DOI: 10.1534/genetics.108.089805

Molecular Evolution of the *Pi-ta* Gene Resistant to Rice Blast in Wild Rice (*Oryza rufipogon*)

Chun-Lin Huang,*,† Shih-Ying Hwang,‡ Yu-Chung Chiang§ and Tsan-Piao Lin*,[1]

*Institute of Plant Biology, National Taiwan University, Taipei 106, Taiwan, †Department of Botany, National Museum of Natural Science, Taichung 404, Taiwan, ‡Department of Life Science, National Taiwan Normal University, Taipei 116, Taiwan and §Department of Life Science, Pingtung University of Science and Technology, Pingtung 912, Taiwan

Manuscript received March 31, 2008
Accepted for publication May 14, 2008

ABSTRACT

Rice blast disease resistance to the fungal pathogen *Magnaporthe grisea* is triggered by a physical interaction between the protein products of the host *R* (resistance) gene, *Pi-ta,* and the pathogen *Avr* (avirulence) gene, *AVR-pita.* The genotype variation and resistant/susceptible phenotype at the *Pi-ta* locus of wild rice (*Oryza rufipogon*), the ancestor of cultivated rice (*O. sativa*), was surveyed in 36 locations worldwide to study the molecular evolution and functional adaptation of the *Pi-ta* gene. The low nucleotide polymorphism of the *Pi-ta* gene of *O. rufipogon* was similar to that of *O. sativa*, but greatly differed from what has been reported for other *O. rufipogon* genes. The haplotypes can be subdivided into two divergent haplogroups named H1 and H2. H1 is derived from H2, with nearly no variation and at a low frequency. H2 is common and is the ancestral form. The leucine-rich repeat (LRR) domain has a high π_{non}/π_{syn} ratio, and the low polymorphism of the *Pi-ta* gene might have primarily been caused by recurrent selective sweep and constraint by other putative physiological functions. Meanwhile, we provide data to show that the amino acid Ala-918 of H1 in the LRR domain has a close relationship with the resistant phenotype. H1 might have recently arisen during rice domestication and may be associated with the scenario of a blast pathogen–host shift from Italian millet to rice.

SPECIFIC host–pathogen interaction models describing induced defense responses in plants are mediated by gene-for-gene interactions as originally reported by Flor (1971). In these specific host–pathogen interactions, resistance to a particular pathogen is conditional on the presence of a specific *Avr* (avirulence) gene of the pathogen and a specific *R* (resistance) gene in the plant host. Functional *R* genes thus far isolated encode resistance to bacterial, viral, fungal, oomycete, and even nematode and insect pathogens with very different lifestyles and occur outside or inside plant cells (Dangl and Jones 2001). Most *R* genes characterized to date encode products that contain a nucleotide-binding site (NBS) and a series of leucine-rich repeats (LRRs) (Hulbert *et al.* 2001). NBS–LRR proteins can be subdivided on the basis of deduced N-terminal structural features: the Drosophila Toll and mammalian interleukin-1 receptor homology region (TIR) and coiled-coil (CC) region (Hulbert *et al.* 2001). They are abundant in plant genomes, with ~150 described in Arabidopsis (Meyers *et al.* 2003) and ~500 in rice (Bai *et al.* 2002; Zhou *et al.* 2004). TIR-class genes account for most of the NBS–LRR genes in Arabidopsis, although they have not been found in rice sequences. Since the TIR is homologous to the intracellular signaling domain of animals, the predominance of the CC class in cereals might indicate that cereal *R* genes signal through fewer or simpler pathways (Bai *et al.* 2002). The LRR domain has been implicated in protein–protein interactions and may determine resistance specificity (Ellis *et al.* 2000). However, direct interactions between an avirulent protein and its cognate R protein have been demonstrated in only a few host–pathogen systems (Martin *et al.* 2003).

Many *R* genes are physically clustered on the plant genome, forming diverse multigene families, such as *Rpp1*, *Rpp5*, and *Rpp8* in Arabidopsis, *Cf9/4* in tomato, *RGC2* in lettuce, and *Mla* in barley (Martin *et al.* 2003). Polymorphic studies of this category of *R* gene have revealed extensive sequence exchanges and excesses of nonsynonymous over synonymous substitutions among paralogs. These genes are believed to be undergoing balancing selection (Bergelson *et al.* 2001; de Meaux and Mitchell-Olds 2003).

Single-locus *R* genes are also common in plants, *e.g.*, *Rpp13*, *Rps4*, and *Rps2* in Arabidopsis and the *L* locus in flax (Martin *et al.* 2003). However, the intraspecific polymorphisms of these loci greatly differ. For example, an excess of amino acid polymorphism segregation is located within the LRR domain of *Rpp13*. It was

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. EU346955–EU347006.

[1]*Corresponding author:* Institute of Plant Biology, National Taiwan University, 1 Roosevelt Rd., Section 4, Taipei 106, Taiwan.
E-mail: tpl@ntu.edu.tw

suggested as being maintained through continual reciprocal selection between the host and pathogens (Rose *et al.* 2004). However, the relative lack of divergence between resistant and susceptible alleles of *Rps4* may be due to a recent selective sweep (Bergelson *et al.* 2001). Furthermore, some *R* genes exhibit presence/absence (P/A) polymorphisms, which are present in some ecotypes but absent in others within species, *e.g.*, *Rpm1* and *Rps5*. Nucleotide diversities around the deletion junction of these loci were high between the presence and absence accessions. Balancing selection was suggested to play an important role in the molecular evolution of P/A polymorphisms (Shen *et al.* 2006).

Rice blast disease, caused by the fungus *Magnaporthe grisea* (Hebert) Barr. (Rossman *et al.* 1990), is one of the most serious diseases of cultivated rice (*Oryza sativa* L.) worldwide. Despite dozens of "major" disease-resistance (*R*) loci, which are known as *Pi*, being available (Zeigler *et al.* 1994), only six, *Pib*, *Pi-ta*, *Pi36*, *Pi9*, *Pi2*, and *Piz-t*, have been cloned and characterized so far (Liu *et al.* 2007). The *Pi-ta* gene commonly used in rice breeding worldwide originated from *indica* cultivars and was introgressed into *japonica* cultivars to control rice blast disease in the 1950s (Rybka *et al.* 1997). Several dominant rice cultivars carry *Pi-ta*, *e.g.*, IR8, IR36, IR64, and IR72 in Asia and Katy in America (Jia *et al.* 2003; Fukuta *et al.* 2007). It is a single-copy gene near the centromere of chromosome 12. Pi-ta encodes a predicted CC–NBS–LRR-type protein (Bryan *et al.* 2000). In the LRR domain of cultivated rice, a single amino acid difference between resistant and susceptible alleles of *Pi-ta* was identified (Bryan *et al.* 2000). Further DNA sequence analysis revealed unusually low DNA polymorphism of the *Pi-ta* allele among rice cultivars (Jia *et al.* 2003). Its cognate, *AVR-Pita* of *M. grisea*, is a putative metalloprotease possessing properties similar to bacterial effector proteins (Jia *et al.* 2000). Rice cultivars carrying the resistant *Pi-ta* allele have been confirmed in the Chinese field isolate O-137 (Bryan *et al.* 2000), American field isolates ZN57 and ZN67 (Jia *et al.* 2004), and Philippine field isolate IK81-25 (Fukuta *et al.* 2007). This means that Pi-ta can recognize some AVR-pita variations from field isolates. Pi-ta/Avr-pita is one of a very few examples so far that interacts directly in yeast two-hybrid and *in vitro* experiments (Bryan *et al.* 2000).

If we can understand the causes of resistance maintenance, perhaps it can be applied to agricultural practice. Efficient resistance breeding depends largely on understanding patterns of variations of germplasm. However, reductions in the genetic variability of crops frequently occur during the process of domestication associated with severe bottlenecks and artificial selection. Therefore, it is difficult to reconstruct the evolutionary history of the adaptive significance of the *R* gene from domesticated crops. Wild ancestors generally contain higher genetic variations than their domesticated descendants (Rakshit *et al.* 2007; Zhu *et al.* 2007). Wild rice (*O. rufipogon* Griff.) is the ancestor of cultivated rice (Khush 1997). Molecular population genetic analysis of wild rice might provide more information on the selection forces maintaining resistance and leading to the evolution of new specificities in natural populations. The goals of this study were (1) to determine if the *O. sativa* resistance gene shares identity with that of its wild relative (*O. rufipogon*), (2) to analyze the polymorphism patterns of the resistance gene and explain its molecular evolution, and (3) to determine the selective forces shaping the *Pi-ta* gene in wild rice.

MATERIALS AND METHODS

Plant materials: Rice seeds were obtained from the International Rice Research Institute (IRRI) (Los Banos, the Philippines); the National Plant Genetic Resources Center (Taichung, Taiwan); and the National Institute of Genetics (Mishima, Japan). Thirty-six Asian wild rice (*O. rufipogon*) accessions were chosen to produce a worldwide sample, four African wild rice (*O. barthii*) accessions were included as outgroup species, and two *O. sativa* cultivars were used for blast inoculation (Table 1).

Polymerase chain reaction and sequencing: DNA was extracted from fresh leaves or silica gel-dried leaves of *O. rufipogon* and *O. barthii* using the Plant Genomic DNA Extraction Miniprep System (Viogene). Primers that amplified the complete open reading frame of *Pi-ta* were designed from cultivated rice *Pi-ta* alleles (GenBank accession nos. AF207842 and AY196754) (Table 2). *Pi-ta* gene sequences were obtained by directly sequencing the polymerase chain reaction (PCR) amplification products. Allelic fragments were identified on the basis of the synteny of multiple (overlapping) PCR fragments. Heterozygotes were detected as double peaks at polymorphic sites in the chromatogram. The identity of the two *Pi-ta* haplotypes within a heterozygote was inferred through haplotype subtraction (Clark 1990), which deduces haplotypes by a comparison of heterozygote sequences to haplotypes commonly observed in the global sample. When the chromatogram quality did not permit this procedure, PCR products were cloned using the pGEM-T Vector System II (Promega), and the two haplotypes were determined separately. Singletons were checked by sequencing multiple clones. Newly determined sequences were deposited in the NCBI GenBank under accession nos. EU346955–EU347006.

Assessment of disease phenotypes: Wild rice (*O. rufipogon*) is an outcrossing, perennial, tufted, scrambling grass with nodal tillering. Seeds of the same *O. rufipogon* accession may be heterogeneous. Therefore, instead of using several isogenic-line seedlings for blast inoculation, tillers of only one individual of each *O. rufipogon* accession were propagated by node cutting in this experiment. By adopting this method, we could ensure that no ambiguous linkage existed between the genotype and phenotype. Tillers at the four- or five-leaf stage of *O. rufipogon* grown under natural light in a nursery were used to inoculate whole plants with the blast fungus (Figure 1). Two *O. sativa* cultivars, Tadukan (*Pi-ta*) and Tsuyuake (*pi-ta*), were bred to the tiller stage for blast inoculation as positive (disease-resistant) and negative (disease-sensitive) controls, respectively (Bryan *et al.* 2000).

M. grisea isolate IK81-25, which is avirulent to *Pi-ta*, is the standard blast isolate for assaying rice cultivar resistance in the IRRI. Many rice cultivars carrying the *Pi-ta* locus have been confirmed using this isolate. The *O. rufipogon* accessions were analyzed for their disease reaction phenotypes in response to the blast isolate IK81-25 (*AVR-pita*), which was obtained from

TABLE 1

Accessions of *O. rufipogon*, *O. barthii*, and *O. sativa* and their *AVRpita*-dependent resistance phenotypes

Code	Germplasm accession no.	Source country	Source[a]	Haplogroup[b]	Phenotype category[c]
OR1	80433	India	IRRI	H2	S
OR2	80629	India	IRRI	H2	S
OR3	80742	Myanmar	IRRI	H2	R
OR4	80774	Philippines	IRRI	H2	S
OR5	81802	Indonesia	IRRI	H2	R
OR6	81881	India	IRRI	H2/H2	S
OR7	81884	India	IRRI	H2/H2	S
OR8	81892	India	IRRI	H2/H2	R
OR9	103844	Bangladesh	IRRI	H2	S
OR10	104599	Sri Lanka	IRRI	H2	S
OR11	105325	India	IRRI	H-OB	S
OR12	105388	Thailand	IRRI	H1	R
OR13	105426	Sri Lanka	IRRI	H2/H2	S
OR14	105709	India	IRRI	H-OB	S
OR15	105735	Cambodia	IRRI	H1	R
OR16	105760	Thailand	IRRI	H2	S
OR17	105767	Thailand	IRRI	H1	R
OR18	105805	Thailand	IRRI	H2/H2	S
OR19	105890	Bangladesh	IRRI	H1/H2	R
OR20	105898	Bangladesh	IRRI	H2	S
OR21	105953	Indonesia	IRRI	H2	S
OR22	106036	Malaysia	IRRI	H1/H2	R
OR23	106042	India	IRRI	H1	R
OR24	106078	India	IRRI	H2	S
OR25	106145	Laos	IRRI	H2/H2	S
OR26	106262	Papua New Guinea	IRRI	H2	S
OR27	106343	Myanmar	IRRI	H1/H2	R
OR28	106404	Myanmar	IRRI	H1	R
OR29	106409	Vietnam	IRRI	H2	S
OR30	106428	Vietnam	IRRI	H2	S
OR31	106450	Thailand	IRRI	H2	S
OR32	106453	Indonesia	IRRI	H2	S
OR33	106523	Papua New Guinea	IRRI	H2	S
OR34	00287070-1	Taiwan	NPGRC	H2	S
OR35	W1715	China	NIG	H2/H2	S
OR36	W2078	Australia	NIG	H2	R
OB1	89146	Zambia	IRRI	H-OB/H-OB	ND
OB2	104121	Chad	IRRI	H2/H2	ND
OB3	104138	Cameroon	IRRI	H-OB	ND
OB4	104287	Mali	IRRI	H-OB	ND
OS-Tadukan	97A00421		NPGRC	H-OS	R
OS-Tsuyuake	97A06059		NPGRC	H-OS	S

OR, *O. rufipogon*; OB, *O. barthii*; OS, *O. sativa*.

[a] IRRI, International Rice Research Institute (Los Banos, the Philippines); NPGRC, National Plant Genetic Resources Center (Taichung, Taiwan); NIG, National Institute of Genetics (Mishima, Japan).

[b] The haplogroup of heterozygote accessions is indicated by a slash.

[c] S, susceptible; R, resistant; ND, not determined.

the IRRI. The blast mycelium was incubated on yeast starch medium (2 g yeast extract, 10 g soluble starch, 17 g agar, and 1000 ml distilled water) to produce conidia (Chiu *et al.* 1965). For fungal inoculation of rice leaves, conidia were harvested in sterile 0.25% (w/v) gelatin and 0.02% Tween 20 (vol/vol). A conidial suspension (1×10^5 conidia/ml) of *M. grisea* IK81-25 was sprayed onto the leaves using an air sprayer. Plants were inoculated inside a plastic bag. After 24 hr of inoculation in low light, the plants were removed from the bags and placed in a growth chamber. Inoculated plants were kept in a humidity chamber at 28°/23° and 14/10 hr light/dark. The syndrome was recorded 7 days after inoculation. The inoculation experiments were repeated three times. Results from the infection assays were documented by taping diseased leaf tissue using transparent tape and maintaining a reliable record of symptoms (Jia *et al.* 2003). Disease severity was scored on a scale of 0–5 on the basis of lesion types defined by Valent *et al.* (1991). Depending on the presence or absence of lesion formation,

TABLE 2

Primers used for amplification and sequencing of the *Pi-ta* gene

Name[a]	Sequence
2322F	GGGCGATCCATGCTGTCAAATC
2391F	CAGCTAGCGCCGGCGAGCTG
3241F	GCCTGACATGACGAAGATCC
3259F	CCTCACCGACATGCTGTCACAGC
3307F	TTCGGATGTTTGGGAGGTTG
3370R	CCTTTTATCTTGCAAATGCGTCCG
4409F	AGCAGGTTATAAGCTAGGCC
4473F	CCTACAGATCTGTAGCCAGC
4596F	GATTTGGAGCTAGTAGTCGGC
4615R	GCCGACTACTAGCTCCAAATC
4741F	CAGCATGCTATCCCACGTATAGC
4901R	TCAGGCAAACCACGGCTAAC
5877F	CCAAGGACTACAACACTTGC
6012R	GTACCTGTGACAGTGAGGGAGC
6041R	CCAGTCCATTTGGGGATGCT
6922R	GTTCTTTGATCCAAGTGTTAGG

[a] The numbers are referenced to the *Pi-ta* gene of rice (GenBank accession no. AF207842). F, forward; R, reverse.

two categories were used: resistant (R; types 0 and 1, no lesion formation), and susceptible (S; types 2–5, lesion formation).

Data analysis: Sequences were aligned using ClustalX version 1.8 (Thompson *et al.* 1997) and manually edited using BioEdit version 6.0.5 (Hall 1999). Orthologs of *O. barthii* were used as the outgroup in the analyses. The best-fitting substitution model was estimated using ModelTest version 3.7 (Posada and Crandall 1998). The estimated parameters were then incorporated into PAUP* version 4.0 (Swofford 2002) to generate a maximum-likelihood gene tree. Levels of silent-site nucleotide diversities per site were estimated as π (Nei 1987) and θ ($4N_e\mu$) (Watterson 1975). Genetic parameters and a sliding window of π_{non}/π_{syn} and K_a/K_s analyses were conducted using DnaSP version 4.0 (Rozas *et al.* 2003). DnaSP 4.0 was also used to perform tests of selection, including Tajima's *D* test (Tajima 1989), Fu and Li's *D* test (Fu and Li 1993), Fay and Wu's *H* test (Fay and Wu 2000), and the McDonald–Kreitman test (MK) (McDonald and Kreitman 1991).

RESULTS

Levels and patterns of nucleotide variations in the *Pi-ta* gene: The *Pi-ta* sequences were examined, and 10 of 36 *O. rufipogon* accessions and two of four *O. barthii* accessions were determined to be heterozygous (Table 1). These two species have outcrossing mating systems. The heterozygous ratio was comparable to the 30–50% outcrossing rate of *O. rufipogon* estimated by Barbier (1989). Figure 1 shows polymorphic sites of the *Pi-ta* genes among *O. rufipogon*, *O. barthii*, and *O. sativa*. The length of the complete alignment was 4306 bp, including sites with alignment gaps. Numbering began from the first position of the start codon. Asian and African wild rice each carry a different haplotype from the other (Figures 1 and 2). We suggest that this was caused by introgression. We identified these accessions according to their morphological characteristics (data not shown); OR11 and OR14 were from *O. rufipogon*, and OB2 was from *O. barthii*. The *Pi-ta* sequences of the three accessions (OR11, OR14, and OB2) were excluded from further analyses.

The alleles showed numerous indels and nucleotide polymorphisms. In total, 91 nucleotide polymorphic sites and 18 indels (ranging from 1 to 73 bp in length) were detected in this 4.3-kb region (Figure 1). Only one 9-bp indel occurred in the exon region of the gene (OR13b and OR19b), but did not cause a frameshift. The predicted proteins encoded by individual alleles were between 925 and 928 amino acids long. All alleles had the same overall domain structure, and there were no truncated genes.

The *Pi-ta* gene showed lower nucleotide polymorphism ($\pi_{silent} = 0.00235$; this study) in comparison with the 10 reference genes surveyed from *O. rufipogon*, which had an average value of $\pi_{silent} = 0.0095$ (Zhu *et al.* 2007). Tang *et al.* (2006) also reported an average π_{silent} of 0.00583 for 10 reference loci of *O. rufipogon*. This low-variation average is comparable with that of *O. sativa*, which has a nucleotide diversity (π_{silent}) in the range of 0.0011–0.0035 (Tang *et al.* 2006; Zhu *et al.* 2007). The nucleotide polymorphism of the coding region (π = 0.00089) was shown to be much lower than that of the noncoding region (π = 0.00345, Table 3). This was caused by the low polymorphisms of the NBS (π = 0.00050) and LRR domains ($\pi = 0.00060$) (Table 3).

From the distribution of polymorphic variations, two distinct sequence types were defined in the population of *O. rufipogon*. Among 44 alleles, we identified eight identical or nearly identical sequences (differing by four singleton nucleotide sites; Figure 1). This group of sequences was named haplogroup H1, and seven of them encoded the same protein sequence (Figures 1 and 2). The remaining sequences (haplogroup H2) harbored 36 different haplotypes with 64 segregating sites (65 mutations) and 14 indels, and 19 of them encoded the same protein sequence (Figures 1 and 2).

There were 19 nonsynonymous sites, and only 6 synonymous polymorphisms in the coding region. The *Pi-ta* sequences of *O. barthii* were used to determine an ancestral form of the single-nucleotide polymorphism. H2 sequences likely contain the ancestral form and preexisting neutral variations. Three of 10 accessions carrying the heterozygote (OR19, OR22, and OR27) were of the H1/H2 type, and the others were of the H2/H2 type. Two of eight fixed differences at nucleotide positions 17 and 4271 between the H1 and H2 haplogroups were nonsynonymous sites. We found that these two nucleotides in the *Pi-ta* gene of *O. barthii* were G and T, respectively. Therefore, the ancestral forms were serine-6 (Ser-6) and Ser-918, and the derived forms were isoleucine-6 (Ile-6) and alanine-918 (Ala-918) in the *Pi-ta* gene of *O. rufipogon* (Figure 1). The Pi-ta amino acids of *O. sativa* were previously surveyed, and Ala-918 was correlated with the gene-for-gene specificity char-

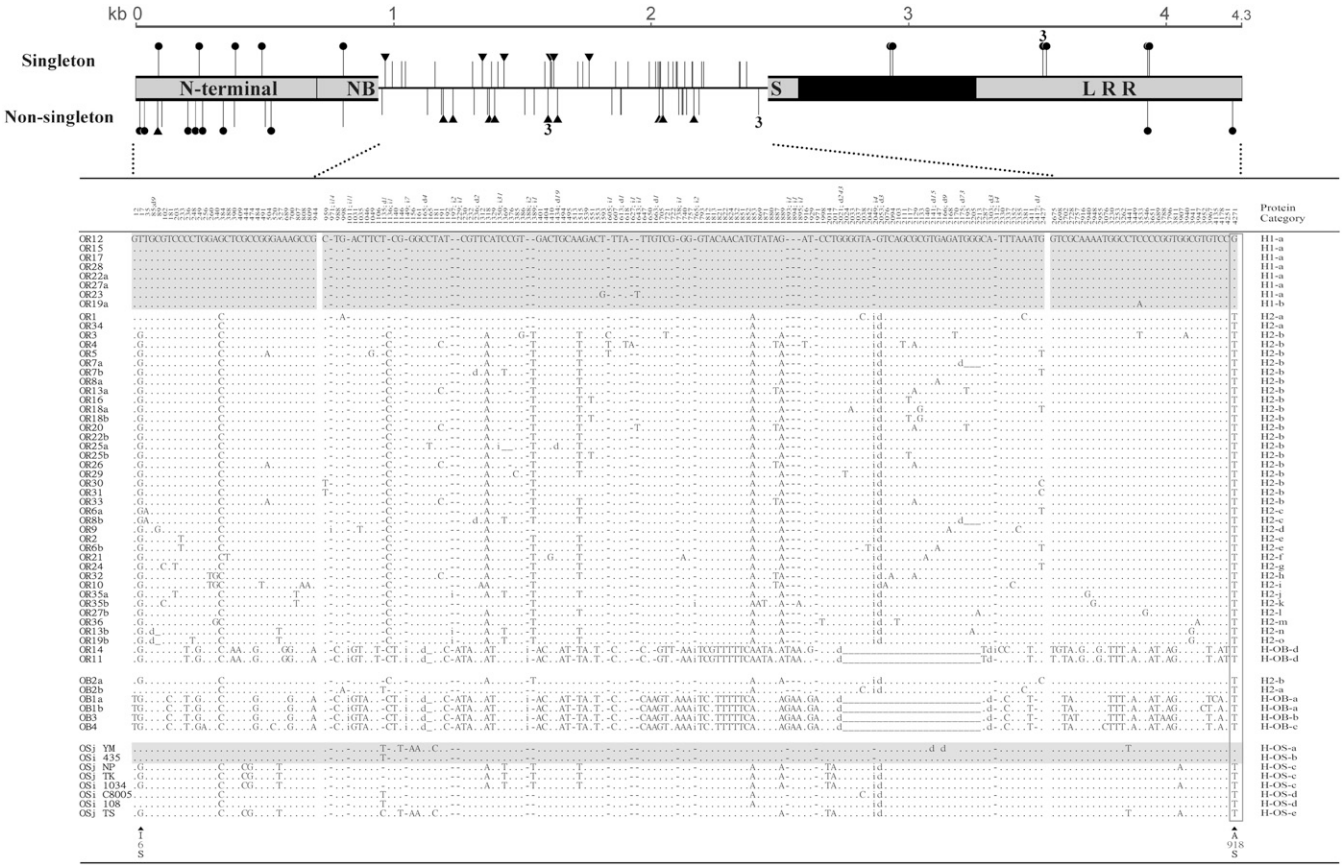

FIGURE 1.—Summary of DNA variations in the 4.3-kb *Pi-ta* region of *O. rufipogon*, *O. barthii*, and *O. sativa*. (Top) The gene model of *Pi-ta*. Boxes indicate exons, and shaded boxes represent different domains. Polymorphic sites found in two or more accessions of *O. rufipogon* are indicated by lines below the gene model, whereas single polymorphisms are indicated by lines above the gene model. Indels are indicated by solid triangles. Nonsynonymous polymorphisms are indicated by a solid circle. 3, three-nucleotide change. (Bottom) DNA variations are summarized. Site 1 corresponds to the first position of the start codon. Shaded sequences represent haplogroup H1 and the resistant *Pi-ta* gene of rice. Dots represent nucleotide variants identical to the first sequence. Dashes indicate the absence of the corresponding variant. d, deletion; i, insertion; d#, deletion of number of base pairs; i#, insertion of number of base pairs. The boxed nucleotide is the nonsynonymous polymorphism that encodes the Ala-918 amino acid. Groups of identical deduced proteins are in the right column. Eight *O. sativa* sequences were adopted from GenBank. OSi & j, *indica* and *japonica* rice, respectively. Rice cultivar abbreviations: YM, Yashiro-mochi (*Pi-ta*); NP, Nippobare; TK, Takasagowase; TS, Tsuyuake (*pi-ta*). Two fixed nonsynonymous sites between H1 and H2 are marked at the bottom.

acteristic of the *Pi-ta/AVR-pita* system, and Ser-6 being replaced by Ile-6 did not affect the disease resistance activity (BRYAN *et al.* 2000).

Tests of positive selection: The frequency spectrum of the variation found in the *Pi-ta* region of all accessions was tested with Tajima's *D* and Fu and Li's *D*. The values of these statistics were all negative and significantly deviated from the neutral model (Table 3). But the value of Fay and Wu's *H* test was −0.83721, which did not significantly differ from zero (Table 3). The H2 haplogroup exhibited a trend similar to the entire accessions of *O. rufipogon* (Table 3). The H1 haplogroup did not deviate from the expectation of neutrality.

The MK test was also used to determine if the level of nonsynonymous polymorphisms observed at the *Pi-ta* locus exceeded that expected under neutrality. Under neutrality, the levels of intraspecific polymorphism and interspecific divergence are expected to be correlated (MCDONALD and KREITMAN 1991). The MK test failed to provide evidence of positive selection (Table 4). Since studies of other *R* genes revealed a pattern of diversifying selection acting on the LRR region and functional studies indicated that pathogen-recognition specificity is encoded by this region (HULBERT *et al.* 2001), we tested whether this region alone was responsible for the deviation from neutral expectations. The analyses of different partitions also showed no deviation from neutrality (Table 4). These results indicated no excess amino acid polymorphism segregation, even in the LRR region, relative to neutral expectations (Table 4). Sliding-window analyses were used to characterize the patterns of polymorphism and divergence across the *Pi-ta* gene. There were obvious nonsynonymous polymorphisms peaking in the LRR compared with nonsynonymous divergence (Figure 3). The π_{non}/π_{syn} ratio of 1.839 in this region contrasts with a K_a/K_s value of 0.678 for the

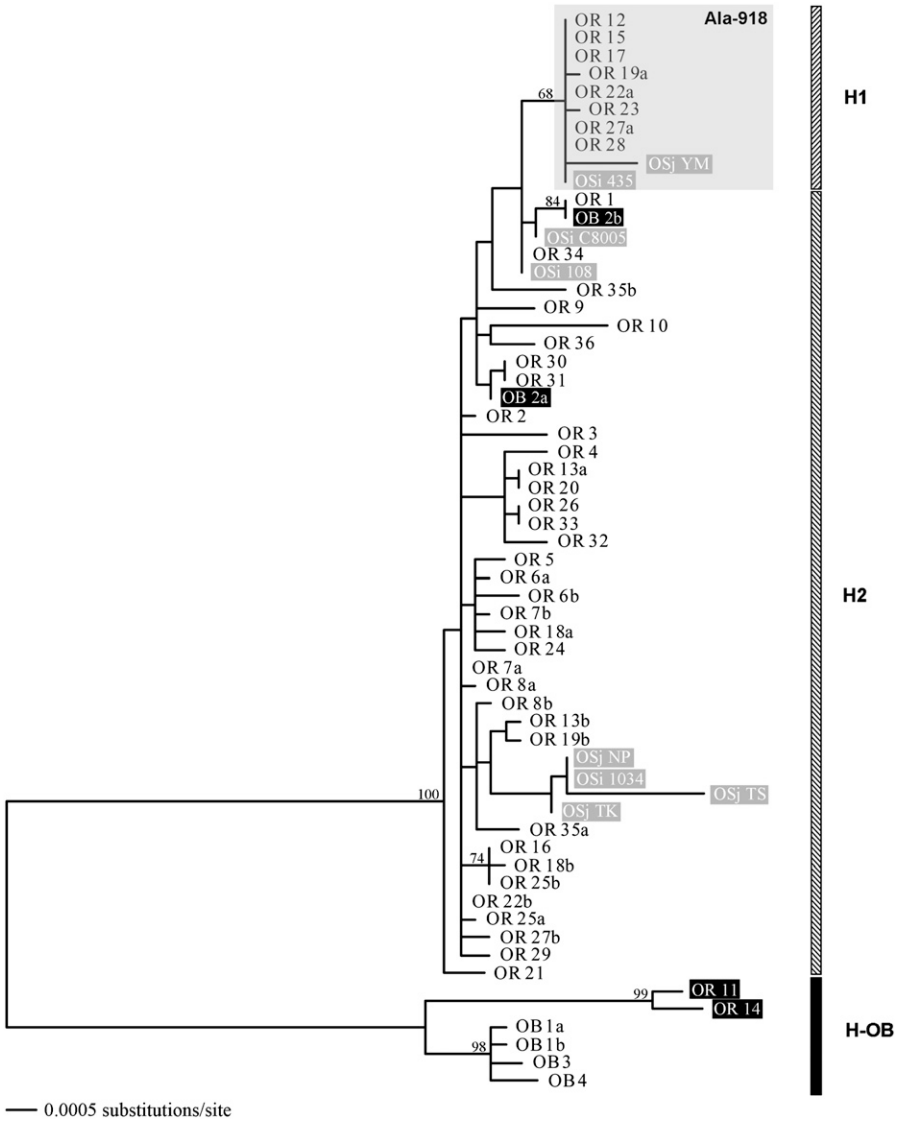

FIGURE 2.—Maximum-likelihood tree of the complete DNA sequence of *Pi-ta* alleles. The *O. barthii* sequence was used as the outgroup. Bootstrap proportions of 1000 bootstrap replicates >65 are indicated above the branches. Solid boxes mark alleles that might have been caused by admixture. The shaded clade highlights alleles with the Ala-918 amino acid. Shaded boxes are alleles of *O. sativa*. H1, haplogroup H1; H2, haplogroup H2; H-OB, haplogroup of *O. barthii*. The rice cultivar abbreviations are the same as those given in the legend to Figure 2.

interspecific comparison (Table 5). Although the level of nonsynonymous divergence exceeded synonymy in the NBS domain ($K_a/K_s = 7.348$; Table 5), it was reported to be the most conserved region of the *R* gene.

Assessment of disease phenotypes: Rice seedlings were used in the blast inoculation because younger plants are quite susceptible. Resistance of rice plants to blast varies with the growth stage, and they become resistant after the bolting stage (SUZUKI 1975). Resistance to *M. grisea* also increases as the leaf age of rice plants increases on the same tillers. In most inoculation results, lesions formed only on the expanding leaf. Therefore, lesions on expanding leaves of each tiller were used to evaluate the resistance to blast (Table 1, Figure 4). Twelve accessions showed the typical resistant phenotype. The accessions belonging to haplogroup H1 all exhibited resistance. Haplogroup H1 is distributed sporadically around the rim of the Indian Ocean (Figure 5). All but four accessions (OR3, OR5, OR8, and OR36) of haplogroup H2 were susceptible to *M. grisea* IK81-25. However, no unique replacement site was found in the Pi-ta amino acid sequence of these four accessions. In a typical rice blast inoculation assay, the well-characterized *M. grisea* isolates sometimes showed unexpected results that suggested that certain additional *AVR/R* gene interactions might mask the identification (JIA *et al.* 2003; FUKUTA *et al.* 2007). Alternatively, these lines may harbor an *R* gene that confers resistance to an unknown *Avr* gene, other than *AVR-Pita*, of *M. grisea* IK81-25.

DISCUSSION

Admixtures between Oryza species: The AA-genome group of Oryza, also called the *O. sativa* complex, possesses various barriers against hybridization among group members, but all of these barriers are not complete in their effect (OKA 1988). Some genes of this group show a hybrid pattern possibly caused by introgression (TANG *et al.* 2006; RAKSHIT *et al.* 2007).

TABLE 3

Polymorphism and neutral test of different groups of the *Pi-ta* gene

	Total sites	*S*	π	θ	Tajima's *D*	Fu and Li's *D* (with outgroup)	Fay and Wu's *H*
				All *O. rufipogon*			
Entire gene	4306	72	0.00196	0.00410	*−1.86739**	*−2.60794**	−0.83721
Coding	2787	24	0.00089	0.00207	*−1.89395**	−2.08664	2.17759
N-terminal	705	13	0.00229	0.00429	−1.43175	−0.00616	1.37209
NBS	357	3	0.00050	0.00193	−1.57449	−1.72555	0.17336
LRR	1032	6	0.00060	0.00156	−1.66947	*−2.71653**	0.54334
Intron	1519	44	0.00345	0.00760	*−1.90876**	*−2.47935**	−3.01480
				H1			
Entire gene	4306	2	0.00012	0.00018	−1.31009	−1.26208	0.42857
Coding[a]	2787	1	0.00009	0.00014	−1.05482	−2.08993	0.21429
LRR	1032	1	0.00024	0.00037	−1.05482	−1.26208	0.21429
Intron	1519	1	0.00017	0.00026	−1.05482	−1.26208	0.21429
				H2			
Entire gene	4306	60	0.00147	0.00357	*−2.19527***	*−2.39754**	−1.36508
Coding	2787	22	0.00068	0.00191	*−2.18652**	−1.98819	1.79365
N-terminal	705	12	0.00178	0.00416	−1.80016	−0.05768	1.15873
NBS	357	3	0.00061	0.00203	−1.56135	−1.64357	0.20952
LRR	1032	5	0.00032	0.00117	*−1.89379**	*−2.66924**	0.31746
Intron	1519	42	0.00321	0.00759	*−2.07182**	−2.22887	−3.15873

Significance levels were determined by 10,000 random coalescent simulations on the basis of the number of alleles and the observed number of segregating sites. *S*, number of segregating sites; π, nucleotide diversity; θ, Watterson's estimator, $4N_e\mu$. Italics indicate the significant statistics. $*P < 0.05$; $**P < 0.01$.

[a] The N-terminal and NBS of H1 have no variation, so these two parts were not included.

Two groups of highly diversified haplotypes have consistently been identified in the 10 selected unlinked loci of the AA-genome group of Oryza. This was suggested to have been caused by recent admixture, possibly as a result of human migration coupled with evolving agricultural practices (Tang *et al.* 2006). Two groups of divergent haplotypes also occurred in the *Pi-ta* gene in *O. rufipogon* and *O. barthii* (Figures 1 and 2). They carry different haplotypes from each other, which could be explained simply as a result of admixture. The admixture effect increases intraspecific polymorphism and decreases interspecific divergence. The samples, including OR11 and OR14, would influence many neutral tests (data not shown). To make a reasonable inference, three accessions (OR11, OR14, and OB2) were excluded from the examination.

The two alleles of OB2 are distributed within the tree clade of *O. rufipogon* (Figure 2). This might reflect

TABLE 4

Summary of the McDonald–Kreitman test

	Polymorphic sites within *O. rufipogon*	Fixed differences between *O. rufipogon* and *O. barthii*	Fisher's exact test (two-tailed)
Entire gene			
Synonymous	39	37	$P = 0.229181$
Nonsynonymous	23	13	
Coding region			
Synonymous	9	6	$P = 0.79407$
Nonsynonymous	23	13	
N-terminal			
Synonymous	3	2	$P = 0.597523$
Nonsynonymous	12	4	
NBS			
Synonymous	1	0	$P = 1.000000$
Nonsynonymous	2	1	
LRR			
Synonymous	4	2	$P = 1.000000$
Nonsynonymous	6	5	

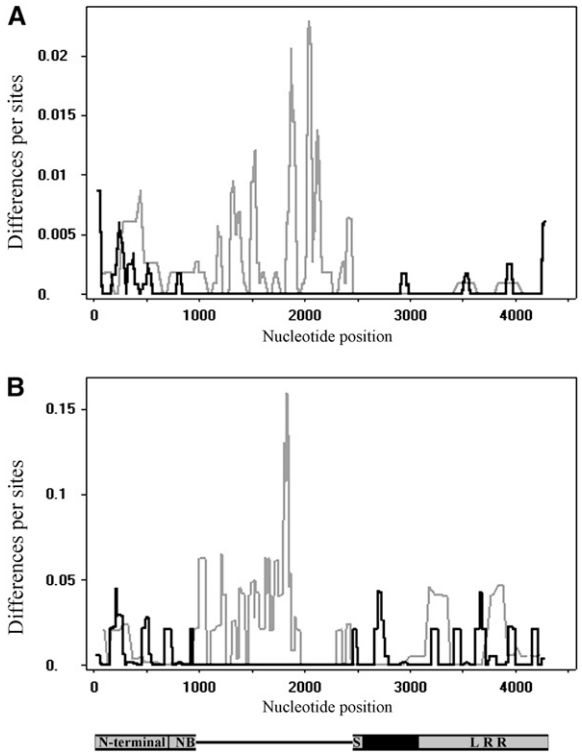

FIGURE 3.—Sliding-window analyses: (A) Average number of differences per site between *Pi-ta* alleles within *O. rufipogon* and (B) between *O. rufipogon* and *O. barthii*. Shaded lines are synonymous variations, while solid lines are nonsynonymous ones. Values are midpoints of the 25-bp windows. The positions of the different domains are indicated below the plots.

admixture with Asian rice in the recent past. However, alleles of OR11 and OR14 are a sister group of haplotype OB (Figure 2) and were collected from southern India just at the margin of *O. rufipogon*'s distribution and a site closest to Africa (Figure 5). This specific geographical area might be a hybridization zone between African and Asian wild rice plants. The alleles of OR11 and OR14 have seven unique segregating sites, greater than the intraspecific polymorphism of *O. barthii*, which possibly reflects ecological adaptations to different environments (Figures 1 and 2). We inferred that the introgression was not a recent event and has already persisted for a long time.

TABLE 5

Nonsynonymous over synonymous polymorphism and divergence of the *Pi-ta* gene of wild rice

	π_{non}/π_{syn} [a]	K_a/K_s [b]
Coding region	0.787	0.649
N-terminal	0.696	0.754
NBS	0.326	7.348
LRR	1.835	0.677

[a] Ratio of nonsynonymous site diversity over synonymous site diversity of *O. rufipogon*.

[b] Evolutionary distances between orthologs in coding sequences between species.

Variations of *Pi-ta* genes may primarily have been caused by recurrent selective sweeps: The significantly negative values of Tajima's *D* and Fu and Li's *D* statistics of the entire gene, including coding and noncoding regions, indicate an excess of low-frequency variants that may have been caused by positive selection or population expansion (NORDBORG and INNAN 2002) (Table 3). If one attempts to test the hypothesis that a given gene has been the target of positive selection, the challenge consists of differentiating between the effects of demography and selection on genetic variations. We adopted six reference loci from the same 15 accessions of *O. rufipogon* germplasm used in this study, including three loci published by LONDO *et al.* (2006) and three loci from Y.-C. CHIANG (unpublished data), and these were pooled for the analyses (Table 6). The nonsignificant negative Tajima's *D* statistics might have caused skewing due to population expansion (ZHU *et al.* 2007). Furthermore, recent surveys of genomewide polymorphism in *O. rufipogon* also showed that the data do not fit standard neutral models. Examples such as the average 10 reference genes (TANG *et al.* 2006), 9 of 10 nuclear loci (ZHU *et al.* 2007), and 111 randomly chosen gene fragments (CAICEDO *et al.* 2007) all had negative Tajima's *D* values (Table 6). There is a systematic shift across the genome toward lower-than-expected allelic frequencies, further supporting the scenario of population expansion.

However, *Pi-ta* with a significantly negative Tajima's *D* value and very low polymorphism cannot satisfactorily be explained only by population expansion. We suggest that some selective forces are responsible for the deviation in the *Pi-ta* gene from a neutral model. Using an outgroup to polarize mutations as ancestral or derived, Fay and Wu's *H* statistic can detect an excess of high-frequency-derived variants that is a unique pattern produced by hitchhiking relative to neutral expectations. The theoretical basis of this effect has been studied for individual recent selective sweeps. Neither all of the *O. rufipogon* alleles nor just haplogroup H2 had a significantly negative Fay and Wu's *H* statistic. However, no significant *H* tests indicated old or recurrent selective sweeps (PRZEWORSKI 2002; KIM 2006). Therefore, we cannot reject the lack of a significant result for the *Pi-ta* gene being caused by reduced statistical power of that test or because the current selective scenario is more complex than that assumed in the statistical test. There are two sources of rare alleles in DNA sequences that have experienced a selective sweep. First, a selective sweep may generate a star-shaped genealogy with many short outer branches. Mutations mapping onto those branches will be found at low frequency in the sample. Second, when only one selective sweep has occurred in the recent past, one or two lineages may escape coalescence during hitchhiking events and generate long

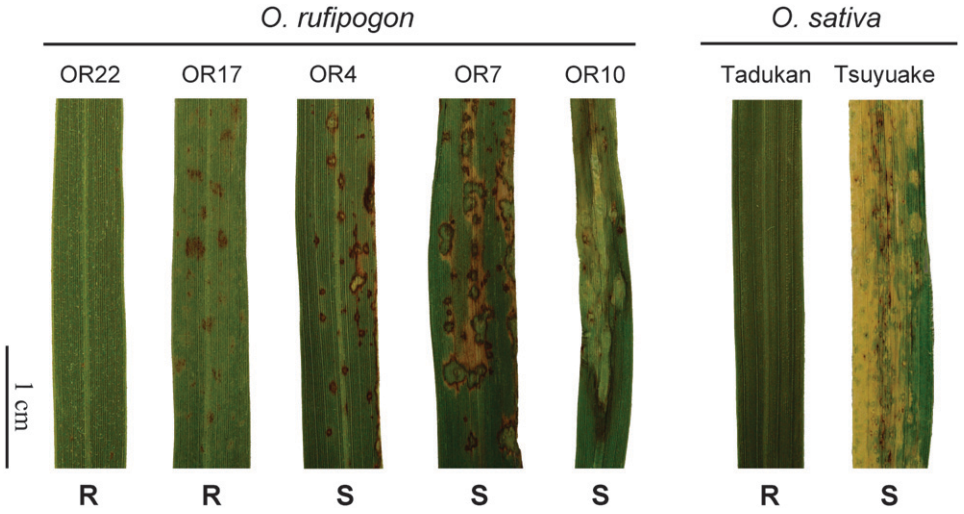

FIGURE 4.—Disease reaction to *M. grisea* IK81-25 (*AVR-pita*) on wild rice and rice cultivars. R, resistant; S, susceptible.

inner branches (KIM 2006). We found those characteristics in the *Pi-ta* gene. On the basis of the genealogical tree of the *Pi-ta* gene, the ancestral H2 might have been shaped by an older selective sweep within or near the gene, and H1 was recently derived from H2 (Figure 2). We suggest that the low polymorphism of *Pi-ta* genes may have been caused mainly by recurrent selective sweeps.

Furthermore, the LRR domain has a high π_{non}/π_{syn} ratio (1.839), indicating relatively more nonsynonymous polymorphisms than synonymous ones. All of the replacements are at the solvent-explored amino acid in the xxLxxLxxLxxLxLxxxx motif (ZHOU *et al.* 2004). One of five polymorphisms, Ala-918 has been confirmed to be correlated with the gene-for-gene specificity characteristic of the *Pi-ta/AVR-Pita* system of *O. sativa* (BRYAN *et al.* 2000). On the basis of inoculation assays, we show that the amino acid Ala-918 of H1 in the LRR domain has a close relationship with the resistant phenotype. The almost complete lack of polymorphisms within haplogroup H1, even in noncoding regions, is consistent with the selective sweep hypothesis (MAYNARD-SMITH and HAIGH 1974). A similar pattern was also found in *Rps4* of Arabidopsis, and it was argued to be the result of a recent selective sweep (BERGELSON *et al.* 2001; BAKKER *et al.* 2006). However, the relative frequency of haplogroup H1, in only 8 of 36 accessions, is very low. The genomewide survey of *R* gene polymorphisms in Arabidopsis showed that worldwide selective sweeps are uncommon (BAKKER *et al.* 2006).

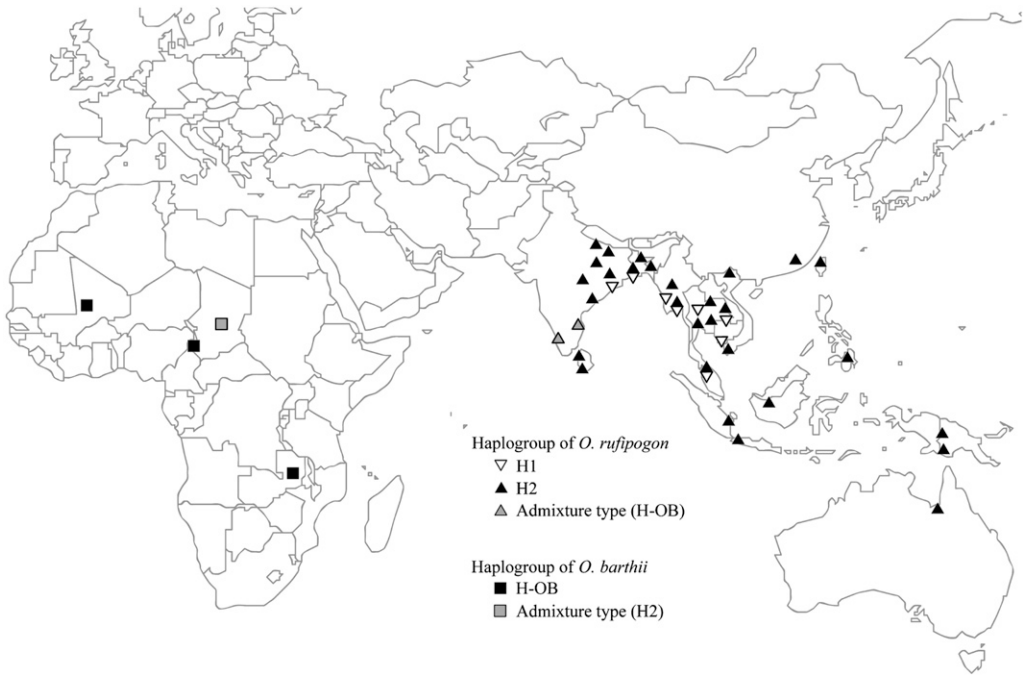

FIGURE 5.—Geographic distributions of the surveyed haplogroups of the *O. rufipogon* and *O. bathii* accessions.

TABLE 6

Tajima's *D* values of reference loci

Gene	Length (bp)	Tajima's *D*
Methionine synthetase B[a]	4649	−0.94230
DC1[a]	2041	−1.47828
SAM (chromosome 1)[a]	1076	−1.76405
SAM (chromosome 5)[b]	1166	−1.24622
atpB_rbcL[b]	800	−0.38329
pVATPase[b]	1264	−1.46108
Average		−1.21254
Average 10 reference genes[c]		−1.205
Average 10 reference genes[d]		−0.5946
Average 111 random gene fragments[e]		−0.2710

[a] Subsamples from Y.-C. Chiang (unpublished data). DC1, *S*-adenosylmethionine decarboxylase 1; SAM, *S*-adenosyl methionine synthetase.
[b] Subsamples from Londo *et al.* (2006).
[c] Tang *et al.* (2006).
[d] Zhu *et al.* (2007).
[e] Caicedo *et al.* (2007).

Some resistance alleles with intermediate frequency exhibiting a partial selective sweep might have resulted from historical allele-frequency fluctuations (Stahl *et al.* 1999). This may be one of the causes of the partial selective sweep of H1.

Thirty-two of 36 alleles of haplogroup H2 encode the same amino acid sequence in the LRR domain (protein categories H2-a–k) (Figure 1). These are opposite patterns for diversifying selection relative to most other *R* genes. In addition to the immune receptor function, some NBS–LRR proteins are involved in signaling cascades important for additional cellular processes, such as drought tolerance, development, and photomorphogenesis (Tameling and Joosten 2007). One example can give us a clue to the linkage between low nucleotide polymorphism and functional constraint of the LRR domain. Rps4 of Arabidopsis not only has conferred resistance to *Pseudomonas syringae* carrying the effector AvrRPS4 (Gassmann *et al.* 1999), but also is involved in phyB signaling (Faigon-Soverna *et al.* 2006). A relative lack of polymorphism between resistant and susceptible alleles was also found in *Rps4*; only a single amino acid polymorphism and no synonymous differences were detected in its LRR region (Bergelson *et al.* 2001; Bakker *et al.* 2006). Therefore, we suggest that a functional constraint of the LRR domain of *Pi-ta* might have occurred if it is also associated with other physiological functions in addition to resistance, about which we know nothing.

The NBS domain having an unexpectedly high K_a/K_s ratio probably is not because of a great divergence in the function of signal transduction between the two species, but instead was caused by a lack of variation in the synonymous site. A genomewide survey of rice NBS–LRR genes showed that "expansion of diversity" occurred not only in the LRR domain but also in the N-terminal domain (Zhou *et al.* 2004). We found that the greatest variation occurred in the N-terminal of the *Pi-ta* gene.

Evolutionary history and distribution of the *Pi-ta* gene resistant to rice blast: Epidemic diseases could not have existed before the origins of agriculture, because they can sustain themselves only in large dense populations that did not exist before agriculture; hence, they are often called "crowded diseases" (Diamond 2002). For example, the origin of the fungal wheat pathogen *Mycosphaerella graminicola* coincided with the known domestication of wheat in the Fertile Crescent ~8000–9000 bc (Stukenbrock *et al.* 2007). The virulent factor *AVR-pita* was present in both the Oryza and Setaria clades (Couch *et al.* 2005). Couch *et al.* (2005) suggested that rice blast arose from a single origin of rice infection, following a host shift from Italian millet (*Setaria italica*). Both Italian millet and rice were domesticated and appeared to have co-occurred early in the history of agriculture in Asia. Bayesian-derived estimates suggested an early origin of the rice-infecting pathogen ~9000 years ago, which may have been associated with rice domestication (Couch *et al.* 2005). Since wild Oryza and Setaria grasses grow in different habitats—wetlands and relatively dry land, respectively—we suggest a scenario of the new resistant allele arising during domestication. The widespread and dense distributions of Italian millet and rice presented frequent opportunities for contact, and a host shift occurred from the Setaria blast pathogen to rice. The relatively lower genetic variation of cultivated rice after artificial selection and dense planting on rice farmland allowed the rice blast to become an epidemic disease. Meanwhile, rice and its wild relatives are distributed sympatrically, and so the newly arising rice blast pathogen was also transferred to wild rice nearby. Wild rice faced a new virulent factor, the AVR-Pita of blast, that it had not encountered before. H1 was the mutation that allowed wild rice to fight the pathogen in the recent past.

The DNA sequences of the resistant *Pi-ta* allele of most rice cultivars distributed in different areas are identical, *e.g.*, Yashiro-mochi, K1, Reiho, Tetep, Tadukan, Katy, and Drew (Jia *et al.* 2003); the rice cultivar 435 (GenBank no. AB364491) has some polymorphisms that differ from those of others (Figure 1). *O. rufipogon* and *O. sativa* have very similar patterns of resistant *Pi-ta* alleles, and no fixed polymorphism exists between H1 and *Pi-ta* alleles of cultivated rice (Figure 1). This characteristic further supports the above scenario. Cultivated rice might have directly acquired the resistant *Pi-ta* allele from its wild ancestor accompanying domestication or through gene flow between each other. Some differences in resistant *Pi-ta* alleles might have arisen from different standing genetic variations of *O. sativa* and *O. rufipogon*, or mutations may have recently accumulated.

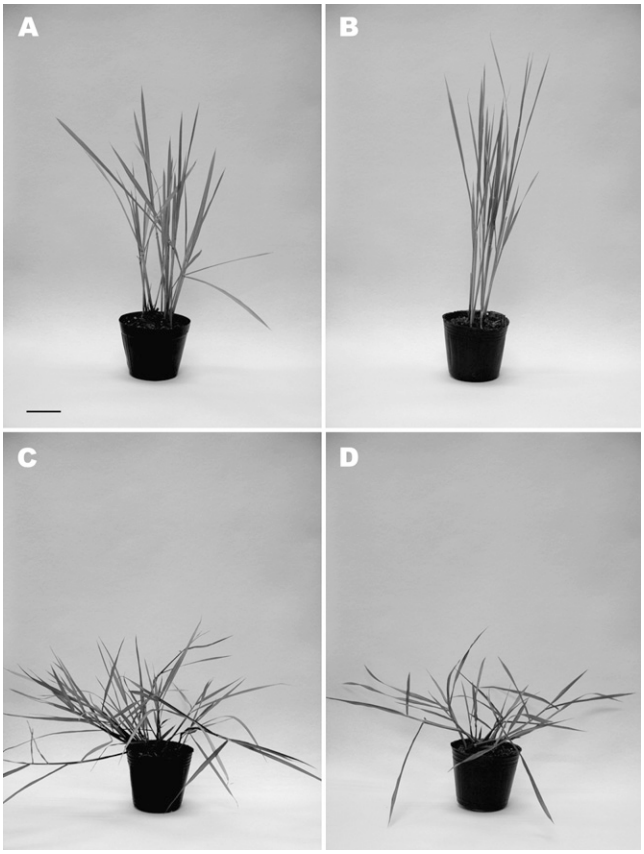

FIGURE 6.—Tillers of *O. rufipogon*. (A) OR19; (B) OR22; (C) OR21; and (D) OR31. (A and B) Accessions show the erect culms, which are blast resistant. (C and D) Accessions show the spreading culms, which were blast susceptible in this study. Bar, 5 cm.

Furthermore, we found that wild rice individuals with the H1 allele are patchily distributed (Figure 5). One possibility for the limited and specific areas of occurrence of H1 is that the selective forces were either temporally or spatially incomplete (HUDSON *et al.* 1994, 1997), perhaps due to meteorological factors of different regions. Rice blast enters an epidemic phase when a new race of the pathogen appears or when meteorological conditions activate the pathogen. Meteorological conditions, *e.g.*, rain and wind, act directly on the pathogen in the prepenetration stage, during formation of appressoria, and at initial colonization. Further development of the pathogen in host tissues is more heavily influenced by the genetic resistance/susceptibility of the rice plant (SUZUKI 1975). These multiple pathogenic conditions might have caused the mosaic distribution of *Pi-ta*-resistant and -susceptible alleles around the rim of the Indian Ocean. In addition, the typical growth characteristic of *O. rufipogon* is a spreading culm (MOLDENHAUER and GIBBONS 2003). However, six of eight accessions containing haplogroup H1, excluding OR23 and OR28, have erect culms (Figure 6). The genes that control the erect culm phenotype might have some linkage with the *Pi-ta* gene.

We suggest here that the recurrent selective sweep and a recently arising novel mutation might be major evolutionary events causing the polymorphism of *Pi-ta* in *O. rufipogon*. It seems that a conventional arms race model can explain the polymorphism of the *Pi-ta* gene (DE MEAUX and MITCHELL-OLDS 2003). However, even though the Pi-ta/AVR-pita interaction is well known, the exact function of the Pi-ta protein of susceptible line that exhibits compatibility with AVR-pita is not yet known. The putative selective sweep of the ancestral H2 may have been caused by complex events such as recognition of another AVR protein of pathogens, hitchhiking associated with internal or nearby selective sites, or influence of cross talk between different signal transductions. A more comprehensive examination of Pi-ta, which will help answer these questions, is needed. Furthermore, the strength of the selective effect can be determined by examining genomic regions near *Pi-ta* and allelic frequencies within populations. These analyses should provide a more complete molecular evolutionary history of this resistance gene.

We thank the International Rice Research Institute, Los Banos, the Philippines; the National Institute of Genetics, Mishima, Japan; and the National Plant Genetic Resources Center, Taichung, Taiwan, for providing cultivated and wild rice samples. We are grateful to members of the Agriculture Research Institute, Taichung, Taiwan, including C. G. Chern, C. P. Li, and W. S. Jwo, for help with the wild rice breeding and L. J. Shieh for help with the rice blast inoculation. We also thank S. J. Chang, Miaoli District Agricultural Research and Extension Station, Miaoli, Taiwan, for providing leaf samples of wild rice. This work was supported by the National Science Council, Taiwan (NSC96-2621-B-002-002).

LITERATURE CITED

BAI, J., L. A. PENNILL, J. NING, S. W. LEE, J. RAMALINGAM *et al.*, 2002 Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res. **12:** 1871–1884.

BAKKER, E. G., C. TOOMAJIAN, M. KREITMAN and J. BERGELSON, 2006 A genome-wide survey of *R* gene polymorphisms in *Arabidopsis*. Plant Cell **18:** 1803–1818.

BARBIER, P, 1989 Genetic variation and ecotypic differentiation in the wild rice species *Oryza rufipogon*. II. Influence of the mating system and life-history traits on the genetic structure of populations. Jpn. J. Genet. **64:** 273–285.

BERGELSON, J., M. KREITMAN, E. A. STAHL and D. TIAN, 2001 Evolutionary dynamics of plant *R*-genes. Science **292:** 2281–2285.

BRYAN, G. T., K.-S. WU, L. FARRALL, Y. JIA, H. P. HERSHEY *et al.*, 2000 A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell **12:** 2033–2046.

CAICEDO, A. L., S. H. WILLIAMSON, R. D. HERNANDEZ, A. BOYKO, A. FLEDEL-ALON *et al.*, 2007 Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet. **3:** e163.

CHIU, R.-J., C.-C. CHIEN and S.-Y. LIN, 1965 Physiologic races of *Piricularia oryzae* in Taiwan, pp. 245–255 in *The Rice Blast Disease Proceedings Symposium*, edited by C.-C. TU, C.-W. TSAI, C.-C. CHIEN, W.-H. TSAI and Y.-C. CHANG. Taiwan Agricultural Research Institute Special Publication no. 32, Taichung, Taiwan.

CLARK, A. G., 1990 Inference of haplotypes from PCR-amplified samples of diploid populations. Mol. Biol. Evol. **7:** 111–122.

COUCH, B. C., I. FUDAL, M.-H. LEBRUN, D. THARREAU, B. VALENT *et al.*, 2005 Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. Genetics **170:** 613–630.

Dangl, J. L., and J. D. G. Jones, 2001 Plant pathogens and integrated defence responses to infection. Nature **411:** 826–833.

de Meaux, J., and T. Mitchell-Olds, 2003 Evolution of plant resistance at the molecular level: ecological context of species interactions. Heredity **91:** 345–352.

Diamond, J., 2002 Evolution, consequences and future of plant and animal domestication. Nature **418:** 700–707.

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

Faigon-Soverna, A., F. G. Harmon, L. Storani, E. Karayekov, R. J. Staneloni *et al.*, 2006 A constitutive shade-avoidance mutant implicates TIR-NBS-LRR proteins in Arabidopsis photomorphogenic development. Plant Cell **18:** 2919–2928.

Fay, J. C., and C.-I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Flor, H. H., 1971 Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. **9:** 275–296.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Fukuta, Y., L. A. Ebron and N. Kobayashi, 2007 Genetic and breeding analysis of blast resistance in elite indica-type rice (*Oryza sativa* L.) bred in International Rice Research Institute. Jpn. Agric. Res. Q. **41:** 101–114.

Gassmann, W., M. E. Hinsch and B. J. Staskawicz, 1999 The Arabidopsis *RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant J. **20:** 265–277.

Hall, T. A., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41:** 95–98.

Hudson, R. R., K. Bailey, D. Skarecky, J. Kwiatowski and F. J. Ayala, 1994 Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. Genetics **136:** 1329–1340.

Hudson, R. R., A. G. Saez and F. J. Ayala, 1997 DNA variation at the *Sod* locus of *Drosophila melanogaster*: an unfolding story of natural selection. Proc. Natl. Acad. Sci. USA **94:** 7725–7729.

Hulbert, S. H., C. A. Webb, S. M. Smith and Q. Sun, 2001 Resistance gene complexes: evolution and utilization. Annu. Rev. Phytopathol. **39:** 285–312.

Jia, Y., S. A. McAdams, G. T. Bryan, H. P. Hershey and B. Valent, 2000 Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO J. **19:** 4004–4014.

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

Jia, Y., Z. Wang, R. G. Fjellstrom, K. A. K. Moldenhauer, M. A. Azam *et al.*, 2004 Rice *Pi-ta* gene confers resistance to the major pathotypes of the rice blast fungus in the United States. Phytopathology **94:** 296–301.

Khush, G. S., 1997 Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. **35:** 25–34.

Kim, Y., 2006 Allele frequency distribution under recurrent selective sweeps. Genetics **172:** 1967–1978.

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

Londo, J. P., Y.-C. Chiang, K.-H. Hung, T.-Y. Chiang and B. A. Schaal, 2006 Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc. Natl. Acad. Sci. USA **103:** 9578–9583.

Martin, G. B., A. J. Bogdanove and G. Sessa, 2003 Understanding the functions of plant disease resistance proteins. Annu. Rev. Plant Biol. **54:** 23–61.

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

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

Meyers, B. C., A. Kozik, A. Griego, H. Kuang and R. W. Michelmore, 2003 Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell **15:** 809–834.

Moldenhauer, K. A. K., and J. H. Gibbons, 2003 Rice morphology and development, pp. 103–127 in *Rice: Origin, History, Technology, and Production*, edited by C. W. Smith and R. H. Dilday. John Wiley & Sons, New York.

Nei, M, 1987 *Molecular Evolutionary Genetics*. Columbia University, New York.

Nordborg, M., and H. Innan, 2002 Molecular population genetics. Curr. Opin. Plant Biol. **5:** 69–73.

Oka, H. I., 1988 *Origin of Cultivated Rice*. Japan Scientific Societies Press, Tokyo/Elsevier, Amsterdam.

Posada, D., and K. A. Crandall, 1998 MODELTEST: testing the model of DNA substitution. Bioinformatics **14:** 817–818.

Przeworski, M., 2002 The signature of positive selection at randomly chosen loci. Genetics **160:** 1179–1189.

Rakshit, S., A. Rakshit, H. Matsumura, Y. Takahashi, Y. Hasegawa *et al.*, 2007 Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. Theor. Appl. Genet. **114:** 731–743.

Rose, L. E., P. D. Bittner-Eddy, C. H. Langley, E. B. Holub, R. W. Michelmore *et al.*, 2004 The maintenance of extreme amino acid diversity at the disease resistance gene, *RPP13*, in *Arabidopsis thaliana*. Genetics **166:** 1517–1527.

Rossman, A. Y., R. J. Howard and B. Valent, 1990 *Pyricularia grisea*, the correct name for the rice blast disease fungus. Mycologia **82:** 509–512.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496–2497.

Rybka, K., M. Miyamoto, I. Ando, A. Saito and S. Kawasaki, 1997 High resolution mapping of the *indica*-derived rice blast resistance genes. II. *Pi-ta*^{2} and *Pi-ta* and a consideration of their origin. Mol. Plant Microbe Interact. **10:** 517–524.

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

Stahl, E. A., G. Dwyer, R. Mauricio, M. Kreitman and J. Bergelson, 1999 Dynamics of disease resistance polymorphism at the *RPM1* locus of *Arabidopsis*. Nature **400:** 667–671.

Stukenbrock, E. H., S. Banke, M. Javan-Nikkhah and B. A. McDonald, 2007 Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. Mol. Biol. Evol. **24:** 398–411.

Suzuki, H., 1975 Meteorological factors in epidemiology of rice blast. Annu. Rev. Phytopathol. **13:** 239–256.

Swofford, D. L., 2002 *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods)*, 4.0 Beta. Sinauer Associates, Sunderland, MA.

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

Tameling, W. I. L., and M. H. A. J. Joosten, 2007 The diverse roles of NB-LRR proteins in plants. Physiol. Mol. Plant Pathol. **71:** 126–134.

Tang, T., J. Lu, J. Huang, J. He, S. R. McCouch *et al.*, 2006 Genomic variation in rice: genesis of highly polymorphic linkage blocks during domestication. PLoS Genet. **2:** e199.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins, 1997 The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876–4882.

Valent, B., L. Farrall and F. G. Chumley, 1991 *Magnaporthe grisea* genes for pathogenicity and virulence identified through a series of backcrosses. Genetics **127:** 87–101.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Zeigler, R. S., S. A. Leong and P. S. Teng, 1994 *Rice Blast Disease*. Commonwealth Agricultural Bureau International, Wallingford, UK.

Zhou, T., Y. Wang, J.-Q. Chen, H. Araki, Z. Jing *et al.*, 2004 Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol. Gen. Genomics **271:** 402–415.

Zhu, Q., X. Zheng, J. Luo, B. S. Gaut and S. Ge, 2007 Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol. Biol. Evol. **24:** 875–888.

Communicating editor: J. A. Birchler