surveyed the sequences of the *Pi-ta* locus and its flanking regions in 159 accessions composed of seven AA genome *Oryza* species. A total of 33 new *Pi-ta* haplotypes and 18 new Pi-ta protein variants were identified, suggesting the *Pi-ta* gene has evolved under an extensive selection pressure during crop breeding.

We generated the genomic sequences of the *Pi2/9* locus in four wild *Oryza* species representing three different genomes, i.e., AA, BB, and CC. The sequences of the entire *Pi9* locus in the four wild species spanning nearly 100 kb were revealed and compared. This genus-wide sequence data-set, as well as the previously established reference sequence in cultivated rice species (Zhou et al. 2007) has provided the most extensive and unique genomic sequence layout so far for an R-gene complex cluster across the *Oryza* genus. The comparative analysis of the *Pi2/9* locus in wild rice species in this study permits us to dissect the genomic dynamics and evolutionary mechanism of the broad-spectrum *R* locus across the *Oryza* genus.

Both unequal recombination at intra- and intergenic levels and tandem duplication have been proposed as major genetic events contributing to the generation of a tandem array of *R*-genes (Leister 2004). Unequal recombination is likely to occur if homologues are highly sequence related, which in turn will result in their homogenization, a phenomenon known as concerted evolution (Walsh 1987; Hichey et al. 1991). The frequency of unequal recombination varies significantly among different R complexes in plants. For example, it is quite common at the maize *Rp1* and *Rp3* loci but rare at the *Dm3* locus of lettuce (Meyers et al. 1998b; Sun et al. 2001; Webb et al. 2002). The *Pi2/9* locus in different wild *Oryza* species is composed of a variable number of NBS-LRR genes, which can be clearly divided into eight VGs (Fig. 1; Table 2). The genes within each VG from different species are more similar to each other than to their homologues within the same species, suggesting that unequal recombination has not much impact on the evolution of the *Pi2/9* locus. Divergence in sequence and organization of the *Pi2/9* gene family members by sequence mutation (point mutation, insertion and deletion), gene translocation and inversion, and insertion of TEs has contributed greatly to the variation of the *Pi2/9* locus at the gene and locus levels. Divergence of the *Pi2/9* locus at the intra- and inter-species level facilitates the independent evolution of each VG within a species.

Many NBS-LRR-type *R*-genes are subjected to positive selection, and the sites are predominantly situated in the LRR region, e.g., *RPS2*, *RPP5*, and *RPP8* loci in *Arabidopsis* and the *L* and *P* loci in flax (McDowell et al. 1998; Ellis et al. 1999; Noel et al. 1999; Dodds et al. 2001; Maricio et al. 2003). At the *Pi2/9* locus in cultivated rice, positive selection was also detected (Zhou et al. 2007). Among the wild species, we found that VG4 and VG7, but not VG1 and VG5 are subjected to positive selection, suggesting that the VGs have been driven differentially by evolutionary forces. Intriguingly, VG4 contains all three functional *R*-genes at this locus, i.e., *Pi9*, *Piz-t*, and *Pi2* (Fig. 2, *Pi2* and *Piz-t* were not included in the figure, because they are highly homologous to *Pi9*, Zhou et al. 2007). We also found that VG4 and VG7 but not VG1 and VG5 exhibit more frequent nonsynonymous substitution sites at their LRR region (Fig. 4). Therefore, we believe that positive selection has played an important role in favoring the generation of novel resistance specificity at the *Pi2/9* locus. Given that VG7 displays a similar pattern with respect to both positive selection and frequency of nonsynonymous substitution sites at the LRR region, we speculate that some of the genes in VG7 may confer novel resistance specificities to rice blast.

Recently, Ammiraju et al. (2008) and Lu et al. (2009) completed comparative sequence analyses of the Adh1–Adh2 and MOC1 loci, respectively, providing the most comprehensive vertical sequence data-sets for revealing the architectural complexities and dynamic evolution of these two loci across the *Oryza* genus. These studies found that TEs are the major driving force behind the expansion of the *Oryza* genomes. In the current study, we identified 25 TEs at the *Pi2/9* locus in the wild species. Most of the TEs are species specific, and some of them are different even between the same genomes in diploid and tetraploid species (BB/BB-OM and CC/CC-OM as listed in Table 3). In this case, the difference in TE content and size of intergenic space at the *Pi2/9* locus indicates that insertions and deletions of TEs contributed to reshaping the *Pi2/9* locus after speciation.

Acknowledgments This project was supported by the “973” Project (2006CB101904), the “948” Project (2006-G61), the Henye Project of Ministry of Agriculture, National Natural Science Foundation of China (30828022), and the NSF-Plant Genome Research Program (#0605017).

References

Amante-Bordeos A, Sitch LA, Nelson R, Damacio RD, Oliva L, Aswidinnoor H, Leung H (1992) Transfer of bacterial blight and blast resistance from the tetraploid wild rice *Oryza minuta* to cultivated rice, *Oryza sativa*. Theor Appl Genet 84:345–354

Ameline-Torregrosa CB, Wang B, O’Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB (2008) Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. Plant Physiol 146:5–21

Ammiraju JSS, Lu F, Sanyal A, Yu Y, Song X, Jiang J, Pontaroli AC, Rambo T, Currie J, Collura K, Talag J, Fan C, Goicoechea JL, Zuccolo A, Chen J, Bennetzen JL, Chen M, Jackson S, Wing RA (2008) Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. Plant Cell 20:3191–3209

Bai JL, Pennill A, Ning J, Lee SW, Ramalingam J, Webb CA, Zhao B, Sun Q, Nelson JC, Leach JE, Hulbert SH (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res 12:1871–1884

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

Caracuel-Rios Z, Talbot NJ (2007) Cellular differentiation and host invasion by the rice blast fungus *Magnaporthe grisea*. Curr Opin Microbiol 10:339–345

Dodds PN, Lawrence GJ, Ellis JG (2001) Six amino acid changes confined to the leucine-rich repeat β-strand/β-turn motif determine the difference between the *P* and *P2* rust resistance specificities in flax. Plant Cell 13:163–178

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

Ewing B, Green P (1998) Base calling of automated sequencer traces using Phred II. Error probabilities. Genome Res 8:186–194

Ewing B, Hillier L, Wendl M, Green P (1998) Base calling of automated sequencer traces using Phred I. Accuracy assessment. Genome Res 8:175–185

Fukuoka S, Saka N, Koga H, Ono K, Shimizu T, Ebana K, Hayashi N, Takahashi H, Hirochika A, Okuno K, Yano M (2009) Loss of function of a proline-containing protein confers durable disease resistance in rice. Science 325:998–1001

Geffroy V, Macadre C, David P, Pedrosa-Harand A, Sevignac M, Dauga C, Langin T (2009) Molecular analysis of a large subtelomeric nucleotide-binding-site-leucine-rich-repeat family in two representative genotypes of the major gene pools of *Phaseolus vulgaris*. Genetics 181:405–419

Gordon D, Abajian C, Green P (1998) Consed, a graphical tool for sequence finishing. Genome Res 8:195–202

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

Hichey DA, Bally-Cuif L, Abukashawa S, Payant V, Benkel BF (1991) Concerted evolution of duplicated protein-coding genes in *Drosophila*. Proc Natl Acad Sci USA 88:1611–1615

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

Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol 39:285–312

Kruijt M, Brandwagt BF, De Wit PJGM (2004) Rearrangements in the *Cf-9* disease resistance gene cluster of wild tomato have resulted in three genes that mediate Avr9 responsiveness. Genetics 168:1655–1663

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

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

Leister D (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. Trends Genet 20:116–122

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

Lu F, Ammiraju JS, Sanyal A, Zhang S, Song R, Chen J, Li G, Sui Y, Song X, Cheng Z, de Oliveira AC, Bennetzen JL, Jackson SA, Wing RA, Chen M (2009) Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. Proc Natl Acad Sci USA 106:2071–2076

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

Maricio R, Stahl EA, Korves T (2003) Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. Genetics 163:735–746

Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol 54:23–61

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

McHale L, Tan X, Koehl P, Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. Genome Biol 7:212

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998a) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817–1832

Meyers BC, Shen KA, Rohani P, Gaut BS, Michelmore RW (1998b) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell 10:1833–1846

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

Monosi B, Wisser RJ, Pennill L, Hulbert SH (2004) Full-genome analysis of resistance gene homologues in rice. Theor Appl Genet 109:1434–1447

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

Parniske M, Jones JD (1999) Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. Proc Natl Acad Sci USA 96:5850–5855

Perriere G, Gouy M (1996) WWW-Query: an on-line retrieval system for biological sequence banks. Biochimie 78:364–369

Qu S, Liu G, Zhou B, Bellizzi M, Zeng L, Dai L, Han H, Wang GL (2006) The broad-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

Shang J, Tao Y, Chen X, Zou Y, Lei C, Wang J, Li X, Zhao X, Zhang M, Lu Z, Xu J, Cheng Z, Wan J, Zhu L (2009) Identification of a new rice blast resistance gene, Pid3, by genomewide comparison of paired nucleotide-binding site-leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. Genetics 182:1303–1311

Shimamoto K, Kyozuka J (2002) Rice as a model for comparative genomics of plants. Annu Rev Plant Biol 53:399–419

Song WY, Pi LY, Wang GL, Gardner J, Holsten T, Ronald P (1997) Evolution of the rice *Xa21* disease resistance gene family. Plant Cell 9:1279–1287

Sun Q, Collins NC, Ayliffe M, Smith SM, Drake J, Pryor T, Hulbert SH (2001) Recombination between paralogues at the rp1 rust resistance locus in maize. Genetics 158:423–438

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acids Res 25:4876–4882

Valent B (1990) Rice blast as a model system for plant pathology. Phytopathology 80:33–36

Vitte C, Panaud O (2003) Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. Mol Biol Evol 20:528–540

Walsh JB (1987) Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? Genetics 117:543–557

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

Webb CA, Richter TE, Collins NC, Nicolas M, Trick HN, Pryor T, Hulbert SH (2002) Genetic and molecular characterization of the maize *rp3* rust resistance locus. Genetics 162:381–394

Wei F, Wing RA, Roger PW (2002) Genome dynamics and evolution of the *Mla* (*Powdery mildew*) resistance locus in barley. Plant Cell 14:1903–1917

Wicker T, Yahiaoui N, Keller B (2007) Contrasting rates of evolution in *Pm3* loci from three wheat species and rice. Genetics 177:1207–1216

Wing RA, Ammiraju JS, Luo M, Kim H, Yu Y, Kudrna D, Zuccolo A, Ammiraju JS, Luo M, Nelson W, Ma J, SanMiguel P, Hurwitz B, Ware D, Brar D, Mackill D, Soderlund C, Stein L, Jackson S (2005) The oryza map alignment project: the golden path to unlocking the genetic potential of wild rice species. Plant Mol Biol 59:53–62

Wisser RJ, Qi S, Hulbert SH, Kresovich S, Nelson RJ (2005) Identification and characterization of regions of the rice genome associated with broad-spectrum, quantitative disease resistance. Genetics 169:2277–2293

Xiao S, Emerson B, Ratanasut K, Patrick E, O'Neill C, Bancroft I, Turner JG (2004) Origin and maintenance of a broad-spectrum disease resistance locus in *Arabidopsis*. Mol Biol Evol 21:1661–1672

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

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

Yoshida K, Miyashita NT (2009) DNA polymorphism in the blast disease resistance gene Pita of the wild rice *Oryza rufipogon* and its related species. Genes Genet Syst 84(2):121–136

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