The Plant Cell, Vol. 21: 1573–1591, May 2009, www.plantcell.org © 2009 American Society of Plant Biologists

Association Genetics Reveals Three Novel Avirulence Genes from the Rice Blast Fungal Pathogen *Magnaporthe oryzae* [W] [OA]

Kentaro Yoshida,[a,1] Hiromasa Saitoh,[a,1] Shizuko Fujisawa,[a] Hiroyuki Kanzaki,[a] Hideo Matsumura,[a] Kakoto Yoshida,[a] Yukio Tosa,[b] Izumi Chuma,[b] Yoshitaka Takano,[c] Joe Win,[d] Sophien Kamoun,[d] and Ryohei Terauchi[a,2]

[a] Iwate Biotechnology Research Center, Kitakami, Iwate, 024-0003 Japan
[b] Laboratory of Plant Pathology, Kobe University, Kobe, 657-8501 Japan
[c] Laboratory of Plant Pathology, Kyoto University, Kyoto, 606-8502 Japan
[d] The Sainsbury Laboratory, John Innes Centre, Norwich, NR4 7UH United Kingdom

To subvert rice (*Oryza sativa*) host defenses, the devastating ascomycete fungus pathogen *Magnaporthe oryzae* produces a battery of effector molecules, including some with avirulence (AVR) activity, which are recognized by host resistance (R) proteins resulting in rapid and effective activation of innate immunity. To isolate novel avirulence genes from *M. oryzae*, we examined DNA polymorphisms of secreted protein genes predicted from the genome sequence of isolate 70-15 and looked for an association with AVR activity. This large-scale study found significantly more presence/absence polymorphisms than nucleotide polymorphisms among 1032 putative secreted protein genes. Nucleotide diversity of *M. oryzae* among 46 isolates of a worldwide collection was extremely low ($\theta = 8.2 \times 10^{-5}$), suggestive of recent pathogen dispersal. However, no association between DNA polymorphism and *AVR* was identified. Therefore, we used genome resequencing of Ina168, an *M. oryzae* isolate that contains nine *AVR* genes. Remarkably, a total of 1.68 Mb regions, comprising 316 candidate effector genes, were present in Ina168 but absent in the assembled sequence of isolate 70-15. Association analyses of these 316 genes revealed three novel *AVR* genes, *AVR-Pia*, *AVR-Pii*, and *AVR-Pik/km/kp*, corresponding to five previously known *AVR* genes, whose products are recognized inside rice cells possessing the cognate *R* genes. *AVR-Pia* and *AVR-Pii* have evolved by gene gain/loss processes, whereas *AVR-Pik/km/kp* has evolved by nucleotide substitutions and gene gain/loss.

INTRODUCTION

Arms-race coevolution dramatically impacts the genome of pathogens and plants. Resistance often follows the gene-for-gene model in which plant resistance (*R*) gene products recognize avirulence (AVR) proteins, a subset of pathogen-secreted virulence proteins known as effectors, to trigger hypersensitive cell death and immunity. Genome-wide analyses indicate that *R* genes are the most polymorphic class of genes in plants (Clark et al., 2007). Pathogen effectors are also rapidly evolving and in a few cases have been reported in regions with high genome plasticity (Orbach et al., 2000; Gout et al., 2006). However, genome-wide analysis of variation of effector candidate genes in plant pathogenic fungi and genome-wide DNA polymorphism information for the identification of *AVR*s by association genetics approach are still limited (Armstrong et al., 2005).

Rice blast caused by the ascomycete fungus *Magnaporthe oryzae* (Couch and Kohn, 2002) is the most devastating fungal disease of rice (*Oryza sativa*; Zeigler et al., 1994; Talbot, 2003). Understanding the function of *M. oryzae* effectors, their host targets, and *AVR-R* gene interactions is important to devise effective means to control the disease. Molecular identification of *M. oryzae AVR* genes will tremendously facilitate race identification of blast fungus and help rapid and effective deployment of *R* genes in rice cultivation. Furthermore, identification and analysis of *M. oryzae AVR* genes will help to elucidate fungal mechanisms of pathogenesis and shed light on the mechanisms involved in coevolution of fungal effectors and their host targets. Therefore, we aimed to identify and functionally analyze *M. oryzae AVR*s and effectors.

To date, >25 *R* genes encoding proteins that recognize *M. oryzae* AVRs have been mapped on the rice genome (Wang et al., 1994), and six *R* genes against different races of *M. oryzae* have been cloned from rice (Wang et al., 1999; Bryan et al., 2000; Qu et al., 2006; Lin et al., 2007; Ashikawa et al., 2008; Lee et al., 2009). By contrast, only four *AVR* genes have been isolated from *M. oryzae*. The AVR genes *PWL1* and *PWL2*, which were isolated by map-based cloning, are genes responsible for the nonpathogenicity of rice pathogenic strains of *M. oryzae* against *Eragrostis curvula*, weeping lovegrass (Kang et al., 1995; Sweigard et al., 1995). PWL1 and PWL2 share 75% amino acid identity and encode Gly-rich hydrophobic proteins with secretion signal

[1] These authors contributed equally to this work.
[2] Address correspondence to terauchi@ibrc.or.jp.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Ryohei Terauchi (terauchi@ibrc.or.jp).
[W] Online version contains Web-only data.
[OA] Open Access articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.109.066324

sequences. *AVR-Pita*, which was also isolated by map-based cloning, confers resistance to rice cultivars harboring the *Pita R* gene (Orbach et al., 2000). *AVR-Pita* encodes a putative secreted protein with similarity to metalloproteases, which is recognized by Pita inside rice cells (Jia et al., 2000). The fourth *AVR* to be cloned was *ACE1*, encoding a putative hybrid protein of a polyketide synthase and a peptide synthase (Bohnert et al., 2004). *ACE1* is remarkable in that its secondary metabolite enzyme product, not the ACE1 protein itself, is the avirulence determinant recognized by the host plant. Judging by the large number of rice *R* genes against *M. oryzae*, there still are many AVRs awaiting characterization.

The whole-genome draft sequence of isolate 70-15, a laboratory strain of *M. oryzae*, was published (Dean et al., 2005). The genome assembly consists of 37.8 Mb DNA encoding 11,109 predicted protein coding genes. In this study, we set out to identify novel *AVR*s and effectors from *M. oryzae* using genome-wide DNA polymorphisms based initially on the 70-15 genome sequence. However, after finding that DNA polymorphisms among the 1032 candidate effector genes in the 70-15 genome did not show any association with *AVR*s, we performed the resequencing of the genome of a field isolate Ina168. This revealed that a remarkable 1.68 Mb of DNA was present in Ina168 but absent from the assembled genome sequence of 70-15 and allowed us to identify 316 additional candidate effector genes. Association analysis of these 316 genes identified three novel *AVR*s, *AVR-Pia*, *AVR-Pii*, and *AVR-Pik/km/kp*, which encode products that are recognized inside cells of rice plants carrying the cognate *R* genes.

RESULTS

Low Levels of Nucleotide Polymorphisms versus Higher Levels of PCR Amplified/Nonamplified Polymorphisms in *M. orzyae* Secreted Protein Genes

Since the majority of known effectors of filamentous fungi are secreted proteins that can be predicted computationally (Ellis et al., 2007; Kamoun, 2007), we focused on *M. oryzae* genes encoding secreted proteins based on the published whole-genome draft sequence of strain 70-15 (Dean et al., 2005). The predicted proteome of 70-15 (11,109 proteins; magnaporthe_grisea_2.3_proteins; http://www.broad.mit.edu/cgi-bin/ annotation/magnaporthe/download_license.cgi) was screened using a previously described bioinformatics pipeline outlined in Methods (Lee et al., 2003; Torto et al., 2003), resulting in 1306 putative secreted proteins (see Supplemental Data Set 1 online).

To identify genes harboring polymorphisms associated with AVR phenotypes, we performed association genetic analyses of DNA polymorphisms in the selected *M. oryzae* genes and avirulence on a previously defined set of resistant rice cultivars (Kiyosawa et al., 1986). PCR primers (see Supplemental Data Set 1 online for primer sequences) were designed to amplify a DNA fragment (1 to 2 kb) of each of the 1306 loci to reveal presence/absence polymorphisms of PCR products in the 23 *M. oryzae* isolates of rice pathogens consisting of 22 field isolates collected mainly from Japan and one experimental strain 70-15 (Table 1). Among these isolates, the presence or absence of *AVR* genes in each *M. oryzae* isolate has been previously phenotypically determined for the eleven genes *AVR-Pia*, *AVR-Pii*, *AVR-Pik*, *AVR-Pikm*, *AVR-Pikp*, *AVR-Piz*, *AVR-Pita*, *AVR-Pita2*, *AVR-Piz-t*, *AVR-Pib*, and *AVR-Pit*. The isolates were sprayed onto the leaves of diagnostic rice cultivars differentially harboring 11 known resistance (*R*-) genes, *Pia*, *Pii*, *Pik*, *Pik-m*, *Pik-p*, *Piz*, *Pita*, *Pita2*, *Piz-t*, *Pib*, and *Pit* cognate to the *AVR* genes (Kiyosawa et al., 1986), and compatible or incompatible interaction between them was recorded. It is notable that in our experiments, the reaction between isolate 70-15 and these 11 rice *R*-genes could not be precisely defined since the disease symptoms caused by 70-15 were not clear (see Supplemental Figure 1 online). Typical incompatible reactions cause no or very small reddish hypersensitive response (HR) lesions if present, whereas typical compatible reactions cause development of brown spindle-shaped necrotic lesions (see Supplemental Figure 1 online). By contrast, 70-15 shows poor virulence. It causes intermediate responses in most of rice cultivars tested: infection causes reddish lesions of various sizes, but they do not further develop into necrotic lesions. Judging from these observations, we decided not to define each reaction involving 70-15 as compatible or incompatible (Table 1). The ambiguous disease symptoms caused by the 70-15 strain may be related to its origin: it is derived from a cross between two isolates of *M. oryzae*, one of which is a rice pathogen and the other a weeping lovegrass pathogen (Chao and Ellingboe, 1991). In addition to the PCR-amplified/nonamplified polymorphism analysis to address the presence/absence of the genes, EcoTILLING (Comai et al., 2004), a high-throughput technique for the detection of DNA polymorphisms based on heteroduplex mismatch cleavage by an endonuclease CELI and gel electrophoresis (Till et al., 2003), was used to identify base substitutions and short insertion/deletions (indels) exhibiting association with *AVR*s. In the EcoTILLING experiment, a total of 46 *M. oryzae* rice isolates (see Supplemental Table 1 online) representing a worldwide collection was used. The 46 rice isolates include the 21 isolates (Table 1) used for the PCR amplified/nonamplified polymorphism study.

Out of 1306 loci, 1032 were successfully PCR amplified from the isolate 70-15, whereas the rest were not, presumably because of suboptimal design of PCR primers (see Supplemental Data Set 1 online). For subsequent analysis, we focused on these 1032 loci. PCR-amplified/nonamplified polymorphism was observed in 394 out of 1032 (38.2%) loci among the 21 isolates (Table 2). EcoTILLING detected polymorphisms in 227 out of 1032 (22.0%; Table 2) loci among the 46 isolates. Thus, there were significantly more PCR-amplified/nonamplified polymorphisms, presumably caused by presence/absence of the genes, than base changes/short indels in the putative secreted protein genes of *M. oryzae* rice pathogen ($\chi^2 = 34.7$, $P < 3.9e^{-9}$). EcoTILLING revealed low levels of average nucleotide diversity (Watterson 1975; $\theta = 8.2 \times 10^{-5}$) among the 46 isolates of *M. oryzae*. The majority (78%) of the genes were monomorphic (see Supplemental Figure 2 online). All the DNA polymorphism data were used to infer phylogenetic relationships among the 21 isolates of *M. oryzae*, revealing three well-separated clades J-A, J-B, and J-C among the isolates (Figure 1). However, none of the detected polymorphisms (see

Table 1. Twenty-Three *M. oryzae* Isolates Used for PCR Screen and Genetic Association Test of Putative Secreted Protein Genes

Code No.[a]	Isolate	Host	Race	Origin	*R* Gene[b]											PCR Screen	EcoTILLING Screen[a]
					Pia	*Pii*	*Pik*	*Pik-m*	*Piz*	*Pita*	*Pita2*	*Piz-t*	*Pik-p*	*Pib*	*Pit*		
1	Ina168	*O. sativa*	101.1	Japan	R	R	R[1]	R[1]	R	S	R	R	S	R	R	+	+
2	70-15	*O. sativa*	Unknown	–	?	?	?	?	?	?	?	?	?	?	?	+	+
3	84R-62B	*O. sativa*	447.0	Japan	S	S	R	R	S	R	R	S	R	R	R	+	–
4	Y93-245c-2	*O. sativa*	337.1	China	S	S	S	S	R	S	S	R	S	R	R	+	–
5	Shin85.86	*O. sativa*	001.0	Japan	R	R	R	R	R	R	R	R	R	R	R	+	+
6	Ina72	*O. sativa*	031.1	Japan	R	R	S	S	R	R	R	R	S	R	R	+	+
7	TH68-140	*O. sativa*	035.1	Japan	R	S	S	S	R	R	R	R	S	R	R	+	+
8	TH69-8	*O. sativa*	071.1	Japan	R	R	S	S	R	R	R	R	S	R	R	+	+
9	1836-3	*O. sativa*	033.1	Japan	S	R	S	S	R	R	R	R	S	R	R	+	+
10	TH68-126	*O. sativa*	033.1	Japan	S	R	S	S	R	R	R	R	S	R	R	+	+
11	22-4-1-1	*O. sativa*	107.0	Japan	S	S	R	R	R	S	R	R	R	R	R	+	+
12	9505-3	*O. sativa*	037.1	Japan	S	S	S	S	R	R	R	R	S	R	R	+	+
13	Sasa2	*O. sativa*	037.1	Japan	S	S	S	S	R	R	R	R	S	R	R	+	+
14	TH78-15	*O. sativa*	177.1	Japan	S	S	S	S	S	S	R	R	S	R	R	+	+
15	Br18	*O. sativa*	176.5	Brazil	S	S	S	S	S	S	R	R	S	R	S	+	+
16	TH87-20-BII	*O. sativa*	007.2	Japan	S	S	R	R	R	R	R	R	R	S	R	+	+
17	Hoku1	*O. sativa*	007.0	Japan	S	S	R	R	R	R	R	R	R	R	R	+	+
18	Ina86-137	*O. sativa*	007.0	Japan	S	S	R	R	R	R	R	R	R	R	R	+	+
19	2012-1	*O. sativa*	007.4	Japan	S	S	R	R	R	R	R	R	R	R	S	+	+
20	2403-1	*O. sativa*	007.4	Japan	S	S	R	R	R	R	R	R	R	R	S	+	+
21	88A	*O. sativa*	433.5	Japan	S	R	S	S	R	R	R	S	S	R	S	+	+
22	Br10	*O. sativa*	403.4	Brazil	S	R	R	R	R	R	R	S	R	R	S	+	+
23	P-2b	*O. sativa*	303.1	Japan	S	R	R	R	R	S	S	R	S	R	R	+	+

[a]EcoTILLING screen of polymorphisms was not carried out for the two isolate code numbers 3 and 4, making the total number of isolates studied by both PCR and EcoTILLING screens 21.

[b]R, rice host is resistant; S, rice host is susceptible; ?, it is difficult to judge whether rice host is resistant or susceptible; [1], the interaction between Ina168 and *Pik*/*Pik-m* was not stable (i.e., Ina168 exhibits virulence or avirulence phenotypes on rice harboring *Pik*/*Pik-m* depending on the experiments).

Supplemental Data Set 1 online) showed a significant association with AVR phenotypes.

Genome Resequencing of *M. oryzae* Isolate Ina168 Reveals 1.11 Mb Absent from Isolate 70-15

Because we could not detect association between AVR and DNA polymorphisms in the putative secreted protein genes of 70-15, we hypothesized that the majority of the *AVR* genes tested in the 22 isolates (Table 1) are lacking in the draft sequence of 70-15. Therefore, we performed whole-genome resequencing of Ina168, a Japanese isolate known to carry nine avirulence genes, *AVR-Pia*, *AVR-Pii*, *AVR-Pik*, *AVR-Pikm*, *AVR-Piz*, *AVR-Pita2*, *AVR-Pizt*, *AVR-Pib*, and *AVR-Pit* (Table 1).

Using the 454 sequencing technology (Margulies et al., 2005) (454 Life Sciences), we sequenced the Ina168 genome to ~10× coverage (491.6 Mb) (Table 3). The sequence was assembled into 4582 contigs corresponding to 38.0 Mb. Of this sequence, 36.3 Mb could be aligned between Ina168 and 70-15 with an average nucleotide divergence $d = 2.8 \times 10^{-4}$ along the lines of the nucleotide polymorphism levels revealed by EcoTILLING. Remarkably, a total of 1.68 Mb in multiple regions did not align with the 70-15 contig sequence based on the threshold given in 454 FLX reference mapper (minimum overlap length 40 bp; minimum overlap identity 90%). Of these Ina168 unmapped sequences, 1.11 Mb did not match even raw sequence reads of 70-15 (BLASTN threshold of E-value > $1 \times e^{-5}$) and were defined as Ina168 specific. Conversely, a total of 5.09 Mb regions of 70-15 genome sequences did not match to Ina168 sequences with the threshold given in 454 FLX reference mapper. Since our Ina168 sequence reads may not cover the entire genome, the actual size of regions present in 70-15 but absent in Ina168 may be < 5.09 Mb. The 1.68 Mb unmapped Ina168 DNA regions contain 316 open reading frames (ORFs) coding for putative secreted proteins larger than 50 amino acids; we named these putative secreted proteins pex1 to pex316 (see Supplemental Data Set 2 online). The disproportionally large number (316) of putative secreted protein genes identified in the 1.68-Mb Ina168 unmapped regions compared with that in 70-15 whole-genome sequence (1306 putative secreted protein genes in the 37.8 Mb regions) is likely due to the different gene prediction methodologies used in our study compared with the 70-15 genome annotation: (1) for Ina168 unmapped sequence, we set the length threshold at > 50 amino acids to recover as many candidate proteins as possible, whereas in 70-15, most of predicted proteins are larger than 100 amino acids (see Supplemental Figure 3 online); (2) any ORFs were considered as candidate genes in Ina168 unmapped sequence, whereas gene prediction

Table 2. Summary of PCR-Amplified/Nonamplified Polymorphisms versus Base Change Polymorphisms as Detected by EcoTILLING in 1032 Putative Secreted Protein Genes of *M. oryzae*

	PCR-Amplified/ Nonamplified Polymorphisms among the 21 Isolates[a]	Base Changes among the 46 Isolates[a]
No. of loci studied	1032	1032
No. of polymorphic loci	394	227

Details of polymorphisms in each locus are given in Supplemental Data Set 1 online. $\chi^2 = 34.7$; P = 3.93e-9.

[a]The 46 isolates studied for base changes (see Supplemental Table 1 online) include the 21 isolates (Table 1) studied for PCR-amplified/ nonamplified polymorphisms.

software (FGENSH and GENWISE) was employed for stringent gene annotation in the 70-15 genome (Dean et al., 2005). Presence/absence polymorphisms of these 316 Ina168 pex ORFs were tested by PCR using the 23 *M. oryza* isolates (Table 1). Remarkably, 113 ORFs out of 316 could be PCR amplified from 70-15 DNA, suggesting possible incomplete coverage of the genome by the 70-15 draft sequence. A total of 173 ORFs segregated among the 22 field isolates of known AVR phenotypes (Figure 1). Segregation patterns of the ORFs could be roughly grouped into those generally conforming to the phylogeny of the isolates (ORFs in the left part of the red/black panel in Figure 1) and those not associated with the phylogeny (ORFs in the right part of the panel). To identify novel AVR genes, we next examined three ORFs where the presence of the ORF was significantly correlated with the function of a specific AVR gene.

Presence/Absence Polymorphisms Reveal Association of Three ORFs with AVR Phenotypes

The presence of the PCR product corresponding to the 85–amino acid pex22 protein (Figure 2A) showed a perfect association with *AVR-Pia* function (Figure 2B) among the 22 *M. oryzae* isolates tested. The probability of observing this result by chance is $P = 3.8 \times 10^{-5}$ (Fisher's exact test). pex22 is located on the contig264 of 2.7 kb in size (Figure 2C). Sequences upstream of pex22 showed similarity to a Pot3 transposase (Hamer et al., 1989; Farman et al., 1996). The association with AVR-Pia function also held for the regions upstream (Region 2) and downstream (Region1) of pex22 (Figures 2B and 2C). The predicted pex22 protein sequence showed no similarity to known protein domains. DNA gel blot analysis using the pex22 ORF as probe (Figure 2B) confirmed that the presence/absence of PCR products is indeed caused by the presence/absence of the pex22 sequence in the genomes of respective isolates. The *pex22* gene was shown to be actively transcribed during infection of *M. oryzae* isolate Ina168 of leaf sheath cells of susceptible rice cultivar (cultivar Shin-2) as revealed by SuperSAGE (Matsumura et al., 2003) and 3′-rapid amplification of cDNA ends (RACE) RT-PCR (see Supplemental Figure 4 online).

The presence of the PCR product corresponding to the 70–amino acid pex33 protein (Figure 3A) did not conform to the phylogeny (Figure 1) but showed a perfect association with *AVR-Pii* (Figure 3B), which is highly significant (Fisher's exact test; $P = 2.0 \times 10^{-6}$). The pex33 region contained another ORF for a putative secreted protein, pex279, in the antisense strand. DNA gel blot analysis using pex33 ORF as probe detected DNA bands only in the isolates that showed PCR amplification (Figure 3B), which confirms that presence/absence of PCR amplification is caused by the presence/absence of pex33 sequence in the genomes of the isolates. The pex33/279 ORFs are located inside the 2.0-kb contig389 (Figure 3C). Sequence upstream of pex33 showed a similarity to the *gag* gene of Maggy transposon (Farman et al., 1996), and the sequence downstream of pex33 contained a region similar to a telomere-like sequence. Perfect association with *AVR-Pii* was lost when the fragments located upstream (pex33 upstream) or downstream (pex90) of pex33 were tested (Figure 3B). Transcription of *pex33*, but not *pex279*, during infection of *M. oryzae* isolate Ina168 to leaf sheath cells of the susceptible rice cultivar Shin-2 was confirmed by SuperSAGE and 3′-RACE RT-PCR (see Supplemental Figure 4 online). BLAST (Altschul et al., 1997) search of the National Center for Biotechnology Information (NCBI) nonredundant (nr) protein database using the predicted pex33 protein sequence as query identified three *M. oryzae* proteins (XP_366338.2, XP_001407225.1, and XP_364190.1) exhibiting similarity with the expectation value of $E < 0.02$ (Figure 3D). These proteins show a higher conservation in the predicted signal peptide region compared with the mature protein region where the sequences are highly diverged from each other. Interestingly, MEME analysis (Bailey et al., 2006) identified two conserved motifs present in all four protein sequences: motif-1, [LI]xAR[SE][DSE]; and motif-2, [RK]CxxCxxxxxxxxxxxH, the latter exhibiting similarity to the C2H2 zinc finger motif (Figure 3D; Evans and Hollenberg, 1988).

Presence/absence of PCR product of another ORF, pex31 coding for a protein of 113 amino acids (Figure 4A), showed an incomplete but significant association with both *AVR-Pik* and *AVR-Pikm* (Fisher's exact test: P = 0.0039) and *AVR-Pikp* (P = 0.017; Figure 4B). In our panel of 22 isolates with known races, *AVR-Pik* and *AVR-Pikm* are always linked, so that we hereafter use *AVR-Pik/km* to indicate either of these two *AVR* genes. Presence/absence of *AVR-Pikp* is associated with that of *AVR-Pik/km* with the exceptions of the two isolates Ina168 and P-2b that harbor *AVR-Pik/km* but not *AVR-Pikp* (Figure 4B, lanes 1 and 23). DNA gel blot analysis with pex31 probe suggests that presence/absence of the PCR product corresponds to the presence/absence of a genomic region harboring this ORF

Table 3. Summary of Resequencing of *M. oryzae* Ina168 Genome

Total length of sequence read	491.6 Mb
Total size of contigs	38.0 Mb
Number of contigs	4,582
The size of Ina168 DNA regions unmapped to 70-15 contigs	1.68 Mb
The size of Ina168 DNA regions without match to 70-15 raw read sequences	1.11 Mb
Divergence between Ina168 and 70-15 *d* (Jukes & Cantor)	2.8×10^{-4}
SD	$\pm 1.3 \times 10^{-2}$

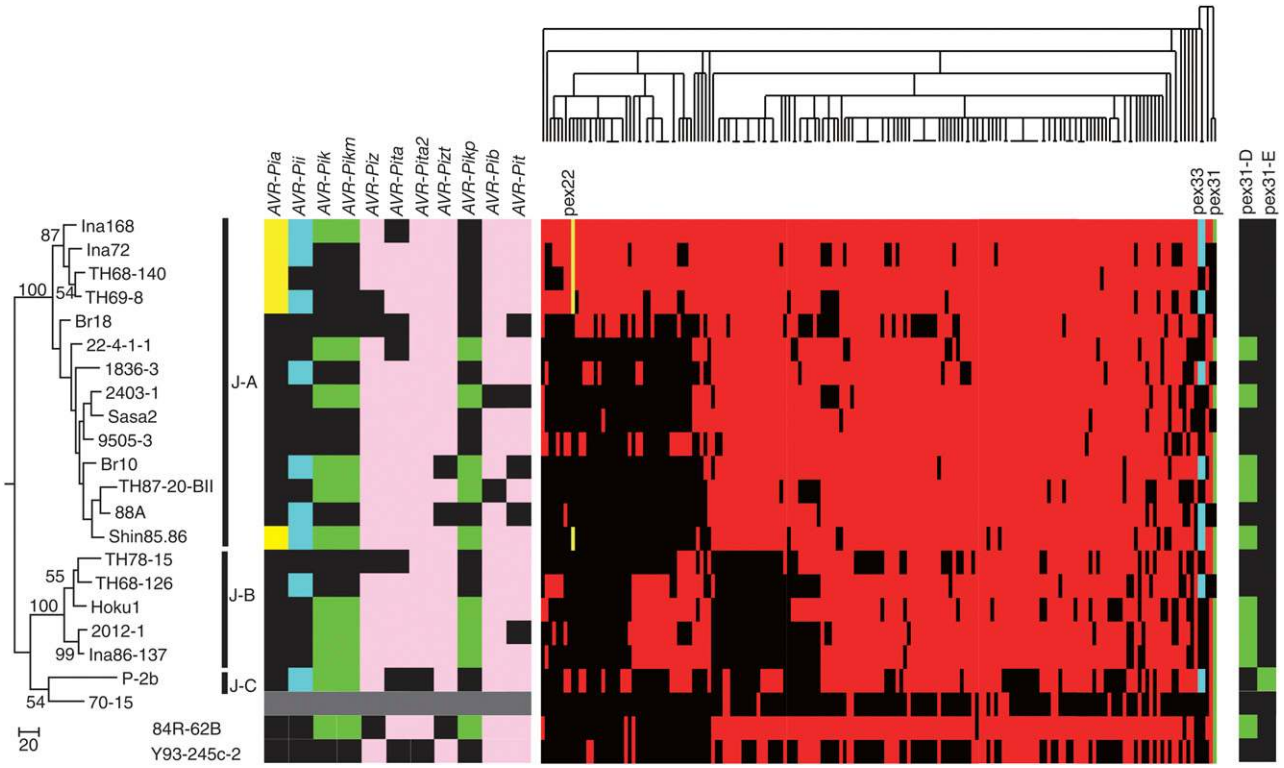

Figure 1. Distribution of *AVR* Genes and Ina168pex ORFs among the 23 Isolates of *M. oryzae* Rice Pathogen.

A phylogenetic tree of 21 *M. oryzae* isolates reconstructed based on 123 presence/absence and 419 nucleotide polymorphisms. Three distinct clades are indicated by J-A, J-B, and J-C (left). The tree was rooted by an outgroup (finger millet isolate of *M. oryzae*). The numbers indicate bootstrapping probabilities after 100 replications. Presence or absence of an avirulence gene is indicated by a colored or black tile for the nine *AVR* genes across the 23 *M. oryzae* isolates. Gray tiles indicate that the presence/absence of *AVR* was not determined (middle). Presence or absence of each of the Ina168pex ORFs is indicated by a colored or black tile for the 23 *M. oryzae* isolates (right). Presence/absence patterns were hierarchically clustered (top).

(Figure 4B). Transcription of the *pex31* gene during Ina168 infection of leaf sheath cells of susceptible rice cultivar Shin-2 was confirmed by 3′-RACE RT-PCR (see Supplemental Figure 4 online). The predicted pex31 protein sequence showed no similarity to known protein domains. DNA sequencing of amplified PCR products revealed five alleles, A to E, differing by nonsynonymous nucleotide substitutions corresponding to one to four amino acid changes (Figure 4D). The level of nucleotide variation (θ) in the pex31 ORF (339 bp) among the five alleles was 7.1×10^{-3}, which was two orders of magnitude higher than the nucleotide polymorphism levels shown by EcoTILLING of 1032 secreted protein genes ($\theta = 8.2 \times 10^{-5}$). Since we did not detect any synonymous changes among the alleles we examined, the number of synonymous changes in this locus must be extremely small. Thus, the average ratio between the number of nonsynonymous substitutions per nonsynonymous site and the number of synonymous substitutions per synonymous site (dN/dS; Nei and Gojobori, 1986) among the alleles approaches infinity, suggestive of a strong positive selection imposed on this DNA fragment. Among the five changes showing nonsynonymous variations, three caused amino acid changes involving charge differences, 46^{N} (neutral) to 46^{H} (positive), 48^{D} (negative) to 48^{G} (neutral), and 67^{A} (neutral) to 67^{D} (negative) (Figure 4D). Some isolates contained two alleles A and D (Figure 4B). The isolates with the alleles D or E invariably harbored *AVR-Pik/km*, and the isolates with the allele D invariably had *AVR-Pikp* (Figure 4B). The isolates without the pex31 ORF and the isolates with pex31 alleles A, B, and C corresponded to the isolates lacking *AVR-Pik/km* except for Ina168. The outcome of Ina168 infection of plants carrying the *Pik*/*Pik-m R* genes is variable between experiments and may be affected by environmental conditions (thus, *AVR-Pik*/*Pikm* presence indicated as +/−).

Genetic Complementation of *M. oryzae* AVR Phenotypes

To test whether pex22 is indeed *AVR-Pia*, we performed genetic complementation in the *M. oryzae* isolate Ina86-137 that lacks *AVR-Pia* and thus can infect rice cultivars possessing the *Pia* R gene (Figure 2D). A 2.2-kb fragment of contig264 harboring pex22 and a 1.7-kb fragment containing only pex22 and its promoter region (22p:pex22) were used for the transformation of Ina86-137. In contrast with wild-type Ina86-137, transformants failed to cause disease in the rice cultivar Sasanishiki possessing *Pia* (Figure 2D). Both the wild-type strain and the transformants successfully infected rice cultivar Shin-2 that lacks *Pia*, suggesting that their differences in infecting cv Sasanishiki were *Pia*

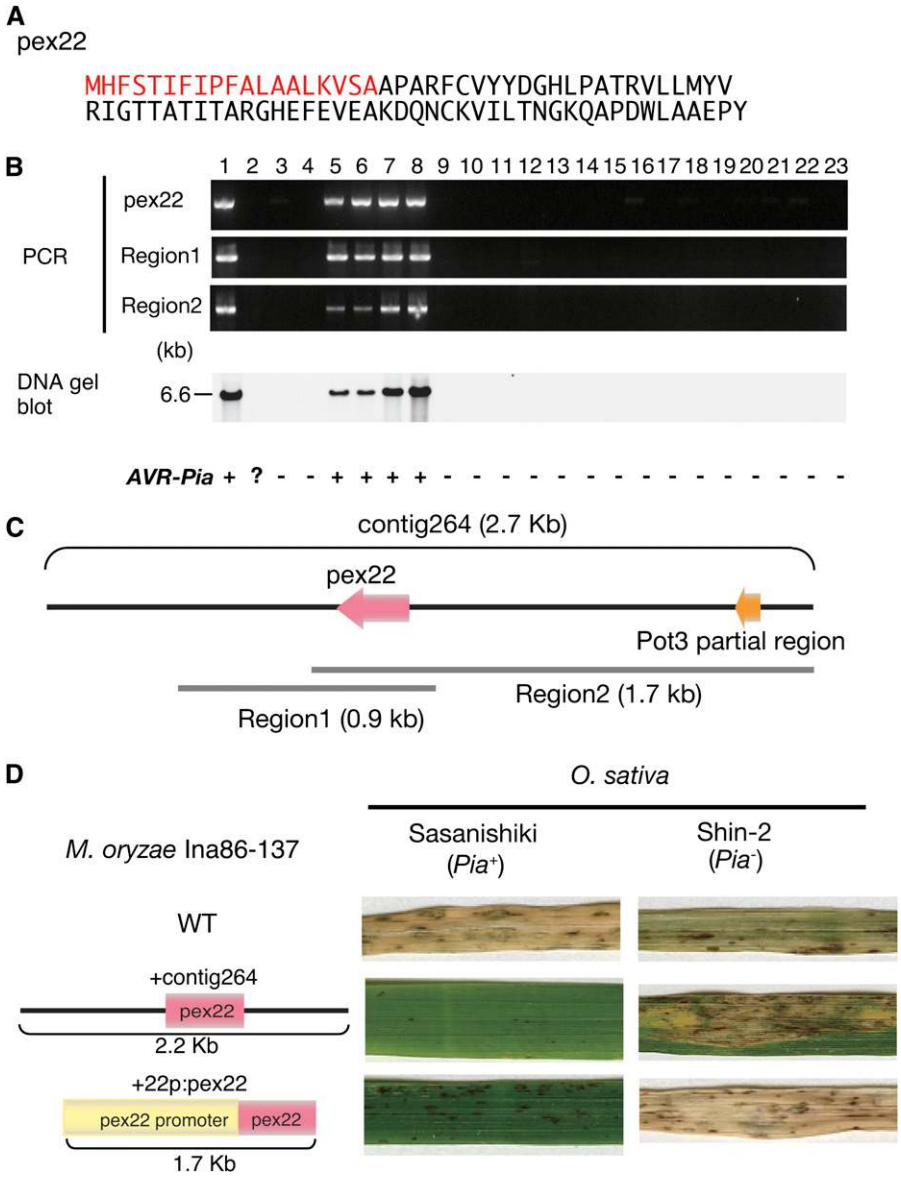

Figure 2. pex22 ORF Is *AVR-Pia*.

(A) Amino acid sequence of pex22 (85 amino acids). The predicted signal peptide is indicated in red.
(B) PCR amplification and DNA gel blot analysis of pex22 ORF. "pex22" corresponds to the pex22 ORF region; "Region 1" and "Region 2" correspond to ORF plus downstream and ORF plus upstream region, respectively, of pex22.
(C) A diagram of contig264 harboring pex22. Positions of Regions 1 and 2 are indicated.
(D) Results of interaction between *M. oryzae* and rice. The isolate Ina86-137 does not have AVR-Pia function and thus can cause disease on Sasanishiki harboring the *R* gene *Pia*. Ina86-137 strains transformed with contig264 (+contig264) or the fragment containing only the pex22 promoter region and ORF (+22p:pex22) became incompatible with Sasanishiki. Both Ina86-137 wild type, Ina86-137 containing contig264, or 22p:pex22 were able to cause disease on a rice cultivar Shin-2 lacking *Pia*, suggesting that the effect of transformation with contig264 and 22p:pex22 is *Pia* dependent.

dependent. Active transcription of the pex22 transgene in the contig264 transformant during infection was confirmed by RT-PCR (see Supplemental Figure 5 online). The same genetic complementation result was obtained with the transformation of another isolate TH68-141 (see Supplemental Figure 6A online). These results suggest that pex22 is actually *AVR-Pia*.

We also conducted similar experiments to validate our identification of pex33/279 as *AVR-Pii*. *M. oryzae* isolate Sasa2 lacking *AVR-Pii* is virulent to the rice cultivar Kakehashi harboring *Pii R*-gene. Sasa2 transformed with the 1.4-kb fragment derived from contig389 became avirulent on the cultivar Kakehashi (Figure 3E). This phenotypic change of *M. oryzae* is *Pii*

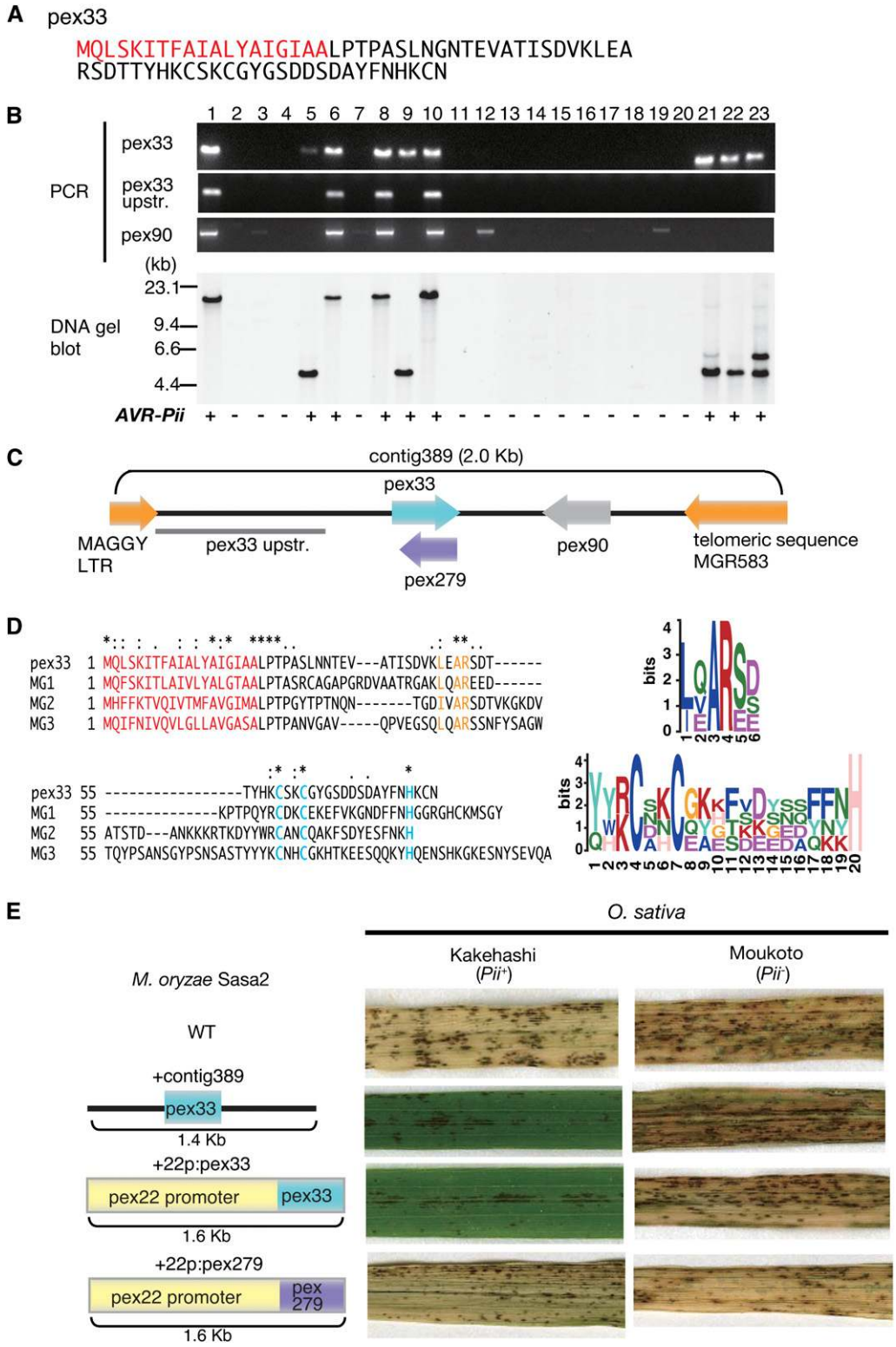

Figure 3. pex33 ORF Is *AVR-Pii*.

(A) Amino acid sequence of pex33 (70 amino acids). The predicted signal peptide is indicated in red.
(B) PCR amplification and DNA gel blot analysis of pex33 ORF. “pex33” corresponds to the pex33 ORF region, “pex33upstr.” to the upstream region of pex33, and “pex90” to an ORF pex90 situated at the downstream of pex33.

dependent since both the wild type and the transformant could infect rice cultivar Moukoto that lacks *Pii*. To determine which of pex33 or pex279 is *AVR-Pii*, we made two constructs in which the promoter region of pex22 was fused to pex33 ORF (resulting in 22p:pex33) or pex279 ORF (22p:pex279). Only 22p:pex33 conferred avirulence to Sasa2 on cv Kakehashi, indicating that pex33 is *AVR-Pii* (Figure 3E). Active transcription of pex33 transgene in the tranformants during infection was confirmed by RT-PCR (see Supplemental Figure 5 online). Similar genetic complementation results were obtained by transformation of another isolate Ina86-137 lacking *AVR-Pii* (see Supplemental Figure 6B online).

To test whether the D allele of pex31 ORF (pex31-D) can function as *AVR-Pik*/*km* and *AVR-Pikp*, we transformed an isolate Sasa2 lacking *AVR-Pik*, *AVR-Pikm*, and *AVR-Pikp* with the DNA fragment containing pex31-D (Figure 4E). Two versions of transgenes were used: 2.2-kb genomic region harboring pex31-D and its promoter region (pex31-D-genome) and pex31-D allele fused with the promoter region of pex22 (22p:pex31-D). These transformants harboring pex31-D allele as well as the wild-type Sasa2 isolate were used for inoculation to the four rice cultivars Kanto51 (*Pik*$^+$), Tsuyuake (*Pik-m*$^+$), K60 (*Pik-p*$^+$), and Shin-2 (*Pik*$^-$, *Pik-m*$^-$, *Pik-p*$^-$). In the two transformants harboring pex31-D-genome and 22p:pex31-D, pex31-D transgene was actively transcribed during infection in rice as revealed by RT-PCR (see Supplemental Figure 5 online). The two transformants and Sasa2 could infect the rice cultivar Shin-2, whereas the two transformants could not infect Kanto51, Tsuyuake, and K60 (Figure 4E), suggesting that pex31-D can function as *AVR-Pik*, *AVR-Pikm*, and *AVR-Pikp*, recognized by the *R* genes *Pik*, *Pik-m*, and *Pik-p*, respectively. Similar genetic complementation results were obtained by transformation of another isolate Ina72 lacking *AVR-Pik*, *AVR-Pikm*, and *AVR-Pikp* (see Supplemental Figure 6C online). Based on these results, we indicate that pex31-D corresponds to *AVR-Pik*/*km*/*kp* in the following explanation.

Expression of the *AVRs* in Rice Protoplasts from Cultivars Containing the Cognate *R* Genes Results in Cell Death

Complementation experiments in *M. oryzae* indicated that pex22, pex33, and pex31-D are *AVR-Pia*, *AVR-Pii*, and *AVR-Pik*/*km*/*kp*, respectively. To determine whether the expression of these ORFs is sufficient to trigger the HR-like cell death in rice possessing the corresponding *R* genes, we performed transient expression of the ORFs in rice protoplast cells (Figure 5A) following the method described by Chen et al. (2006). Each of pex22, pex33, and pex31-D was cloned into a plasmid downstream of the maize (*Zea mays*) ubiquitin-1 gene promoter (ubi-p) (Christensen and Quail, 1996) to serve as an effector plasmid. We created two versions of effector plasmids: an ORF lacking the signal peptide (indicated by -ns) and that with the signal peptide (-s). Proteins without signal peptide (-ns) are retained within rice cells. Fungal secretion signal peptides are known to function in plants (Catanzariti et al., 2006), so that the proteins with *M. oryzae* signal peptides (-s) are assumed to be targeted to the outside of rice cells. To monitor cell viability, we created another plasmid containing the firefly luciferase gene driven by ubi-p (*luc* plasmid). The effector and *luc* plasmids were mixed in an equal ratio and used for the transformation of rice protoplasts. Cell viability was monitored by luciferase activity 40 h after transformation.

The effector plasmid harboring pex22-ns caused a significant reduction in cell viability compared with the empty plasmid when the rice cultivar Sasanishiki harboring the *Pia R* gene was used for transformation (Figure 5B). A similar result was obtained when Akitakomachi, another cultivar with *Pia*, was tested (see Supplemental Figure 7B online). This pex22-mediated cell death was not observed when protoplasts of the rice cultivar Kakehashi lacking the *Pia R* gene were transformed (Figure 5B), suggesting that the observed cell death is *Pia* specific. This result confirms that pex22 is indeed *AVR-Pia* that is recognized by *Pia*. Both pex22-ns and pex22-s effectively killed rice cells 40 h after the transformation (Figure 5F). Likewise, pex33-ns caused cell death in rice protoplasts harboring *Pii* but not in rice cells lacking *Pii* (Figure 5C; see Supplemental Figure 7B online), while pex31-D-ns caused cell death in rice cells harboring *Pik* but not in the cells lacking *Pik* (Figure 5D). These results confirm that pex33 is *AVR-Pii*, and pex31-D is *AVR-Pik*, recognized by the rice *R* genes *Pii* and *Pik*, respectively. Finally, for *AVR-Pia*, *AVR-Pii*, and *AVR-Pik*, ORFs without signal peptide caused cell death, suggesting that they can be recognized inside rice cells. pex31 has five alleles, A to E (Figure 4D). pex31-D-ns caused cell death both in *Pik* and *Pik-p* rice cells (Figure 5E), whereas pex31-E-ns, differing from pex31-D-ns by only one amino acid change (46H → N) (Figure 4D), caused cell death in *Pik* rice cells but not in *Pik-p* rice cells, suggesting that this amino acid change determines the recognition specificity by the *R* gene *Pik-p*. Furthermore, pex31-C-ns differing from pex31-E-ns by only one amino acid (67A → D) caused cell death neither in *Pik* nor *Pik-p* rice cells, suggesting

Figure 3. (continued).

(C) A diagram of contig389 harboring pex33.

(D) Amino acid alignment of pex33 and three putative secreted protein genes of *M. oryzae*: MG1, XP_366338.2; MG2, XP001407225.1; and MG3, XP_364190.1. Putative signal peptide is indicated in red. Conserved residues (motif-1 and motif-2) are indicated in orange and blue. LOGO representation of the conserved motif-1 (right top) and motif-2 (right bottom).

(E) Results of interaction between *M. oryzae* and rice. The isolate Sasa2 can cause disease on a rice cultivar Kakehashi harboring the *R* gene *Pii*. Sasa2 transformed with contig389 (+contig389) as well as Sasa2 transformed with a fragment containing the pex22 promoter fused with pex33 ORF (+22p:pex33) were incompatible with Kakehashi, whereas Sasa2 transformed with pex279 driven by the pex22 promoter (+22p:pex279) was compatible with Kakehashi. Sasa2, [Sasa2(+contig389*)], and [Sasa2(+22p:pex33)] are all compatible with a rice cultivar Moukoto lacking *Pii*, suggesting that the effect of transformation with contig389 and 22p:pex33 is *Pii* dependent.

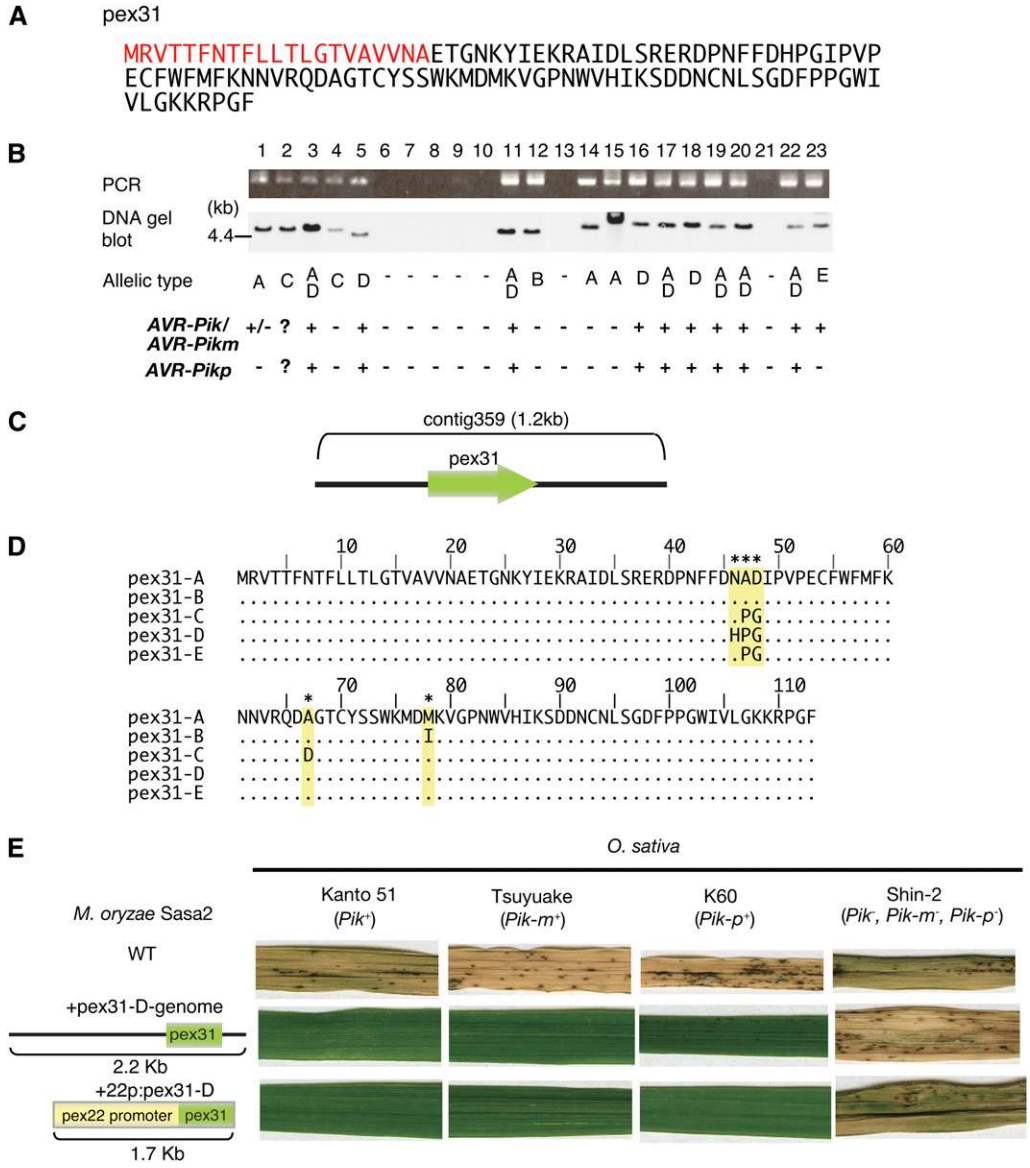

Figure 4. pex31 Is Associated with *AVR-Pik/km/kp*, and pex31-D Functions as *AVR-Pik/km/kp*.

(A) Amino acid sequence of pex31 (113 amino acids). The predicted signal peptide is indicated in red.
(B) Presence/absence of PCR products for pex31, DNA gel blot result with pex31 ORF as probe, allelic type of pex31 possessed by each isolate, and presence/absence of *AVR-Pik/km* and *AVR-Pikp* in the tested isolates.
(C) A diagram of 1.2-kb contig359 harboring pex31 ORF.
(D) Amino acid sequences of pex31 alleles A to E. Number of nonsynonymous polymorphic sites is 5; number of synonymous polymorphic sites is 0.
(E) Genetic complementation of Sasa2 with pex31-D fragment. The isolate Sasa2 is compatible with four rice cultivars, Kanto51 (*Pik*$^+$), Tsuyuake (*Pik-m*$^+$), K60(*Pik-p*$^+$), and Shin-2 (*Pik*$^-$, *Pik-m*$^-$, and *Pik-p*$^-$). Sasa2 strains transformed with the 2.2-kb genomic fragment containing pex31-D allele (+pex31-D-genome) or with a fragment containing pex31-D ORF driven by the pex22 promoter (+22p:pex31-D) were incompatible with Kanto51, Tsuyuake, and K60. Sasa2, [Sasa2(+pex31-D-genome)], and [Sasa2(+22p:pex31-D)] are all compatible with a rice cultivar Shin-2 lacking *Pik*, *Pik-m*, and *Pik-p*.

that this amino acid change determines the recognition specificity by *Pik*. pex33-ns was more effective in causing cell death than was pex33-s (Figure 5F). Cell death caused by pex31-D-ns appeared to be quicker than that by pex31-D-s 40 h after transformation, although the difference was not significant (Figure 5F).

Conditional Expression of the *AVRs* in Rice Plants Harboring the Cognate *R* Genes Causes Cell Death

To further confirm that the interaction between pex22 and *Pia* in rice leaves triggers HR-like cell death, we made stable transgenic rice plants harboring pex22 ORF in a conditional

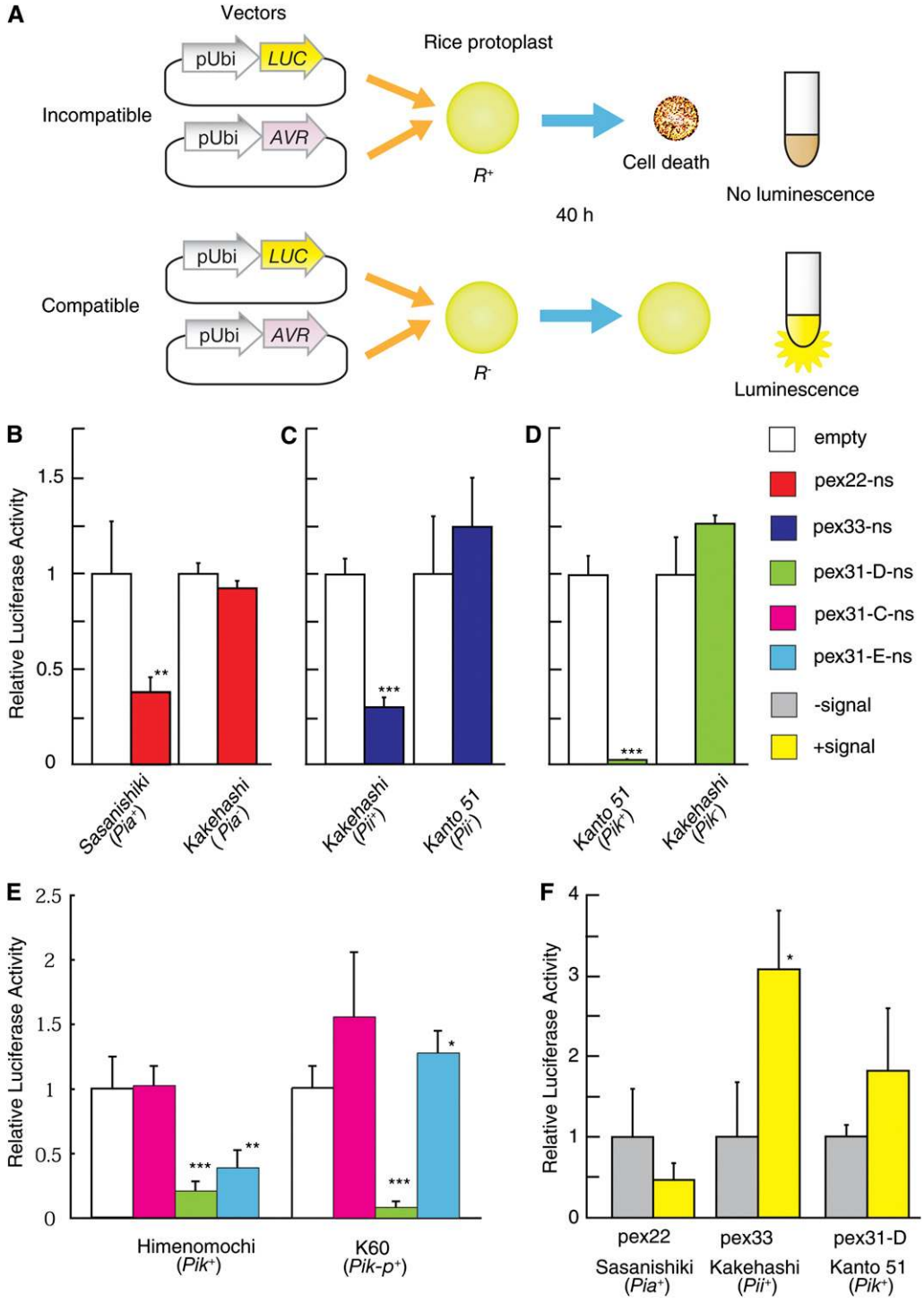

Figure 5. Interaction between AVR Candidates and *R* Genes Causes Cell Death in Rice Protoplasts.

(A) Scheme of experimental protoplast transformation of rice. Interaction between *R* gene and AVR causes rapid cell death, resulting in a reduction in luciferase activity, which is measured by luminescence after the addition of luciferin and ATP.

(B) to **(D)** Relative luciferase activity after transformation with pex22-ns **(B)**, pex33-ns **(C)**, and pex31-D-ns **(D)**. Rice cells with (left) and without (right) cognate *R* genes were tested.

(E) Relative luciferase activity after transformation with three alleles of pex31: pex31-C, pex31-D, and pex31-E.

(F) Comparison of luciferase activity after transformation with AVR candidates with (yellow column) or without (gray column) signal peptides for pex22 (left), pex33 (center), and pex31-D (right). Average values of three or more replications per experiment are shown with SD. Statistical significance is indicated by *, 0.05; **, 0.01; ***, 0.001.

expression vector GVG (Aoyama and Chua, 1997). A gene in the GVG vector is normally suppressed but is strongly induced upon treatment with a glucocorticoid, dexamethasone (DEX). Two versions of pex22 transgenes were prepared: the full ORF containing regions for the signal peptide and mature protein (pex22-s) and N terminus truncated version in which the region corresponding to the signal peptide was truncated (pex22-ns). *Pia*$^{+}$ rice plants harboring GVG-pex22-s and GVG-pex22-ns grew normally. After treatment with DEX, pex22 transcription was induced (Figure 6A). Following the expression of pex22, the plants started to exhibit cell death as indicated by trypan blue staining in leaf blades (Figure 6A) as well as rapid desiccation of leaves (Figure 6A). Transgenic rice cultivar Hitomebore of *Pia*$^{-}$ background overexpressing pex22-s and -ns did not show any cell death phenotype (Figure 6A). These results suggest that pex22 protein is recognized by *Pia* to cause HR-like cell death in rice leaves. It is notable that the cell death symptoms consistently occur more quickly in pex22-ns than in pex22-s transgenic rice, supporting the idea that pex22 is recognized by *Pia* inside rice cells.

Similarly, conditional overexpression of pex33-s and pex33-ns in the rice cultivar Hitomebore harboring *Pii*, as confirmed by an RT-PCR experiment with pex33-specific PCR primers (Figure 6B), caused HR-like cell death, whereas transgenic rice cultivar Sasanishiki of *Pii*$^{-}$ background overexpressing pex33-s and pex33-ns did not give cell death (Figure 6B), suggesting that interaction between pex33 and *Pii* in rice leaves caused HR-like cell death.

We also tested pex31-D in rice leaves. Conditional overexpression of pex31-D-s in a rice cultivar Kanto51 harboring *Pik R*-gene, as confirmed by an RT-PCR experiment with pex31-specific PCR primers (Figure 6C), caused cell death, whereas pex31-D-s expression in a rice cultivar Sasanishiki lacking *Pik* did not (Figure 6C), confirming that pex31-D is recognized by *Pik* in the rice plant and causes HR-like cell death.

Polymorphic Candidate Effector Genes Are Associated with Transposable Elements

Transposon and telomeric sequences were noted in the genomic regions neighboring *AVR-Pia* and *AVR-Pii* (Figures 2C and 3C) and are known to enhance the likelihood of gene loss and horizontal transfer (Silva et al., 2004; Rehmeyer et al., 2006). To test whether the presence of transposable elements affect the polymorphism levels among the 1032 candidate effector genes identified in the genomic sequence of isolate 70-15, we compared the levels of presence/absence polymorphisms of PCR products among the 23 *M. oryzae* isolates between the genes linked and unlinked to transposable elements (Figure 7). Putative secreted protein genes linked to transposons within a distance of 5 kb exhibit significantly higher levels of presence/absence polymorphisms than those unlinked to transposons, suggesting that the linkage to transposons could enhance the likelihood of gain/loss for a given candidate effector gene (Figure 7B). A similar significant difference was observed even when a narrower window of distance (2 kb) between the candidate effector genes and transposons was applied (Figure 7C).

DISCUSSION

Comparison of *M. oryzae* 70-15 and Ina168 Genome Sequences

In this study, we aimed to identify novel avirulence genes from the rice blast pathogen *M. oryzae* using an association genetics approach. Our initial attempts to identify associations using polymorphisms in secreted protein genes mined from the sequence of isolate 70-15 failed. We subsequently hypothesized that 70-15 lacks the targeted *AVR* genes and we sequenced the genome of Ina168, an *M. oryzae* isolate known to contain nine *AVR* genes. Remarkably, 1.68 Mb of genomic DNA were found in Ina168 but absent from 70-15 draft sequences. Conversely, 5.09 Mb regions of 70-15 were lacking from our shotgun sequence of Ina168. In view of the result that 113 out of 316 pex ORFs on the 1.68 Mb Ina168-unmapped genome could be PCR-amplified from 70-15, this discrepancy may partly be explained by incompleteness of the 70-15 draft sequence and the Ina168 shotgun sequence. The total length of all sequence contigs of 70-15 is 37.8 Mb (Dean et al., 2005), whereas that of Ina168 was 38 Mb (Table 3). The actual size of *M. oryzae* genome may be around 40 Mb (Dean et al., 2005). It is also likely that substantial portions of their genomes are not shared between the two isolates. This discordance may be related to the origin of 70-15 isolate, which was derived from a cross between isolate 104-3 from rice and isolate AR4 from weeping lovegrass, followed by subsequent backcrosses to rice isolate Guy11 (Chao and Ellingboe, 1991). It is possible that 70-15 draft genome sequences may lack genomic regions specific to rice pathogens but in turn contain regions specific to weeping lovegrass pathogen. Ambiguous reactions between 70-15 and rice cultivars (see Supplemental Figure 1 online) also point to the possibility that the 70-15 genome lacks *AVR*s recognized by rice *R* genes but contain genes whose products trigger nonhost resistance reactions in rice. These observations imply that further genome sequencing from multiple isolates of *M. oryzae* is absolutely required for characterizing rice pathogenic strains, as key information can only be found by going beyond the "type" genome.

To determine whether Ina168 contains a supernumerary chromosome (minichromosome) that may correspond to the Ina168-specific genomic regions, we performed a pulsed field gel electrophoresis of Ina168 genome DNA (see Supplemental Figure 8 online). Ina168 genome did not contain a supernumerary chromosome (< 2.2 Mb) as described by Talbot et al. (1993) and Chuma et al. (2003), suggesting that the Ina168-specific sequences are not located on such supernumerary chromosomes.

Identification of *AVR-Pia*, *AVR-Pii*, and *AVR-Pik/km/kp*

Presence/absence polymorphisms of three Ina168 ORFs, pex22, pex33, and pex31, tightly associated with *AVR-Pia*, *AVR-Pii*, and *AVR-Pik/km/kp* functions, respectively (Figure 1). Genetic complementation of *M. oryzae* showed that these ORFs indeed function as the respective avirulence determinants (Figures 2 to 4). Furthermore, ectopic expression of pex22, pex33, and pex31-D caused rapid cell death in rice protoplasts (Figure 5) as well as in rice plants (Figure 6) that harbor *Pia*, *Pii*, and *Pik*,

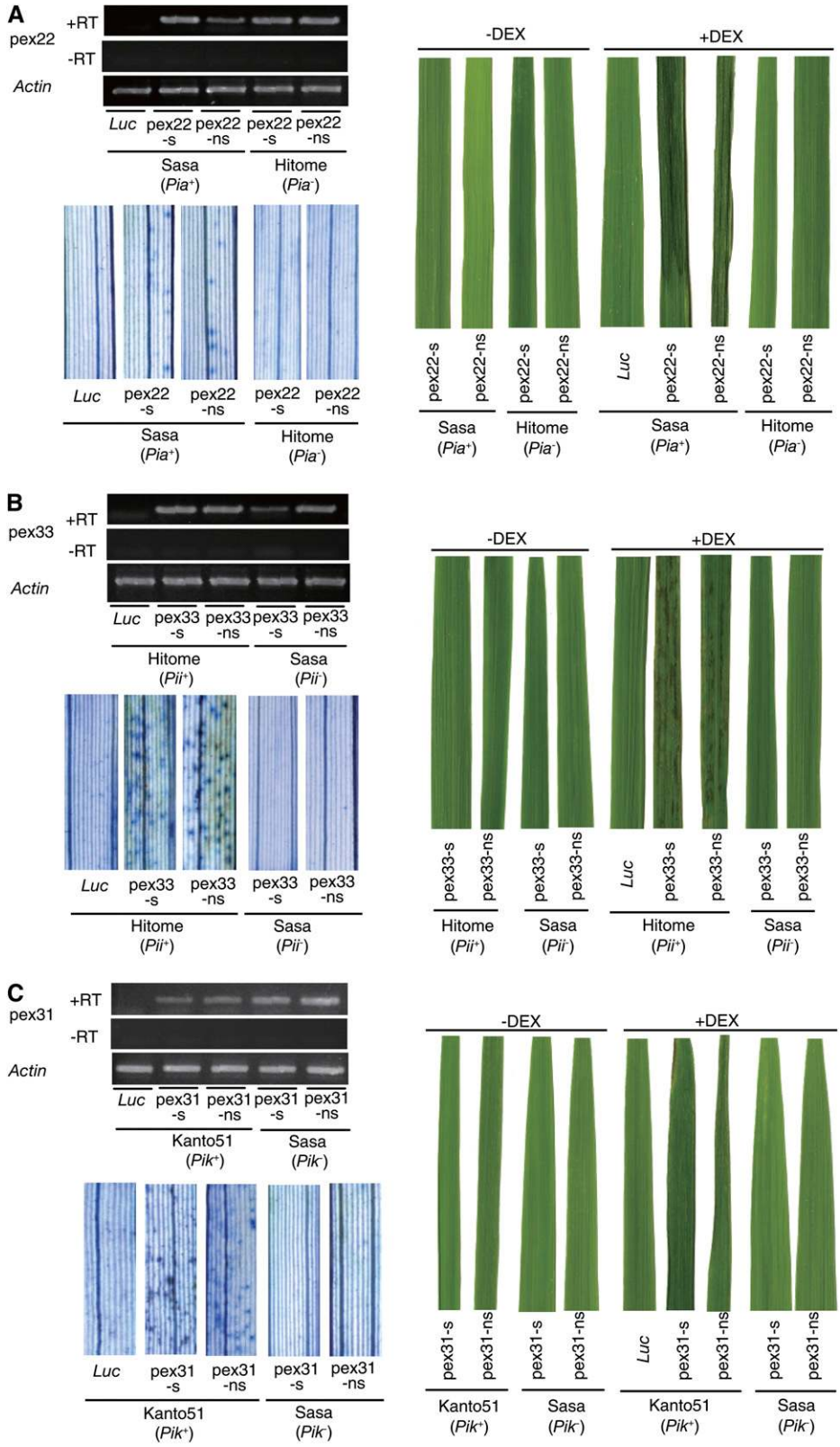

Figure 6. Overexpression of pex22, pex33, and pex31-D in Rice Leaves of the Cultivars Harboring *Pia*, *Pii*, and *Pik*, Respectively, Causes HR-Like Cell Death.

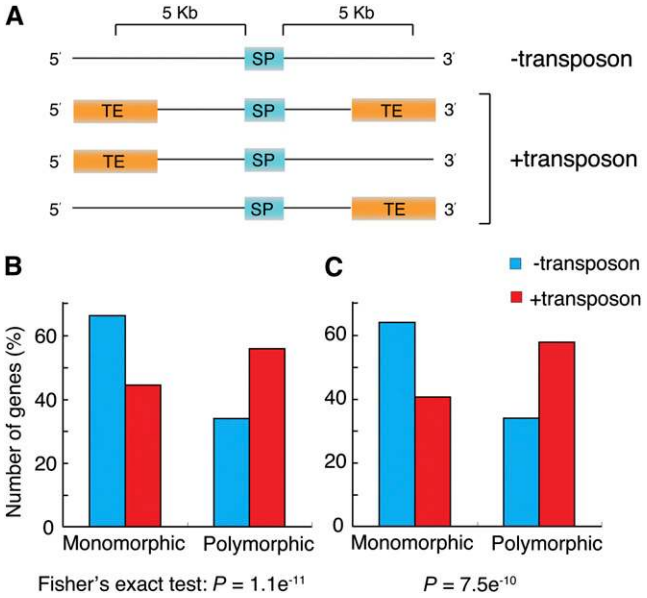

Figure 7. Candidate Effector Genes Linked to Transposable Elements Tend to Be More Polymorphic Than Those Unlinked.

(A) A scheme of linkage of putative secreted protein genes (SP = candidate effector genes) to transposable elements (TE). Putative secreted protein genes (total: 1032) could be categorized to those without transposons (–transposon; 610 genes) or with transposons (+transposon; 422 genes) depending on the presence or absence of transposon sequences within a 5-kb distance from the gene.

(B) Percentage of putative secreted protein loci showing monomorphism (left) or polymorphism (right) with respect to presence/absence polymorphisms. Blue column: putative secreted protein genes without transposons within 5-kb distance. Red column: putative secreted protein genes with transposons within 5-kb distance. A total of 1032 loci predicted from the 70-15 genome sequence were tested. Difference in the level of polymorphisms between –transposon and +transposon putative secreted protein genes is statistically significant ($P = 1.1e^{-11}$).

(C) Same as **(B)** with the distance between secreted protein genes and transposon set to 2 kb. Difference in the level of polymorphisms between –transposon and +transposon putative secreted protein genes is statistically significant ($P = 7.5e^{-10}$). A list of the transposable elements studied is given in Supplemental Table 4 online.

respectively. Altogether, these results demonstrate that pex22 is *AVR-Pia*, pex33 is *AVR-Pii*, and pex31-D functions as *AVR-Pik* (*M. oryzae* transformation and rice transformation results; Figures 4 to 6) as well as *AVR-Pikm* and *AVR-Pikp* (*M. oryzae* transformation result; Figure 4).

The AVR proteins identified are all small, comprising 85 (pex22), 70 (pex33), and 113 (pex31) amino acids. These proteins are smaller than previously reported AVRs of *M. oryzae*: AVR-Pita (223 amino acids; Orbach et al., 2000), PWL proteins (138 to 147 amino acids; Kang et al., 1995). Pex22 and pex31 do not show similarity to any known proteins, whereas pex33 exhibits partial similarity to three *M. oryzae* putative secreted proteins of unknown functions (Figure 3D). Judging from their small sizes and the lack of similarity to known enzymes, we propose that these effector proteins function by physically interacting with other proteins (most probably host proteins) but not by directly mediating a catalytic activity.

Pex33-Like Family in the *M. oryzae* Genome

Pex33 protein showed substantial similarity to at least three *M. oryzae* predicted proteins (XP_366338.2, XP_001407225.1, and XP_364190.1), particularly in the signal peptide (Figure 3D). The four homologs share two conserved amino acid motifs, motif-1 and motif-2, so that it is reasonable to hypothesize that their genes were derived from a common ancestral gene. We tentatively call these proteins the pex33 protein family. A higher amino acid sequence conservation in the signal peptide in contrast with divergent mature protein region has been observed in pathogen effectors (Dodds et al., 2004; Liu et al., 2005; Catanzariti et al., 2006). This pattern could be caused by a strong positive selection in the mature protein regions as opposed to purifying selection imposed on the signal peptide region. Alternatively, recombination among the pex33 paralogs could have given rise to the distinct amino acid sequences in the mature protein region. In any case, we hypothesize that the high level of divergence in the pex33 mature protein region was driven by coevolution with host rice plants. The conserved motifs in pex33 protein family are intriguing. The motif-1 ([LI]xAR[SE][DSE]) corresponds to the LxAR motif recently reported in AVR-Piz-t and other candidate effectors (Bo Zhou, personal communication). The motif-2 ([RK]CxxCxxxxxxxxxxxH) is reminiscent of C2H2 zinc finger motif involved in protein–protein interaction. Among the 1306 *M. oryzae* putative secreted proteins, we found other 11 proteins possessing this motif-2 (see Supplemental Table 2 online), which is mostly located in the C-terminal region. Possibly, motif-2 in these candidate effectors may mediate binding to host target molecules with a shared structure.

Pex31 Allelic Divergence and *R* Gene Recognition Specificity

The patterns of polymorphisms are very different between pex22/pex33 and pex31. pex22 and pex33 show presence/absence polymorphisms without nucleotide polymorphisms, whereas pex31 exhibits both presence/absence and nucleotide polymorphisms. pex31 contained five allelic variants, A to E. DNA

Figure 6. (continued).

Results of conditional overexpression of pex22 **(A)**, pex33 **(B)**, and pex31-D **(C)**. Left top: RT-PCR results showing that ORFs are expressed in transgenic rice leaves after treatment with DEX. Left bottom: trypan blue stain of rice leaves treated with DEX. Right: leaf phenotypes before (−DEX) and after (+DEX) treatment with DEX. -s, ORF with signal peptide region; -ns, ORF without signal peptide region. Abbreviation of cultivar names: Hitome, cv Hitomebore; Sasa, cv Sasanishiki.

polymorphisms in pex31 are all nonsynonymous substitutions, providing strong evidence of positive selection. Moreover, the amino acid differences among the alleles A to E are functionally relevant in the *M. oryzae*–rice interaction. In the rice protoplast transformation assay, pex31-D was recognized by both *Pik* and *Pik-p*, whereas pex31-E was recognized only by *Pik* (Figure 5E) and pex31-C neither by *Pik* nor *Pik-p*. It is likely that pex31 evolution by nonsynonymous substitution is driven by the recognition specificity of rice *R* genes. Indeed, linkage analyses indicate that *Pik*, *Pikm*, and *Pikp R* genes are all located on a similar position of rice chromosome 11, and these *R* genes may be alleles of the same locus or linked paralogs (Hayashi et al., 2006; Ashikawa et al., 2008). This situation is similar to that observed between the *RPP13 R* gene locus of *Arabidopsis thaliana* and *ATR13* of *Peronospora parasitica* (Allen et al., 2004; Rose et al., 2004). It is possible that dispensing with pex31 may have a fitness penalty in *M. oryzae*, so that the fungus is under selection to keep the gene, and evasion of host recognition evolves by rapid amino acid substitutions. Only when the fungus is exposed to rice cultivars that possess all the cognate *R* genes will the fitness gain by losing pex31 be higher than the cost. In such circumstances, *M. oryzae* may adapt to the host by losing pex31.

Recognition of AVR-Pia, AVR-Pii, and AVR-Pik/km/kp Occurs inside Rice Cells

Overexpression of *AVR-Pia*, *AVR-Pii*, and *AVR-Pik/km/kp* constructs lacking the signal peptide in rice cells possessing cognate *R* genes, *Pia*, *Pii*, and *Pik*, respectively, caused rapid cell death (Figures 5B to 5D and 6). This information suggests that the AVRs produced in rice cells are effectively recognized by R proteins. This idea conforms to the recent reports that *R* genes corresponding to the isolated AVRs, *Pik-m* (Ashikawa et al., 2008), *Pii* (Lee et al., 2009), and *Pia* (Y. Okuyama and R. Terauchi, unpublished data), all encode nucleotide-binding site–leucine-rich repeat (NBS-LRR) class R proteins localized inside the rice cytoplasm. The observation that AVRs without signal peptides are equally (AVR-Pia and AVR-Pik) or more (AVR-Pii) effective than those with signal peptides in triggering cell death in rice protoplasts with cognate *R* genes (Figure 5F) is consistent with the hypothesis that all three AVRs are targeted inside rice cells. More enhanced cell death in transgenic rice expressing AVR-Pia without signal peptide than that with signal peptide (Figure 6A) also supports this hypothesis. These observations imply that these AVRs translocate inside rice cells prior to the recognition by their cognate R protein. Future studies by truncation and mutagenesis of the *AVR* genes should clarify the motifs necessary for these AVRs to enter rice cells.

Association Study Identified *M. oryzae AVR* Genes

We employed a PCR-based association genetics approach to identify three *AVR* genes, which were verified by *M. oryzae* complementation and rice transformation studies. Actually, the pex31-D ORF corresponded to the three *AVR*s, *AVR-Pik*, *AVR-Pikm*, and *AVR-Pikp*, indicating successful cloning of five *AVR*s out of nine *AVR*s previously described in the Ina168 isolate, illustrating the efficiency of association analysis in isolation of *AVR*s. Presence/absence polymorphisms as detected by PCR in the three *AVR*s were all verified by DNA gel blot analysis (Figures 2 to 4), suggesting that a rapid PCR survey is effective in addressing the presence/absence of DNA regions in the *M. oryzae* genome. Among the 21 putative secreted protein genes that were predicted in the 70-15 genomic sequence but failed to PCR amplify in Ina168, 17 (81%) genes were genuinely absent from Ina168 sequence reads with the threshold of $E > 2e^{-35}$. We propose that the majority of the polymorphisms detected by PCR indeed reflect presence/absence of the genomic regions. So far, substantial effort has been expended to isolate *AVR* genes by linkage analysis of progeny derived from a sexual cross between isolates with and without a certain *AVR* gene. Apart from a few successful cases, including that of *AVR-Pita* (Orbach et al., 2000), these approaches have not been very fruitful, presumably because of the difficulty in chromosome walking caused by repetitive sequences and in isolating a plasmid clone containing the gene from a genomic library. Association studies do not require genetic crossing, so this approach is applicable to any organism, even those not amenable to linkage analysis. It is interesting to note that the *AVR-Pita* locus (Orbach et al., 2000; Khang et al., 2008) was PCR amplified for all the 21 isolates (see Supplemental Data Set 1 online) but contained a high level of nucleotide polymorphism ($\theta = 0.0029$). One of the nucleotide polymorphisms detected by EcoTILLING exhibited a significant association with the *AVR-Pita* phenotype (see Supplemental Figure 9 online), suggesting that the association genetics approach would have successfully identified *AVR-Pita*. An association study has also been successfully used to isolate genes causing drug resistance in prokaryotes (Andries et al., 2005). In filamentous plant pathogens, *Avr3A* in *Phytophthora infestans* was isolated by this approach (Armstrong et al., 2005). A major complication to the association study is caused by a linkage disequilibrium caused by shared phylogeny among the tested samples (Pritchard et al., 2000; Hirschhorn and Daly, 2005). *AVR-Pia*, *AVR-Pii*, and *AVR-Pik/km/kp* of *M. oryzae* are highly variable across isolates, and their distribution did not conform to the phylogeny of the core genome (Figure 1), so they were effectively in linkage equilibrium with other genomic regions. Large variability and linkage equilibrium with other regions seem to be a general feature of effectors in pathogens (see below), so association genetics should be suitable to isolate *AVR*s from other pathogens as well.

Patterns of DNA Polymorphisms in *M. oryzae* AVR Genes

Distribution of the five AVR genes, *AVR-Pia*, *AVR-Pii*, and *AVR-Pik/km/kp*, does not conform to the phylogeny of *M. oryzae* and suggests multiple gain/loss events (Figure 1). How did these genes evolve considering that sexual reproduction is extremely rare in natural populations of *M. oryzae* (Notteghem and Silue, 1992)? We hypothesize that the effectors were present in the ancestral genotype but were recently lost from a subset of lineages. Alternatively, the genes could have been gained by some isolates through lateral transfer mediated by parasexual exchange of DNA (Zeigler et al., 1997). pex22 corresponding to *AVR-Pia* is adjacent to pot3 transposase sequence, and pex33

for *AVR-Pii* is close to a telomere sequence and Maggy transposon sequence (Figures 2C and 3C). Yasuda et al. (2006) performed linkage analysis of *AVR-Pia* and *AVR-Pii* and showed that DNA markers linked to *AVR-Pia* variably located on either chromosome 5 or 7 depending on the crosses used for the linkage analysis and suggested that *AVR-Pia* is located on a region prone to chromosome rearrangement by the presence of MAGGY transposons. The same authors showed that *AVR-Pii* was tightly linked to a subtelomeric DNA marker (Yasuda et al., 2006). These observations conform to ours, and together suggest that both *AVR-Pia* and *AVR-Pii* are located on unstable chromosome regions that are variable among isolates. Localization of *M. oryzae AVRs* on unstable chromosome regions has been previously reported; *PWL* genes are situated in transposon-rich regions of the genome and are highly polymorphic (Kang et al., 1995; Sweigard et al., 1995). *AVR-Pita* is located close to the telomere of its chromosome, and loss of chromosome tips is one mechanism for frequent gain of virulence (Orbach et al., 2000; Khang et al., 2008). These findings together with ours point to a general tendency of *M. oryzae AVR* genes being located on unstable chromosome regions. It is notable that candidate effector genes linked to transposons (within a 5-kb region) tend to be more polymorphic than those not linked to transposons (Figure 7). This supports the hypothesis that the adjacent transposons help the effectors to be gained or lost during evolution of the pathogen.

Resistance and virulence vary markedly in populations of plants and plant pathogens. A negative trade-off occurs between the pathogen's ability to infect hosts with particular *R* genes and aggressiveness (i.e., the severity of disease on susceptible hosts) (Thrall and Burdon, 2003). Thus, effector genes are under fluctuating selective forces in pathogen populations to maximize infection of variable host genotypes. Many effector genes including *AVR-Pita* of *M. oryzae* display extreme levels of sequence polymorphism and positive selection (Orbach et al., 2000; Allen et al., 2004; Ellis et al., 2007; Kamoun, 2007). This study showed that *AVR-Pik/km/kp* also follows this pattern. The association of *M. oryzae* effectors with plastic genomic loci could confer an alternative mechanism of adaptation to host resistance. Indeed, localization of effectors in unstable genomic regions has been widely noted in plant (van der Does and Rep, 2007; Stavrinides et al., 2008) and animal (Gardner et al., 2002; Berriman et al., 2005), pathogens suggesting a general adaptive feature.

Low Nucleotide Diversity of *M. oryzae* Rice Pathogen

Nucleotide divergence between the genomes of *M. oryzae* Ina168 and 70-15 ($d = 2.8 \times 10^{-4}$) is consistent with the calculated nucleotide diversity for secreted protein genes ($\theta = 8.2 \times 10^{-5}$). These low levels of genetic diversity indicate that rice isolates of *M. oryzae* have a small effective population size that recently originated from a founder population. It is plausible that this spread accompanied the expansion of rice cultivation from its origin in Southeast Asia within the past 7000 years (Couch et al., 2005). Our finding that the majority of the candidate effector loci display low nucleotide diversity while frequently showing presence/absence polymorphisms indicates that, in the arms race against rice, *M. oryzae* effector genes are mainly evolving by a gain/loss process.

METHODS

Identification of Putative Secreted Protein Genes from 70-15 Genome Sequence

Amino acid sequences of a total of 11,109 predicted proteins (Dean et al., 2005) deposited as magnaporthe_grisea_2.3_proteins_fasta http://www.broad.mit.edu/cgi-bin/annotation/magnaporthe/ were used for the identification of secreted proteins (*Magnaporthe* Sequencing Project; Ralph Dean, Fungal Genomics Laboratory at North Carolina State University [http://www.fungalgenomics.ncsu.edu] and Broad Institute of MIT and Harvard [http://www.broad.mit.edu]). These sequences were applied to a pipeline of programs for prediction of secreted proteins (Torto et al., 2003). First, each sequence was tested by the SignalP program (Nielsen et al., 1997) to see whether a signal peptide is present. We retrieved the sequence when the predicted cleavage site is within 28 amino acids from the N terminus and the confidence value of the Hidden Markov Model prediction is equal to or >90%, which resulted in the selection of 1565 proteins. Next, TM-HMM (Krogh et al., 2001) was used to remove transmembrane proteins, resulting in the retention of a set of 1257 proteins. Lastly, TargetP (Emanuelsson et al., 2000) was used to remove proteins that are predicted to be targeted to mitochondria, resulting in the selection of 1206 putative secreted proteins. By application of the PSORT program (http://psort.ims.u-tokyo.ac.jp/), another 100 putative secreted protein genes were selected, making the total number of candidate secreted proteins equal to 1306.

EcoTILLING, Estimation of *M. oryzae* Nucleotide Diversity, and Phylogenetic Analysis

EcoTILLING was performed according to the protocol given by Comai et al. (2004) with modifications (Rakshit et al., 2007). Forty-six isolates infecting *Oryza sativa* were used for EcoTILLING to find polymorphic genes. To increase throughput of the screen, we made six template DNA pools each containing equal concentrations of eight (five pools) or six (one pool) subject DNAs as well as 70-15 reference DNA. The TILLING platform can readily detect one chromosome harboring a mutation among up to 16 chromosomes (Till et al., 2003). Once mutations were detected by EcoTILLING with pooled DNA, DNA of each isolate was treated separately to identify those mutations. Presence/absence of an EcoTILLING fragment was assumed to be caused by a nucleotide change, and nucleotide difference estimated between isolates. Nucleotide diversity θ (Watterson, 1975) was calculated by dividing the estimated number of polymorphic sites with the total length of the examined secreted protein gene regions. Phylogenetic analysis was performed using the MEGA software package (Kumar et al., 2004). Hierarchical cluster analysis was done by Cluster 3.0 for Mac OS X (De Hoon et al., 2004).

454 Sequencing of Ina168 Whole Genome

Sequencing the whole genome of Ina168 isolate and sequence assembly were done with the 454 FLX sequencer by Agencourt Bioscience. We obtained 1,950,918 sequence reads that correspond to ~10× genome coverage (491.6 Mb total sequence; Table 3). De novo assembly trial generated 3345 large (>500 bp) contigs corresponding to a total size of 37.6 Mb. Among the large contigs, average, N50, and the largest contig sizes were 11.2, 28.4, and 206.7 kb, respectively. Using the reference published 70-15 sequence (Supercontig_5), the number of the large contigs reduced to 2300. A total of 132,400 raw sequence reads did not

match to the 70-15 sequence according to the 454 FLX reference mapper criteria (minimum overlap length 40 bp; minimum overlap identity 90%). These reads were subjected to de novo assembly, resulting in 1760 contig fragments corresponding to 1,677,825 bp that cannot be mapped to 70-15 reference sequence and thus were named Ina168 unmapped sequence.

Ina168 unmapped sequences were subjected to BLASTN search against a database of raw read sequences of 70-15 deposited at TRACE database of NCBI, ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/magnaporthe_grisea, resulting in recovery of 734 contigs (568,812bp) finding matches (E-value $\leq 1e^{-5}$) in 70-15 unmapped sequences. A total of 1026 contigs corresponding to 1,109,013 bp Ina168 DNA sequences did not match either 70-15 Supercontig_5 or raw read sequences and thus were defined as Ina168-specific fragments. The pex ORFs (see Supplemental Data Set 2 online) were selected from Ina168 unmapped sequence; therefore, some of the pex ORFs were missing from 70-15 reference sequence (Supercontig_5) but were actually present in 70-15 as revealed by PCR (see Results and Figure 1).

Sequence Divergence between Ina168 and 70-15

Divergence between Ina168 and 70-15 was estimated based on the nucleotide sequence alignment between supercontigs of 70-15 genome and 454 raw reads of Ina168. Divergence was calculated by the Jukes and Cantor (1969) method. The size of DNA region present in both 70-15 and Ina168 excluding indels (36,273,449 bp) was used as the denominator of the calculation of distance.

Genetic Complementation

We followed Sweigard et al. (1997) for the genetic transformation of *M. oryzae*. The lists of plasmids and PCR primers we used are given in Supplemental Tables 3 and 5 online, respectively. For the complementation assay of *M. oryzae* isolates TH68-141 and Ina86-137 that do not have *AVR-Pia* locus with pex22, a 2.2-kb fragment containing the *AVR-Pia* gene was amplified with the primers NotI-pex22-U1 (see Supplemental Table 5 online) and XbaI-pex22-L1 with KOD-Plus- (Toyobo) according to the manufacturer's instructions using total DNA of *M. oryzae* isolate Ina168 as template. The PCR product was digested with *Not*I and *Xba*I, and ligated to pCB1004, which carries the hygromycin phosphtransferase gene (Sweigard et al., 1997), generating pCB1004-pex22.

For the complementation assay of *M. oryzae* isolates Sasa2, Ina86-137, and TH68-140 that do not have *AVR-Pii* locus with pex33/279, a 1.4-kb fragment containing *AVR-Pii* gene was amplified with the primers NotI-pex33-U1 and XbaI-pex33-L1 with KOD-Plus- according to the manufacturer's instructions using total DNA of *M. oryzae* isolate Ina168 as template. The PCR product was digested with *Not*I and *Xba*I, and ligated to the same sites of pCB1531, which carries the bialaphos-resistant gene (Sweigard et al., 1997), generating pCB1531-pex33.

For construction of the *pex22p:pex33 and pex22p:pex279* expression vectors pCB1531-pex22p-pex33 and pCB1531-pex22p-pex279, first we made pCB1531-pex22p-EGFP. A 1.4-kb fragment containing the *pex22* gene promoter was amplified with the primers NotI-pex22-U1 and XbaI-pex22p-L3 using pCB1004-pex22 as template. The PCR product was digested with *Not*I and *Xba*I, and ligated to *Not*I and *Xba*I sites of pBAGFP (Kimura et al., 2001), generating pCB1531-pex22p-EGFP. A 0.2-kb fragment containing *pex33* gene was amplified with the primers pBAFP_kozak_pex33_XbaI_F and pBAFP_pex33_BamHI_R using pCB1531-pex33 as template. The PCR product was digested with *Xba*I and *Bam*HI, and exchanged *EGFP* gene at same sites of pCB1531-pex22p-EGFP, generating pCB1531-pex22p-pex33. A 0.2-kb fragment containing the *pex279* gene was amplified with the primers XbaI-kozak-pex279-U1 and BamHI-pex279-L1 using pCB1531-pex33 as template. The PCR product was digested with *Xba*I and *Bam*HI, and exchanged *EGFP* gene at same sites of pCB1531-pex22p-EGFP, generating pCB1531-pex22p-pex279.

For complementation assay of *M. oryzae* isolates Sasa2 and Ina72 that do not have *AVR-Pik*, *AVR-Pikm*, and *AVR-Pikp* loci with pex31, a 2.2-kb fragment containing *AVR-Pik* gene was amplified with the primers NotI-pex31-U1 and XbaI-pex31-L1 with KOD-Plus- according to the manufacturer's instructions using total DNA of *M. oryzae* isolate Ina86-137 carrying pex31-D allele as template. The PCR product was digested with *Not*I and *Xba*I, and ligated to the same sites of pCB1004, which carries the *hygromycin-resistant gene* (Sweigard et al., 1997), generating pCB1004-pex31-D.

For construction of the *pex22p:pex31-D* expression vector pCB1531-pex22p-pex31-D, a 0.3-kb fragment containing the pex31 gene was amplified with the primers Xba1_kozak_pex31_U1 and pBAFP_pex33_BamHI_R using genomic DNA of Ina86-137 isolate as template. The PCR product was digested with *Xba*I and *Bam*HI, and exchanged *EGFP* gene at same ;sites of pCB1531-pex22p-EGFP, generating pCB1531-pex22p-pex31-D.

SuperSAGE of *M. oryzae* Strain Ina168-Infected Rice Leaves

Leaf sheaths of rice cultivar Shin-2 at the four to five leaf stage were filled with a suspension of spores (3×10^5 spores/mL) of *M. oryzae* isolate Ina168 using a syringe to establish a compatible interaction. The leaf sheaths were incubated at 25°C under dark conditions for 40 h. SuperSAGE library was made from total RNA of the leaf sheaths as described (Matsumura et al., 2003; Terauchi et al., 2008). Di-tag fragments were sequenced by the 454 FLX sequencer (454 Life Sciences). Each 26-bp tag sequence was used for BLASTN search against *M. oryzae* 70-15 genome sequence, Ina168 genome sequence, and *O. sativa* spp *japonica* cultivar Nipponbare genome sequence (see Supplemental Data Set 3 online). A total of 131,228 tags each comprising 26-bp sequence were recovered, of which 82,481 tags showed perfect match to the rice genome (cv Nipponbare; International Rice Genome Sequencing Project, 2005), 18,189 tags to *M. oryzae* 70-15 reference sequences (Dean et al., 2005), and 2598 tags to *M. oryzae* Ina168 unmapped sequences. Number of tags for each of secreted protein genes of *M. oryzae* is given in Supplemental Data Sets 1 and 2 online. To identify the 3′ untranslated region of In168pex22, pex33, and pex31 genes, 3′-RACE RT-PCR was performed using the GeneRacer kit (Invitrogen). The same mRNA isolated for SuperSAGE experiment was used for 3′-RACE RT-PCR of pex22, pex33, and pex31.

Transient Assay of Cell Death in Rice Protoplasts

To make effector plasmids containing putative *AVR* genes under the control of maize (*Zea mays*) ubiqutin promoter, the PCR-amplified fragments of cDNA encoding pex22, pex33, and pex31 proteins with and without signal peptide were cloned into pAHC17 vector, which contains maize ubiquitin promoter and Nos terminator (Christensen and Quail, 1996). For the reporter plasmid to monitor cell viability, firefly luciferease (*luc*) gene under the control of maize ubiquitin promoter was used (*luc* plasmid). Protoplasts from etiolated seedling tissues were isolated as described by Chen et al. (2006). Electroporation was performed for delivery of effector and *luc* plasmids into rice protoplasts. Forty hours after the electroporation, protoplasts were collected by centrifugation. Luciferase assay was conducted using Luciferase Assay System (Promega).

Inducible Expression of pex22, pex33, and pex31-D in Rice

PCR-amplified fragments of cDNA corresponding to pex22, pex33, and pex31-D with and without signal peptide were cloned into the GVG vector pTA7001 (Aoyama and Chua, 1997). These constructs were used for transformation of *Agrobacterium tumefaciens* EHA105 by

electroporation. Transformation and regeneration of rice plants was performed according to Hiei et al. (1994) with a slight modification. After acclimatization, the transformants were grown in soil. Leaves of T0 plants carrying the GVG vectors were cut off and put in DEX (20 μM in 0.2% ethanol) to induce gene expression. To detect cell death on rice leaves, trypan blue staining was conducted as described by Koch and Slusarenko (1990). Detached leaves were boiled for 5 min in lactophenol trypan blue (2.5 mg/mL), cleared in chloral hydrate solution (2.5 g/ml) overnight, and kept in 50% glycerol before observation.

RT-PCR for Detecting Expression of Transgenes in Rice Plants

Total RNA was extracted from leaf tissues of T0 plants carrying GVG vectors 24 h after DEX treatment using Plant RNA Isolation Reagent (Invitrogen), which was subsequently treated with TURBO DNase (Ambion). From 2 μg of the DNase-treated RNA of each sample, single-strand cDNA was synthesized using oligo(dT) primer and ReverTra Ace (Toyobo). To confirm the gene expression of pex22, pex33, pex31, and rice actin gene *RAc7* (X15863), these genes were amplified by PCR with primers as given in Supplemental Table 5 online.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL databases under the following accession numbers: AB498873 (*AVR-Pia*), AB498874 (*AVR-Pii*), AB498875-AB498879 (*AVR-Pik/km/kp*), X15863 (*RAc7*), and MGG_03982 (*Actin*).

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Ambiguous Symptoms Caused by 70-15 Isolate Inoculation of Various Rice Cultivars.

Supplemental Figure 2. Frequency Distribution of Nucleotide Diversity among 1032 Loci of *M. oryzae* Putative Secreted Protein Genes.

Supplemental Figure 3. Size Distribution of Predicted Proteins of *M. oryzae*.

Supplemental Figure 4. pex22, pex33, and pex31 Transcripts Are Expressed during *M. oryzae* Infection of Rice Leaf Sheath as Revealed by SuperSAGE and 3′-RACE RT-PCR.

Supplemental Figure 5. Confirmation of Active Transcription of pex Transgenes by RT-PCR in *M. oryzae* Transformants during Infection.

Supplemental Figure 6. Transformation of *M. oryzae* Isolates with pex22, pex33, and pex31-D Complements *AVR-Pia*, *AVR-Pii*, and *AVR-Pik/km/kp*, respectively.

Supplemental Figure 7. Additional Results of Interactions between *M. oryzae* and Rice.

Supplemental Figure 8. Pulsed Field Gel Electrophoresis Images of Chromosomes of Ina168 and GFS1-7-2 Isolates of *M. oryzae*.

Supplemental Figure 9. EcoTILLING Result of *AVR-Pita*.

Supplemental Table 1. Forty-Six Isolates of *M. oryzae* Used for EcoTILLING and Phylogenetic Analysis.

Supplemental Table 2. A List of *M. oryzae* Putative Secreted Proteins Possessing the [RK]CxxCxxxxxxxxxxxH] Motif.

Supplemental Table 3. A List of Plasmids Used for the Genetic Transformation of *M. oryzae*.

Supplemental Table 4. A List of Transposons Analyzed for the Linkage with Secreted Protein Genes.

Supplemental Table 5. Primers Used for Plasmid Construction and RT-PCR.

Supplemental Data Set 1. A List of 1306 *M. oryzae* Genes Coding for Putative Secreted Proteins Selected on the Basis of Predicted Proteins of 70-15 Isolate.

Supplemental Data Set 2. A List of 316 Ina168-Unmapped Genes Coding for Putative Secreted Proteins.

Supplemental Data Set 3. SuperSAGE Result of *M. oryzae* Isolate Ina168-Infected Rice (cv Shin2) Leaves (Information Given Only for Tags with More Than Four Counts).

ACKNOWLEDGMENTS

This work was supported by the Program for Promotion of Basic Research Activities for Innovative Biosciences (Japan), the Iwate University 21st Centrury COE Program: Establishment of Thermo-Biosystem Research Program, the Ministry of Agriculture, Forestry, and Fisheries of Japan (Genomics for Agricultural Innovation PMI-0010), and Japan Society for the Promotion of Science Grants 18310136 and 19.13042. We thank Matt Shenton for improving the manuscript and Hideki Innan and Koichiro Tsunewaki for invaluable suggestions.

Received February 16, 2009; revised April 18, 2009; accepted April 30, 2009; published May 19, 2009.

REFERENCES

Allen, R.L., Bittner-Eddy, P.D., Grenville-Briggs, L.J., Meitz, J.C., Rehmany, A.P., Rose, L.E., and Beynon, J.L. (2004). Host-parasite coevolutionary conflict between Arabidopsis and downy mildew. Science **306:** 1957–1960.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. **17:** 3389–3402.

Andries, K., et al. (2005). A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. Science **307:** 223–227.

Aoyama, T., and Chua, N.-H. (1997). A glucocorticoid-mediated transcriptional induction system in transgenic plants. Plant J. **11:** 605–612.

Armstrong, M.R., et al. (2005). An ancestral oomycete locus contains late blight avirulence gene *Avr3a*, encoding a protein that is recognized in the host cytoplasm. Proc. Natl. Acad. Sci. USA **102:** 7766–7771.

Ashikawa, I., Hayashi, N., Yamane, H., Kanamori, H., Wu, J., Matsumoto, T., Ono, K., and Yano, M. (2008). Two adjacent nucleotide-binding site-leucine-rich repeat class genes are required to confer pikm-specific rice blast resistance. Genetics **180:** 2267–2276.

Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. **34:** W369–W373.

Berriman, M., et al. (2005). The genome of the African Trypanosome *Trypanosoma brucei.* Science **309:** 416–422.

Bohnert, H.U., Fudal, I., Dioh, W., Tharreau, D., Notteghem, J.L., and Lebrun, M.H. (2004). A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. Plant Cell **16:** 499–513.

Bryan, G.T., Wu, K.S., Farrall, L., Jia, Y., Hershey, H.P., McAdams, S.A., Faulk, K.N., Donaldson, G.K., Tarchini, R., and Valent, B.

(2000). A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell **12:** 2033–2046.

Catanzariti, A.M., Dodds, P.N., Lawrence, G.J., Ayliffe, M.A., and Ellis, J.G. (2006). Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. Plant Cell **18:** 243–256.

Chao, C.-C.T., and Ellingboe, A.H. (1991). Selection for mating competence in *Magnaporthe grisea* pathogenic to rice. Can. J. Bot. **69:** 2130–2134.

Chen, S., Tao, L., Zeng, L., Vega-Sanchez, M.E., Umemura, K., and Wang, G.-L. (2006). A highly efficient transient protoplast system for analyzing defence gene expression and protein-protein interactions in rice. Mol. Plant Pathol. **7:** 417–427.

Chuma, I., Tosa, Y., Taga, M., Nakayashiki, H., and Mayama, S. (2003). Meiotic behavior of a supernumerary chromosome in *Magnaporthe oryzae.* Curr. Genet. **43:** 191–198.

Christensen, A.H., and Quail, P.H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res. **5:** 213–218.

Clark, R.M., et al. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana.* Science **317:** 338–342.

Comai, L., Young, K., Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R., and Henikoff, S. (2004). Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant J. **37:** 778–786.

Couch, B.C., Fudal, I., Lebrun, M.-H., Tharreau, D., Valent, B., van Kim, P., Nottéghem, J.L., and Kohn, L.M. (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.

Couch, B.C., and Kohn, L.M. (2002). A multilocus gene genealogy concordant with host preference indicates segregation of a new species, *Magnaporthe oryzae*, from *M. grisea.* Mycologia **94:** 683–693.

Dean, R.A., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea.* Nature **434:** 980–986.

De Hoon, M.J.L., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics **20:** 1453–1454.

Dodds, P.N., Lawrence, G.L., Catanzariti, A.-M., Ayliffe, M.A., and Ellis, J.G. (2004). The *Melamspsora lini* AvrL567 avirulence genes are expressed in haustroria and their products are recognized inside plant cells. Plant Cell **16:** 755–768.

Ellis, J.G., Dodds, P.N., and Lawrence, G.J. (2007). The role of secreted proteins in diseases of plants caused by rust, powdery mildew and smut fungi. Curr. Opin. Microbiol. **10:** 326–331.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. **300:** 1005–1016.

Evans, R.M., and Hollenberg, S.M. (1988). Zinc fingers: Gilt by association. Cell **52:** 1–3.

Farman, M.L., Tosa, Y., Nitta, N., Leong, S.A., and Leong, S.A. (1996). Maggy, a retrotransposon in the genome of the rice blast fungus *Magnaporthe grisea.* Mol. Gen. Genet. **251:** 665–674.

Gardner, M.J., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum.* Nature **419:** 498–511.

Gout, L., Fudal, I., Kuhn, M.L., Blaise, F., Eckert, M., Cattolico, L., Balesdent, M.H., and Rouxel, T. (2006). Lost in the middle of nowhere: The AvrLm1 avirulence gene of the Dothideomycete *Leptosphaeria maculans.* Mol. Microbiol. **60:** 67–80.

Hamer, J.E., Farrall, L., Orbach, M.J., Valent, B., and Chumley, F.G. (1989). Host species-specific conservation of a family of repeated DNA sequences in the genome of a fungal plant pathogen. Proc. Natl. Acad. Sci. USA **86:** 9981–9985.

Hayashi, K., Yoshida, H., and Ashikawa, I. (2006). Development of PCR-based allele-specific and indel marker sets for nine rice blast resistance genes. Theor. Appl. Genet. **113:** 251–260.

Hiei, Y., Ohta, S., Komari, T., and Kumashiro, T. (1994). Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J. **6:** 271–282.

Hirschhorn, J.N., and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. **6:** 95–108.

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

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

Jukes, T.H., and Cantor, C.R. (1969). Evolution of protein molecules. In Mammalian Protein Metabolism, H.N. Munro, ed (New York: Academic Press), pp. 21–132.

Kamoun, S. (2007). Groovy times: Filamentous pathogen effectors revealed. Curr. Opin. Plant Biol. **10:** 358–365.

Kang, S., Sweigard, J.A., and Valent, B. (1995). The PWL host specificity gene family in the blast fungus *Magnaporthe grisea.* Mol. Plant Microbe Interact. **8:** 939–948.

Khang, C.H., Park, S.-Y., Lee, Y.-H., Valent, B., and Kang, S. (2008). Genome organization and evolution of the *AVR-Pita* avirulence gene family in the *Magnaporthe grisea* species complex. Mol. Plant Microbe Interact. **21:** 658–670.

Kimura, A., Takano, Y., Furusawa, I., and Okuno, T. (2001). Peroxisomal metabolic function is required for appressorium-mediated plant infection by *Colletotrichum lagenarium.* Plant Cell **13:** 1945–1957.

Kiyosawa, S., Mackill, D.J., Bonman, J.M., Tanaka, Y., and Ling, Z.Z. (1986). An attempt of classificaton of world's rice varieties based on reacton pattern to blast fungus strains. Bull. Nat. Inst. Agrobiol. Res. **2:** 13–39.

Koch, E., and Slusarenko, A. (1990). *Arabidopsis* is susceptible to infection by a downy mildew fungus. Plant Cell **2:** 437–445.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete gemomes. J. Mol. Biol. **305:** 567–580.

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

Lee, S.A., Wormsley, S., Kamoun, S., Lee, A.F.S., Joiner, K., and Wong, B. (2003). An analysis of the *Candida albicans* genome database for soluble secreted proteins using computer-based prediction algorithms. Yeast **20:** 595–610.

Lee, S.K., et al. (2009). Rice Pi5-mediated resistance to *Magnaporthe oryzae* requires the presence of two coiled-coil-nucleotide-binding-leucine-rich-repeat genes. Genetics **181:** 1627–1638.

Lin, F., Chen, S., Que, Z., Wang, L., Liu, X., and Pan, Q. (2007). The blast resistance gene *Pi37* encodes a nucleotide binding site-leucine-rice repeat protein and is a member of a resistance gene cluster on rice chromosome 1. Genetics **177:** 1871–1880.

Liu, Z., Bos, J.I.B., Armstrong, M., Whisson, S.C., da Cunha, L., Torto-Alalibo, T., Win, J., Avrova, A.O., Wright, F., Birch, P.R.J., and Kamoun, S. (2005). Patterns of diversifying selection in the phytotoxin-like *scr74* gene family of *Phytophthora infestans.* Mol. Biol. Evol. **22:** 659–672.

Margulies, M., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:** 376–380.

Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P.,

Kahl, G., Reuter, M., Kruger, D.H., and Terauchi, R. (2003). Gene expression analysis of plant host-pathogen interactions by SuperSAGE. Proc. Natl. Acad. Sci. USA **100:** 15718–15723.

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

Nielsen, H., Engelbrecht, J., Brunak, S., and Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:** 1–6.

Notteghem, J.L., and Silue, D. (1992). Distribution of the mating type alleles in *Magnaporthe grisea* populations pathogenic on rice. Phytopathology **82:** 421–424.

Orbach, M.J., Farrall, L., Sweigard, J.A., Chumley, F.G., and Valent, B. (2000). A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta.* Plant Cell **12:** 2019–2032.

Pritchard, J.K., Stephens, M., Rosenberg, N.A., and Donnelly, P. (2000). Association mapping in structured populations. Am. J. Hum. Genet. **67:** 170–181.

Qu, S., Liu, G., Zhou, B., Bellizzi, M., Zeng, L., Dai, L., Han, B., and Wang, G.-L. (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.

Rakshit, S., Rakshit, A., Matsumura, H., Takahashi, Y., Hasegawa, Y., Ito, A., Ishii, T., Miyashita, N.T., and Terauchi, R. (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.

Rehmeyer, C., Li, W., Kusaba, M., Kim, Y.S., Brown, D., Staben, C., Dean, R., and Farman, M. (2006). Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae.* Nucleic Acids Res. **34:** 4685–4701.

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

Silva, J.C., Loreto, E.L., and Clark, J.B. (2004). Factors that affect the horizontal transfer of transposable elements. Curr. Issues Mol. Biol. **6:** 57–72.

Stavrinides, J., Mccann, H.C., and Guttman, D.S. (2008). Host-pathogen interplay and the evolution of bacterial effectors. Cell. Microbiol. **10:** 285–292.

Sweigard, J.A., Carroll, A.M., Kang, S., Farrall, L., Chumley, F.G., and Valent, B. (1995). Identification, cloning, and characterization of *PWL2*, a gene for host species specificity in the rice blast fungus. Plant Cell **7:** 1221–1233.

Sweigard, J.A., Chumley, F., Carroll, A., Farrall, L., and Valent, B. (1997). A series of vectors for fungal transformation. Fungal Genet. Newsl. **44:** 52–55.

Talbot, N.J. (2003). On the trail of a cereal killer. Annu. Rev. Microbiol. **57:** 177–202.

Talbot, N.J., Salch, Y.P., Ma, M., and Hamer, J.E. (1993). Karyotypic variation within clonal lineages of the rice blast fungus, *Magnaporthe grisea.* Appl. Environ. Microbiol. **59:** 585–593.

Terauchi, R., Matsumura, H., Kruger, D.H., and Kahl, G. (2008). SuperSAGE: The most advanced transcriptome technology for functional genomics. In The Handbook of Plant Functional Genomics, G. Kahl and K. Meksem, eds (Weinhaim, Germany: Wiley-VCH), pp. 37–54.

Thrall, P.H., and Burdon, J. (2003). Evolution of virulence in plant host-pathogen metapopulation. Science **299:** 1735–1737.

Till, B.J., et al. (2003). Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res. **13:** 524–530.

Torto, T., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N.A.R., van West, P., and Kamoun, S. (2003). EST mining and functional expression assays identify extracellular effector proteins from *Phytophthora.* Genome Res. **13:** 1675–1685.

van der Does, H.C., and Rep, M. (2007). Virulence genes and the evolution of host specificity in plant-pathogenic fungi. Mol. Plant Microbe Interact. **20:** 1175–1182.

Wang, G.L., Mackill, D.J., Bonman, J.M., Mccouch, S.R., and Nelson, R.J. (1994). RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics **136:** 1421–1434.

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

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

Yasuda, N., Noguchi, M.T., and Fujita, Y. (2006). Partial mapping of avirulence genes *AVR-Pii* and *AVR-Pia* in the rice blast fungus *Magnaporthe oryzae.* Can. J. Plant Pathol. **28:** 494–498.

Zeigler, R.S., Scott, R.P., Leung, A.A., Bordeos, A.A., Kumar, J., and Nelson, R.J. (1997). Evidence of parasexual exchange of DNA in the rice blast fungus challenges its exclusive clonality. Phytopathology **87:** 284–294.

Zeigler, R.S., Tohme, J., Nelson, R.J., Levy, M., and Correa, F.J. (1994). Lineage exclusion: A proposal for linking blast population analysis to resistance breeding. In Rice Blast Disease, R.S. Zeigler, S.A. Leong, and P.S. Teng, eds (CAB International; International Rice Research Institute (IRRI). Los Banos, Laguna, Phillipines), pp. 267–292.

Association Genetics Reveals Three Novel Avirulence Genes from the Rice Blast Fungal Pathogen *Magnaporthe oryzae*

Kentaro Yoshida, Hiromasa Saitoh, Shizuko Fujisawa, Hiroyuki Kanzaki, Hideo Matsumura, Kakoto Yoshida, Yukio Tosa, Izumi Chuma, Yoshitaka Takano, Joe Win, Sophien Kamoun and Ryohei Terauchi

PLANT CELL 2009;21;1573-1591; originally published online May 19, 2009;
DOI: 10.1105/tpc.109.066324