Identification of a New Rice Blast Resistance Gene, *Pid3*, by Genomewide Comparison of Paired Nucleotide-Binding Site–Leucine-Rich Repeat Genes and Their Pseudogene Alleles Between the Two Sequenced Rice Genomes

Junjun Shang,[1],[*],[†] Yong Tao,[1],[*],[†] Xuewei Chen,[1],[*],[†] Yan Zou,[*],[†] Cailin Lei,[‡] Jing Wang,[*],[†] Xiaobing Li,[*] Xianfeng Zhao,[*] Meijun Zhang,[§] Zhike Lu,[§] Jichen Xu,[*] Zhukuan Cheng,[*] Jianmin Wan[‡] and Lihuang Zhu[*],[2]

[*]State Key Laboratory of Plant Genomics and National Plant Gene Research Centre (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China, [‡]Institute of Crop Research, Chinese Academy of Agricultural Sciences, Beijing 100081, China, [†]Graduate School of Chinese Academy of Sciences, Beijing 100101, China and [§]Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 101300, China

ABSTRACT

Rice blast, caused by *Magnaporthe oryzae*, is one of the most devastating diseases. The two major subspecies of Asian cultivated rice (*Oryza sativa* L.), *indica* and *japonica*, have shown obvious differences in rice blast resistance, but the genomic basis that underlies the difference is not clear. We performed a genomewide comparison of the major class of resistant gene family, the nucleotide-binding site–leucine-rich repeat (NBS–LRR) gene family, between 93-11 (*indica*) and Nipponbare (*japonica*) with a focus on their pseudogene members. We found great differences in either constitution or distribution of pseudogenes between the two genomes. According to this comparison, we designed the PCR-based molecular markers specific to the Nipponbare NBS–LRR pseudogene alleles and used them as cosegregation markers for blast susceptibility in a segregation population from a cross between a rice blast-resistant *indica* variety and a susceptible *japonica* variety. Through this approach, we identified a new blast resistance gene, *Pid3*, in the *indica* variety, Digu. The allelic *Pid3* loci in most of the tested *japonica* varieties were identified as pseudogenes due to a nonsense mutation at the nucleotide position 2208 starting from the translation initiation site. However, this mutation was not found in any of the tested *indica* varieties, African cultivated rice varieties, or AA genome-containing wild rice species. These results suggest that the pseudogenization of *Pid3* in *japonica* occurred after the divergence of *indica* and *japonica*.

THE two elite rice varieties with sequenced genomes, Nipponbare from Japan and 93-11 from China, represent the two major Asian cultivated rice subspecies, *Oryza sativa* L. *japonica* and *indica*, respectively. Sequencing these two genomes (GOFF *et al.* 2002; YU *et al.* 2002, 2005; IRGSP 2005) has opened an avenue for exploring the genomic basis of agricultural trait differences between these two subspecies (GARRIS *et al.* 2005). Rice blast caused by *Magnaporthe oryzae* is one of the most devastating diseases in both *indica* and *japonica* varieties. Many rice breeders have observed the difference in blast resistance between the two subspecies. Genetic charac-

terization of blast resistance has suggested that *indica* varieties have more complicated blast resistance patterns than *japonica* varieties (MACKILL *et al.* 1985; YU *et al.* 1987; MACKILL and BONMAN 1992; PAN *et al.* 1996; IMBE *et al.* 1997). In contrast to other Asian countries, China has a long history of growing both *indica* and *japonica* varieties (NORMILE 1997; ZONG *et al.* 2007). When Chinese rice varieties were tested with different Japanese *M. oryzae* strains, most of the *japonica* varieties could be classified into groups according to their respective disease-response patterns to the *M. oryzae* strains. However, when Chinese *indica* varieties were tested with the same stains, it was difficult to analyze their disease-response patterns because some were resistant to all the strains (KIYOSAWA 1984; LING 1984). This suggests that the *indica* and *japonica* varieties have diverged significantly in resistance to *M. oryzae* strains.

Plants face frequent threats of many potential pathogens, including viruses, bacteria, and fungi, in their environments. During the long coevolution with pathogens, plants have evolved sophisticated mechanisms to

perceive and defend against a pathogen's attack (DANGL and JONES 2001). Innate immunity, the most important of the plant defense responses, can be classified into "basal" and "R-gene-mediated" defenses (ABRAMOVITCH *et al.* 2006). The R-gene-mediated defense was originally proposed by H. H. Flor (FLOR 1956). This type of resistance is highly effective and easily manipulated in breeding programs to improve crop resistance. The majority of R proteins contain a nucleotide-binding site and a carboxy-terminal leucine-rich repeat domain (MARTIN *et al.* 2003). Notably, almost all the functionally identified nucleotide-binding site–leucine-rich repeat (NBS–LRR) genes are R genes with only one exception (FAIGON-SOVERNA *et al.* 2006) and thus this class of genes is regarded as the core component of the plant immune system (NIMCHUK *et al.* 2003).

Thus far, 10 rice blast R genes [*Pib* (WANG *et al.* 1999), *Pita* (BRYAN *et al.* 2000), *Pi9* (QU *et al.* 2006), *Pid2* (CHEN *et al.* 2006), *Pi2* (ZHOU *et al.* 2006), *Piz-t* (ZHOU *et al.* 2006), *Pi36* (LIU *et al.* 2007), *Pi37* (LIN *et al.* 2007), *Pikm* (ASHIKAWA *et al.* 2008), and *Pi5* (LEE *et al.* 2009)] have been identified via map-based cloning methods. All the cloned blast resistance genes encode NBS–LRR proteins except for *Pid2*, which encodes a receptor-like kinase (CHEN *et al.* 2006). The large number of published blast R gene isolations reveals the importance of the NBS–LRR gene family in rice blast disease resistance.

Recently, a genomewide investigation of the NBS–LRR genes of the two sequenced rice genomes was reported (YANG *et al.* 2006). However, that study focused only on genetic variations among common R gene sequences in the two genomes, but did not associate the genetic variations with the differential disease resistance between *indica* and *japonica* rice. In view of the difference in blast resistance between *indica* and *japonica* varieties described above, we performed a genomewide comparison of NBS–LRR gene sequences between 93-11 and Nipponbare. Our investigation focused on the systemic changes in the NBS–LRR gene sequences between the two genomes. In the survey, we found many potential NBS–LRR pseudogenes in both 93-11 and Nipponbare genomes due to premature stop codons or frameshift mutations in the ORFs. Notably we found great differences in constitution and distribution of pseudogenes between the two genomes.

The rice genome has been estimated to carry ∼500 NBS–LRR genes (BAI *et al.* 2002; MONOSI *et al.* 2004; ZHOU *et al.* 2004). Although most of the annotated rice NBS–LRR genes have not been functionally identified, we can expect that if a complete NBS–LRR gene functions as a blast R gene in one genome, its allelic pseudogene in the other genome might lose the function of blast resistance. Thus, we believe that the NBS–LRR pseudogene alleles can be used as cosegregation markers for blast susceptibility in an appropriate rice segregation population. In this study, we demonstrated

the feasibility of this approach. We applied this novel strategy to the *indica* variety, *Digu*, and successfully identified a new blast R gene, *Pid3*.

MATERIALS AND METHODS

Plant materials: The *indica* and *japonica* varieties used in this study are listed in supporting information, Table S1. The accessions and geographic origins of the 36 lines of wild rice and 4 varieties of African cultivated rice are listed in Table S2.

Fungal inoculation: The *M. oryzae* strains used in this study were collected and isolated from *indica*- and *japonica*-cultivated regions in China or Japan. The *M. oryzae* stain Zhong-10-8-14 was isolated from a *japonica*-cultivated region in northern China. We expected to use this strain to screen useful *R* genes from the highly resistant *indica* variety, Digu. The isolated *R* genes might be used to improve the blast resistance of Chinese *japonica* varieties.

Rice seedlings at the four-leaf stage were inoculated by spraying a spore suspension (5×10^4 spores/ml) of the *M. oryzae* strain onto the leaves. Subsequently, the inoculated plants were placed into a chamber maintained at 25° and 100% humidity in the dark for 24 hr and then transferred to a greenhouse. The disease reaction was examined 1 week after inoculation with the susceptible variety, TP309, as a control.

Identification of NBS–LRR seed sequences of rice: By searching with the keyword "NBS," we retrieved 68 protein sequences from the Swiss-Prot database (http://www.ebi.ac.uk/swissprot/). A set of 32,127 rice full-length cDNA sequences was downloaded from the KOME database (http://cdna01.dna.affrc.go.jp/cDNA/). The cDNA sequences were translated in six-frame and the longest ORF was reported as the putative protein sequence. A set of 153 sequences with the NBS motif was selected from the complete set of predicted full-length (fl)-cDNA supported rice proteins, using a hidden Markov model (HMM) (EDDY 1998) for the NBS domain from the Pfam database (PF0931; http://pfam.wustl.edu) (BATEMAN *et al.* 2002). The 153 protein sequences were aligned only on the basis of the NBS domain, using hmmalign. This alignment was then used to develop a rice-specific HMM model to identify related sequences. The refined HMM was compared again with the complete set of fl-cDNA sequences, using estwisedb in the wise2 software package (ftp://ftp.ebi.ac.uk/pub/software/unix/wise2/) (BIRNEY *et al.* 2004). All sequences that matched the model with a score of ≥25 were extracted (http://www.ebi.ac.uk/Tools/Wise2/doc_wise2.html#SECTION00055000000000000000). In total, we obtained 166 nonredundancy cDNA sequences that encoded the NBS domain. We identified a subset of them as seed sequences according to three criteria: (1) The cDNA sequences were aligned against the closest matching protein among 68 known NBS proteins, using estwise without obvious ORF-disrupted mutations (such as premature stop codons or frameshifts); (2) the noncoding regions of cDNA were initially searched against 68 known NBS protein sequences, using tblastn with an expectation score of >0.1; and (3) all sequences that met the previous criteria were verified by cross-validation, using estwise without obvious disablements. We found a total of 102 seed proteins, including 99 fl-cDNA–supported proteins and the 3 proteins encoded by *Pib* (WANG *et al.* 1999), *Pita* (BRYAN *et al.* 2000), and *Pi9* (QU *et al.* 2006). Alignment of protein amino acid sequences with nucleotide sequences was performed with the wise2 package.

Identification of NBS–LRR genes and pseudogenes in rice: Nipponbare and 93-11 genome sequences were downloaded from the TIGR Rice Genome Annotation Database (version 4.0, http://rice.plantbiology.msu.edu/) and RISE Database

(http://rise.genomics.org.cn), respectively. To identify the regions that contain the NBS domain, the genomes were scanned with a rice-specific NBS HMM model, using Genewisedb with a score of ≥ 25. A set of 653 regions was detected from the database of Nipponbare, and a set of 553 regions was detected from that of 93-11.

Using the BGI FGF pipeline (http://fgf.genomics.org.cn/), we compared the genome regions containing the NBS domain with the closest matching seed protein, using Genewise. If there was a continuous span of sequence that was $\geq 70\%$ of the length of the closest matching seed protein, we designated that sequence as an NBS–LRR gene; otherwise it was designated as an NBS–LRR gene segment. If an NBS gene contained a premature stop codon or frameshift mutation when compared with the seed protein, it was designated as a pseudogene.

Identification of Nipponbare and 93-11 NBS–LRR allelic pairs: We generated the Nipponbare and 93-11 NBS–LRR allelic pairs by comparing the gene sequences from the two genomes. We conducted a multiple sequence alignment of the 1206 protein sequences that contained an NBS domain from the Nipponbare and 93-11 genomes using ClustalW (CHENNA *et al.* 2003), including 267 NBS–LRR protein segments. Then the least divergent NBS–LRR gene pairs between the Nipponbare and 93-11 genomes were extracted from the phylogenetic tree, which was generated by using PAUP 4.0 (SWOFFORD 2000) based on the neighbor-joining (NJ) method. If they met all the following criteria, the least divergent pairs were assigned as NBS–LRR allelic pairs: (1) neither gene in the pair is an NBS–LRR segment, (2) both genes are located at the corresponding locus of the homologous chromosomes, and (3) the nucleotide sequence identity of the pair is $\geq 82\%$.

Expression analysis of *Pid3*: The four-leaf stage plants of the resistant *indica* variety Digu and the susceptible *japonica* variety TP309 were inoculated with *M. oryzae* strain Zhong-10-8-14. Total RNAs were isolated from the leaf tissue at 0, 3, 6, 12, 24, 48, and 72 hr, respectively, after inoculation. An equal amount of total RNA was pretreated with RNase-free DNase I (Promega, Madison, WI) and subjected to reverse transcription, using the Reverse Transcription System (Promega). The semiquantitative RT–PCR with the *Pid3*-specific primers 5′-TACTACTCATG GAAGCTAGTTCTC-3′ and 5′-ACGTCACAAATCATTCGCTC-3′ was performed for 30 cycles. The RT–PCR with *Actin*-specific primers 5′-CCTCGTCTCGACCTTGCTGGG-3′ and 5′-GA GAACAAGCAGGAGGACGGC-3′ was performed for 25 cycles as a control.

Sequence analysis of *Pid3*: The cDNA sequence of *Pid3* was obtained through RACE–PCR and RT–PCR analysis. RACE–PCR was carried out using the SMART RACE cDNA Amplification kit (Clontech) according to the manufacturer's instructions. The *Pid3*-specific primers 5′-CCATCTTGCACATCCTCTT GAGTG-3′ and 5′-ACATCTTATTGGAGAAGCAC-3′ were used in 5′-RACE and 3′-RACE, separately. The primers 5′-GAAGC TAGCAAGCTATGGCGGAG-3′ and 5′-ACGTCACAAATCATT CGCTC-3′ were used in RT–PCR. The 5′- and 3′-RACE ends and RT–PCR products were cloned into the pGEM-T-easy vector (Promega) prior to sequencing.

The coding regions of *Pid3* alleles in Digu, 93-11, Zaiyeqing8, LTH, Nipponbare, TP309, and the wild rice A4 were amplified from the respective genomic DNAs by using the primers 5′-AGTAACACCCAAGGATAGGATAG-3′ and 5′-GAA CGACAAGTGCGACATGATTG-3′. The PCR products were cloned into the pGEM-T-easy vector (Promega) and confirmed by DNA sequencing.

Analysis of the pseudogenization of the *Pid3* alleles: To examine the *Pid3* nonsense mutation locus, we designed two derived cleaved amplified polymorphic sequence (dCAPS) markers. With dCAPS marker Pid3-dCAPS-1, a 178-bp fragment was amplified using the primers 5′-TACTACTCATGGAAGC TAGTTCTC-3′ and 5′-GCAGCACTTCTTGACTACTGTCTGT-3′ and then digested with *Xba*I. With dCAPS marker Pid3-dCAPS-2, a 178-bp fragment was amplified using the primers 5′-TACTACTCATGGAAGCTAGTTCTC-3′ and 5′-AGCACTTCTT GACTACTGTCTGCCT-3′ and then digested with *Bam*HI.

Vector construction and rice transformation: The 6235-bp genomic sequence of *Pid3* containing the promoter region and the full coding region was amplified from genomic DNA of Digu, using the forward primer 5′-ACC<u>GAATTC</u>CACA CATTGTACACCTACGACCAC-3′ (*Eco*RI recognition site is underlined) and the reverse primer 5′-ACC<u>GTCGAC</u>GAAC GACAAGTGCGACATGATTG-3′ (*Sal*I recognition site is underlined), and then cloned into the binary vector pCAMBIA1300 through the *Eco*RI and *Sal*I cloning sites. After sequence verification the final construct was introduced into *Agrobacterium tumefaciens* LBA4404. The callus of susceptible *japonica* variety TP309 was transformed according to published methods (HIEI *et al.* 1994). The resistance of the primary transgenic lines (T_0) and of the self progenies (T_1) of the T_0 lines was challenged by inoculation with the *M. oryzae* strain Zhong-10-8-14. Segregation of the transgene in T_1 plants was confirmed by dCAPS marker Pid3-dCAPS-1; double bands containing 178- and 153-bp fragments can be checked on the agarose gel if the T_1 plants contained the transgenic fragment.

RESULTS

Genomewide comparison of the NBS–LRR genes between 93-11 and Nipponbare: To understand the genomic difference in blast resistance between *indica* and *japonica* varieties, we compared the NBS–LRR gene sequences between 93-11 and Nipponbare with a focus on the intact genes in one genome and their allelic pseudogenes in the other. We surveyed both 93-11 and Nipponbare genomes and reannotated all the NBS–LRR genes and their pseudogene counterparts, using a modified homology-based method for human pseudogene annotation (ZHANG and GERSTEIN 2003a; ZHANG *et al.* 2006). First, using an optimized HMM for the rice NBS domain, we identified 653 and 553 gene loci with the NBS domain from the 93-11 and Nipponbare genomes, respectively. Second, we defined a set of 102 seed protein sequences as intact NBS–LRR proteins (see details in MATERIALS and METHODS). The seed sequences were derived from 99 fl-cDNA (KIKUCHI *et al.* 2003) supported genes and 3 cloned NBS–LRR genes (WANG *et al.* 1999; BRYAN *et al.* 2000; QU *et al.* 2006). Finally, we selected the other NBS–LRR genes from the NBS domain loci in either the 93-11 or the Nipponbare genome if they had a continuous homology span on the closest matching seed protein sequence in at least 70% of its length. The homology search in the 93-11 genome resulted in 471 NBS–LRR genes among the 653 NBS domain loci. However, 184 of them were further identified as pseudogenes due to frameshift mutation or nonsense mutation or both. Meanwhile, in the Nipponbare genome, we identified 468 NBS–LRR genes from 553 gene loci, of which 150 were considered to be pseudogenes. The detailed information on this analysis
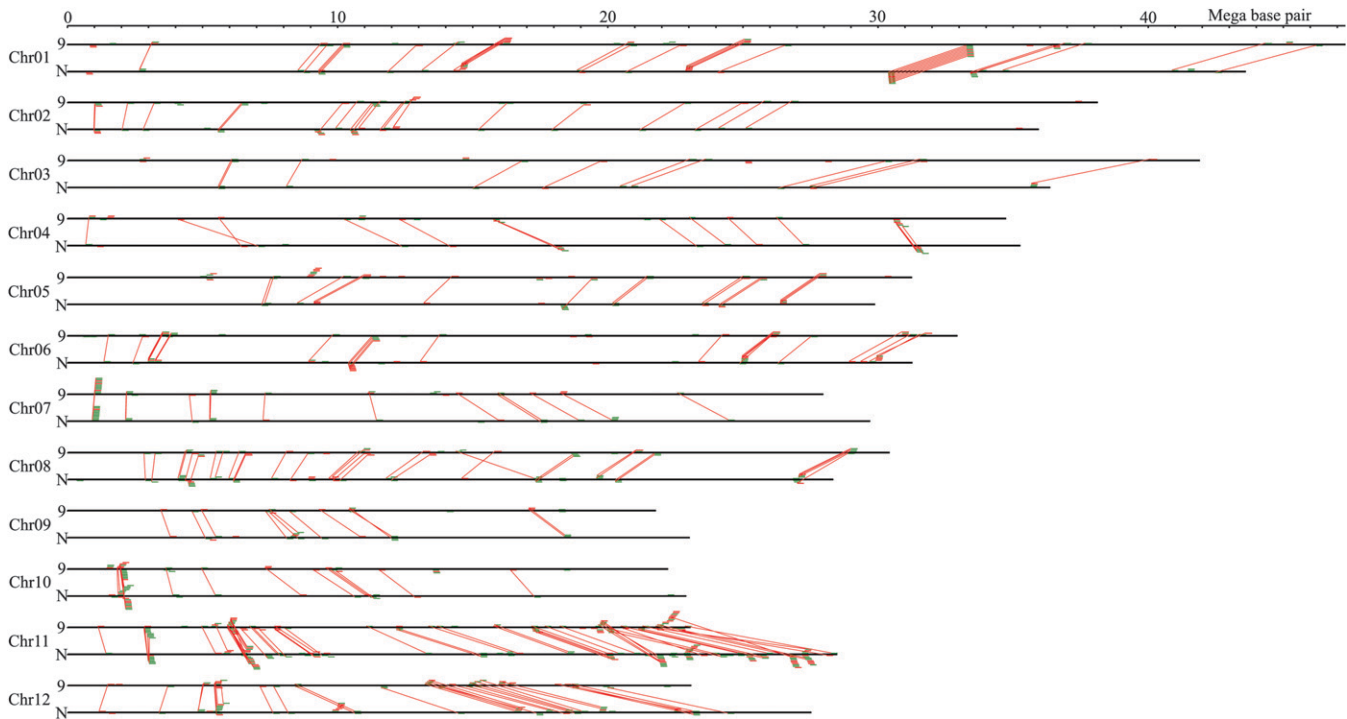
FIGURE 1.—Genomewide comparison of NBS–LRR genes between 93-11 and Nipponbare. The horizontal bars represent the physical maps of the chromosomes of 93-11 (marked with "9") and Nipponbare (marked with "N"). The green boxes represent the intact NBS–LRR genes, and the red boxes represent the NBS–LRR pseudogenes. Genes above and below the chromosomes correspond to forward and reverse orientations, respectively. The linking lines between chromosomes of 93-11 and Nipponbare indicate allelic loci.

is available in Table S3. All of the annotated NBS–LRR genes and pseudogenes have been submitted to GenBank as third party annotation (accession nos.: annotated Nipponbare NBS–LRR intact genes, FJ771051–FJ771368; annotated Nipponbare NBS–LRR pseudogenes, FJ771369–FJ771518; annotated 93-11 NBS–LRR intact genes, FJ771519–FJ771802; and annotated 93-11 NBS–LRR pseudogenes, FJ771803–FJ771989).

Next, we identified the NBS–LRR gene pairs by the most similarity criterion between the two genomes. Each NBS–LRR gene in one genome was paired only once to the closest homolog in the other. However, because of chromosome DNA rearrangements in some regions where some of the NBS–LRR clustered gene family locates, not every NBS–LRR gene of either Nipponbare or 93-11 has its closest match at the same loci in the other genome. Of the total 337 pairs from the two rice genomes, 301 were considered allelic, while the other 36 were nonallelic due to their different chromosome locations (Figure 1). Further analysis showed that there are 147 allelic pairs that contain at least one pseudogene. These 147 allelic pairs are grouped as follows: (1) both alleles are pseudogenes (53 pairs), (2) only the allele in 93-11 is a pseudogene (59 pairs), and (3) only the allele in Nipponbare is a pseudogene (35 pairs). The positions of the 94 allelic NBS–LRR gene pairs with a pseudogene allele either in 93-11 or in Nipponbare on the chromosomes are shown in Figure S1.

We compared our annotated 150 NBS–LRR pseudogenes in Nipponbare with the corresponding genes annotated by TIGR (YUAN *et al.* 2005). In the corresponding TIGR models, the fragments that contain frameshift or nonsense mutation sites in the 90 pseudogenes we annotated and those in the other 34 pseudogenes were taken as introns and untranslated regions (UTRs), respectively. The pseudogenization ratio of the KOME fl-cDNA NBS–LRR genes (27%, 41/152) is similar to that of the genomewide annotated NBS–LRR genes (32%, 150/468) in Nipponbare. These results also suggest that most NBS–LRR pseudogenes are expressed as intact genes.

Identifying an unknown blast R gene locus using an NBS–LRR pseudogene as a molecular marker: To analyze the genetic relationship between the pseudogenization of NBS–LRR genes and rice blast resistance, an F_2 population was developed from a cross between the blast-resistant *indica* variety, Digu, and the susceptible *japonica* variety, TP309. A total of 473 individuals of the F_2 population were inoculated with the *M. oryzae* strain Zhong-10-8-14. This *M. oryzae* strain was isolated from a *japonica*-cultivated area in northern China, and it is compatible to the previously cloned blast resistance gene *Pid2* from the variety, Digu. In the F_2 population, most plants were resistant, but we were able to clearly distinguish 24 susceptible individuals. Thus, the segregation ratio of resistance to susceptibility is 15:1 (resis-
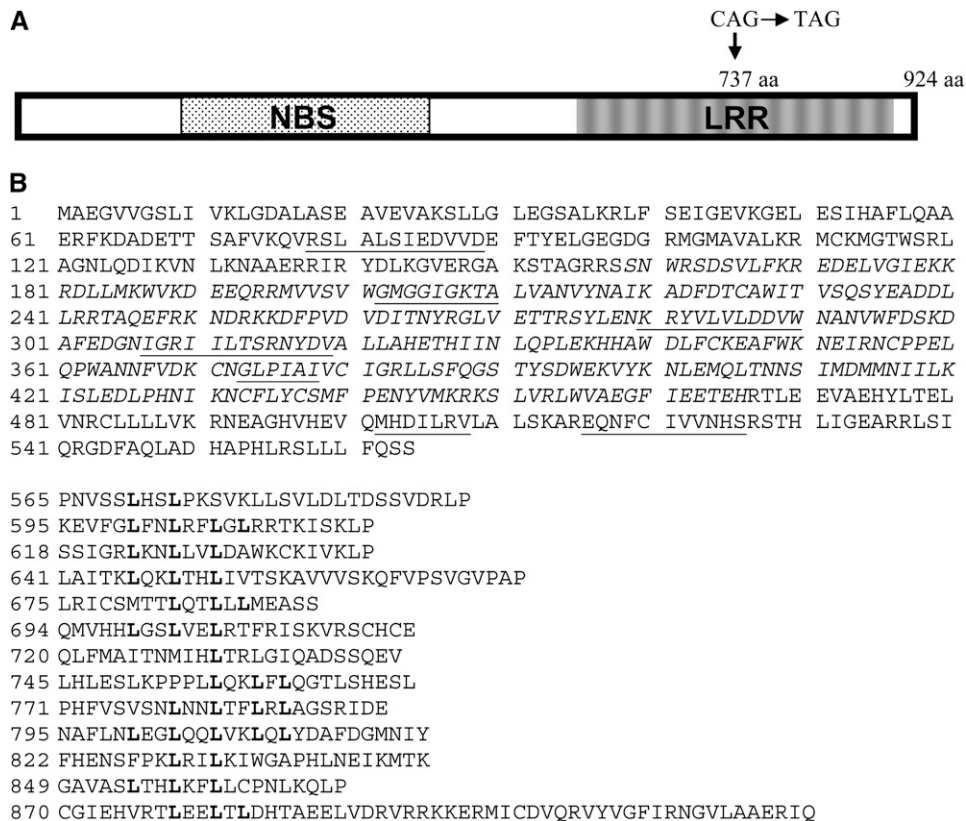
A

B

```
1    MAEGVVGSLI VKLGDALASE AVEVAKSLLG LEGSALKRLF SEIGEVKGEL ESIHAFLQAA
61   ERFKDADETT SAFVKQVRSL ALSIEDVVDE FTYELGEGDG RMGMAVALKR MCKMGTWSRL
121  AGNLQDIKVN LKNAAERRIR YDLKGVERGA KSTAGRRSSN WRSDSVLFKR EDELVGIEKK
181  RDLLMKWVKD EEQRRMVVSV WGMGGIGKTA LVANVYNAIK ADFDTCAWIT VSQSYEADDL
241  LRRTAQEFRK NDRKKDFPVD VDITNYRGLV ETTRSYLENK RYVLVLDDVW NANVWFDSKD
301  AFEDGNIGRI ILTSRNYDVA LLAHETHIIN LQPLEKHHAW DLFCKEAFWK NEIRNCPPEL
361  QPWANNFVDK CNGLPIAIVC IGRLLSFQGS TYSDWEKVYK NLEMQLTNNS IMDMMNIILK
421  ISLEDLPHNI KNCFLYCSMF PENYVMKRKS LVRLWVAEGF IEETEHRTLE EVAEHYLTEL
481  VNRCLLLLVK RNEAGHVHEV QMHDILRVLA LSKAREQNFC IVVNHSRSTH LIGEARRLSI
541  QRGDFAQLAD HAPHLRSLLL FQSS

565  PNVSSLHSLPKSVKLLSVLDLTDSSVDRLP
595  KEVFGLFNLRFLGLRRTKISKLP
618  SSIGRLKNLLVLDAWKCKIVKLP
641  LAITKLQKLTHLIVTSKAVVVSKQFVPSVGVPAP
675  LRICSMTTLQTLLLMEASS
694  QMVHHLGSLVELRTFRISKVRSCHCE
720  QLFMAITNMIHLTRLGIQADSSQEV
745  LHLESLKPPPLLQKLFLQGTLSHESL
771  PHFVSVSNLNNLTFLRLAGSRIDE
795  NAFLNLEGLQQLVKLQLYDAFDGMNIY
822  FHENSFPKLRILKIWGAPHLNEIKMTK
849  GAVASLTHLKFLLCPNLKQLP
870  CGIEHVRTLEELTLDHTAEELVDRVRRKKERMICDVQRVYVGFIRNGVLAAERIQ
```

FIGURE 2.—Sequence analysis of Pid3 protein. (A) Pid3 encodes a 924-amino-acid protein. The susceptible alleles in *japonica* varieties contain a premature stop codon due to a C to T nonsense mutation. This mutation leads to the production of a truncated polypeptide of 737 amino acids, resulting in a disrupted LRR region. (B) Pid3 protein sequence. The NBS domain from amino acids 158 to 466 is shown in italics. The conserved motifs are underlined. The C-terminal LRR region is composed of 13 imperfect LRR repeats that are shown separately.

tant:susceptible, 449:24; $\chi^2_{15:1} = 0.04$), indicating that Digu may contain two genes that confer resistance to the strain Zhong-10-8-14.

We designed the PCR-based molecular markers specific to the 35 Nipponbare pseudogenes in the above-described 35 allelic NBS–LRR gene pairs in the third group. (See the structures of the allelic gene pairs in Figure S2 and see the PCR primers and restriction enzymes used in these markers in Table S4.) We used these primers to survey the 24 susceptible individuals and the parent varieties Digu and TP309. We found a candidate pseudogene on chromosome 6 that could be detected in all 24 susceptible individuals and TP309 but not in Digu. This suggests that the intact NBS–LRR allele in Digu corresponding to this pseudogene is most likely one of the two functional blast resistance genes against Zhong-10-8-14. We named this blast R gene *Pid3*, as we found it has a different chromosomal location from the two previously identified *Pid(t)1* and *Pid2* from Digu (CHEN *et al.* 2004, 2006).

Cloning and functional analysis of *Pid3*: We amplified and sequenced the candidate NBS–LRR gene *Pid3* and its alleles from the genomes of the rice varieties Digu, 93-11, TP309, and Nipponbare. Both the ORFs of the Digu *Pid3* candidate and the allele from 93-11 consist of the same 2772-nucleotide (nt) sequence, encoding a protein with 924 amino acids. However, the alleles in *japonica* varieties TP309 and Nipponbare contain the same premature stop codon due to a C to T

nonsense mutation. This mutation leads to the production of a truncated 737-amino-acid polypeptide with a disrupted LRR region (Figure 2A). The allelic *pid3* candidate in Nipponbare corresponds to the gene Loc_Os06g22460 in TIGR's annotation (http://www.jcvi.org/). On the basis of the annotation, Loc_Os06g22460 contains a 135-nt intron that includes the nonsense mutation to avoid the ORF disruption.

Using the DNA or protein sequences of the candidate *Pid3* as queries, BLAST searches in GenBank (http://www.ncbi.nih.gov/blast) and KOME rice fl-cDNA databases (http://cdna01.dna.affrc.go.jp/cDNA/) were performed, respectively. No corresponding mRNA or fl-cDNA could be found. A 521-nt EST sequence (GenBank accession no. CK083629) was found to match the 3′ region of the candidate *Pid3*, indicating that the candidate *Pid3* is transcribed. We performed semiquantitative RT–PCR analysis of *Pid3* in Digu and TP309 infected with *M. oryzae* Zhong-10-8-14. The expression analysis indicated that *Pid3* was constitutively expressed in both the resistant *indica* variety Digu and the susceptible *japonica* variety TP309 (see Figure S3).

Through RT–PCR and RACE–PCR we obtained the mRNA sequence of the candidate *Pid3*, including a 2772-nt coding sequence, a 63-nt 5′-UTR, and a 135-nt 3′-UTR, with the TIGR-predicted intron around the nonsense mutation also being included in the mature mRNA. The candidate *Pid3* encodes a 924-amino-acid polypeptide that contains a conserved NBS domain in
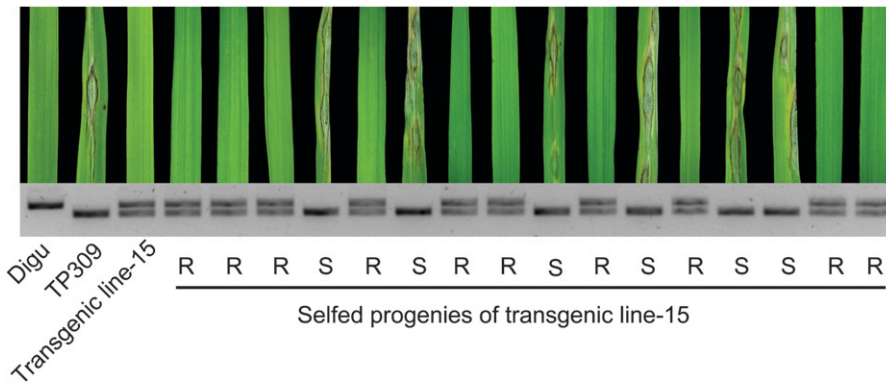
FIGURE 3.—Complementation test with *Pid3*. The primary transgenic line and the selfed progenies were inoculated by *M. oryzae* Zhong-10-8-14 and the genotypes were analyzed by a Pid3 dCAPS marker. The resistant variety Digu and the susceptible variety TP309 are shown as controls.

positions 158–466 from the translation initiation site. The NBS domain includes four sequence motifs, GMGGIGKTA (positions 202–210), KRYVLVLDDVW (positions 280–290), IGRIILTSRNYDV (positions 307–319), and GLPIAI (positions 373–378), corresponding to kinase 1a (p-loop), kinase 2, kinase 3a, and GLPL motifs, respectively. At the C terminus is the LRR region that comprises 13 imperfect LRR repeats. The MHD motif, MHDILRV (positions 502–508), and the NBS–LRR linker motif, EQNFCIVVNHS (positions 516–526), are present between the NBS domain and the LRR region. At the N terminus, there is a conserved motif, RSLALSIEDVVD (positions 78–89), but no TIR or coiled-coil motif is found (Figure 2B).

To confirm that *Pid3* confers blast resistance to the strain Zhong-8-10-14, we performed a genetic complementation test. Using Digu's genomic DNA as template, a 6235-bp DNA fragment, including the entire 2772-bp coding region, a 3010-bp upstream region, and a 451-bp downstream region, was amplified by high-fidelity PCR. After sequencing verification, this fragment was subcloned into the binary vector pCAMBIA1300, which was then introduced into susceptible variety TP309 by Agrobacterium-mediated transformation. We obtained 21 independent primary transgenic lines (T_0), all with resistance to the *M. oryzae* strain Zhong-10-8-14. Cosegregation of the transgene and blast resistance were confirmed in selfed progenies (T_1) of the T_0 lines (Figure 3). We then inoculated the transgenic plants with eight *indica M. oryzae* strains and eight *japonica* strains collected from China. As shown in Table 1, TP309 shows susceptibility to all 16 strains, whereas *Pid3* confers resistance to strains CH43, Sichuan36, JS2001-108, CH706, Zhong-10-8-14, ZK-10-2, 97-3-1, and 99-20-2, of which *Pid2* shows susceptibility to Sichuan36 and Zhong-8-10-14. This result suggests that *Pid3* confers a race-specific resistance to *M. oryzae*. Thus, we successfully identified a new blast *R* gene, *Pid3*, using the NBS–LRR pseudogene markers. (The sequence of *Pid3* has been deposited in GenBank, accession no. FJ745364.)

Comparison of the *Pid3* nonsense mutation among *indica*, *japonica*, and wild rice: To test if the ORF-disrupted mutation of *Pid3* occurs in other rice varieties,

we designed specific dCAPS markers that can distinguish the mutation alleles (see the primers and restriction enzymes used in the dCAPS markers in Table S4). In the *japonica* collection, 29 of 32 varieties contain the same C to T mutation, whereas none of the varieties in the *indica* collection contains this mutation. This result indicates that *Pid3* is replaced by its pseudogene in most tested *japonica* varieties (Figure S4).

The genus Oryza includes 2 cultivated species and 21 wild species (KHUSH 1997). *O. sativa*, Asian cultivated rice, is cultivated all over the world while *O. glaberrima*, African cultivated rice, is grown on a small scale in western Africa (KHUSH 1997). On the basis of gene transfer, the 2 cultivated species, *O. sativa* and *O. glaberrima*, and 6 wild species, *O. rufipogon, O. nivara, O. glumaepatula, O. meridionalis, O. breviligulata*, and *O. longistaminata*, have been grouped into a primary gene pool (KHUSH 1997). These 8 Oryza species all share the AA genome and the gene transfer among them can be accomplished through conventional hybridization and selection procedures (KHUSH 1997). Using the above-mentioned dCAPS markers, we examined whether the nonsense mutation in *Pid3* occurs in four African rice varieties and a collection of 36 wild rice lines belonging to the 6 AA genome wild species. The accessions and geographic origins of the wild rice and African cultivated rice are listed in Table S2. Notably, we found that none of the tested African cultivated varieties or the tested wild rice lines contains the nonsense mutation. These results suggest that the mutation creating the nonfunctional allele of *Pid3* occurred in *O. sativa* subspecies *japonica* after its split from subspecies *indica*.

Next we sequenced the respective allelic *Pid3* coding regions of the *O. rufipogon* wild rice line A4, the *indica* variety Zhaiyeqing 8 (ZYQ8) that had shown resistance to Zhong-10-8-14 (ZHU *et al.* 1993), and the *japonica* variety LTH that had been regarded as an universal blast-susceptible variety lacking all known major blast R genes (LING *et al.* 2000; TSUNEMATSU *et al.* 2000). Protein sequence comparison revealed a few amino acid changes among the *Pid3* proteins of A4, Digu, ZYQ8, 93-11, Nipponbare, TP309, and LTH (see the protein sequence alignment in Figure S5). The sequences

TABLE 1

Inoculation results of *Pid2* and *Pid3* transgenic lines with 16 *M. oryzae* strains

	ZB15	CH43	Sichuan36	JS2001-108	Sichuan26	CH706	CH45	CH704	Zhong10-8-14	ZK-10-2	91-65-1	97-3-1	97-27-2	99-20-2	99-26-1	99-26-2
TP309	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Pid3 transgenic line	S	R	R	R	S	R	S	S	R	R	S	R	S	R	S	R
Pid2 transgenic line	R	R	S	R	R	R	R	R	S	R	S	R	S	R	S	R

S, susceptible; R, resistant.

of the *Pid3* alleles were submitted to GenBank [accession nos. FJ745365 (Pid3-A4), FJ745366 (Pid3-ZYQ8), FJ745367 (Pid3-TP309), FJ745368 (Pid3-LTH), FJ773285 (Pid3-9311), and FJ773286 (Pid3-Nip)].

DISCUSSION

As a major R gene family in plants, NBS–LRR genes have been studied extensively. These investigations, however, have been focused mainly on the functional genes with a complete NBS–LRR coding sequence. Genomewide analyses of NBS–LRR genes in Arabidopsis (Columbia) and rice (Nipponbare) have revealed that a large number of NBS–LRR genes encoded in these two genomes are pseudogenes due to frameshift mutations or premature stop codons (MEYERS *et al.* 2003; MONOSI *et al.* 2004). At least 12 of 149 (∼8%) NBS–LRR genes in Arabidopsis (Columbia) (MEYERS *et al.* 2003) and >100 of the ∼500 (20%) NBS–LRR genes in rice (Nipponbare) (MONOSI *et al.* 2004) were predicted as pseudogenes. In these previous studies, manual reannotation was used to verify the results obtained from automated annotation programs. However, pseudogenes are often misincorporated into gene collections due to their high level of sequence homology to their real gene counterparts (NELSON 2004). It has been demonstrated that either a frameshift mutation or a nonsense mutation in an internal exon may force gene-prediction programs to adjust their splicing patterns and thus transform the pseudogenes into variable gene models (HARRISON *et al.* 2002; ZHANG and GERSTEIN 2004; ZHANG *et al.* 2004, 2006; ZHU *et al.* 2006). In our study, we used a homology-based method derived from a pipeline that has been successfully used in human pseudogene annotation (ZHANG and GERSTEIN 2003b; ZHANG *et al.* 2006) to reannotate the complete sets of NBS–LRR genes and pseudogenes in 93-11 and Nipponbare genomes. According to our annotation, >30% of the NBS–LRR genes are pseudogenes in either the 93-11 or the Nipponbare genomes, greater than the 20% that was reported previously (MONOSI *et al.* 2004).

The number of annotated NBS–LRR genes in 93-11 (471) is similar to that in Nipponbare (468). However, the number of NBS-domain-containing loci in 93-11 (653) is more than that in Nipponbare (553). The higher number of NBS-domain-containing loci in 93-11 could be artificially inflated from the "whole-genome shotgun" strategy applied to the 93-11 genome sequencing. Sequence assembly errors in this method might put fragments from a single locus into multiple loci. Although much effort has been made to obtain a high-quality genome assembly during sequencing of 93-11 and Nipponbare, sporadic mistakes in the genome databases are inevitable and would result in misannotated genes and overestimation of the number of pseudogenes during our analysis using these databases.

Thus far, >70 blast R-gene loci have been identified and mapped on rice chromosomes (BALLINI *et al.* 2008). Some of the genes have been successfully used in blast resistance breeding. When a map-based cloning strategy is used in the identification of R genes, resistance gene analog (RGA) markers that are designed according to the conserved NBS domain could be used to identify the corresponding candidate NBS–LRR resistance gene (MAGO *et al.* 1999; CHAUHAN *et al.* 2002; ZHUANG *et al.* 2002; LEE *et al.* 2003; SALLAUD *et al.* 2003; CHEN *et al.* 2004). In practice, however, successful examples of isolation of an NBS–LRR resistance gene with RGA markers have rarely been reported. The ineffectiveness of this method is possibly due to the complex nature of NBS–LRR genes in the genome. In contrast to the use of RGA markers, we designed molecular markers on the basis of the pseudogenization of NBS–LRR genes. We analyzed the relationship between the NBS–LRR pseudogenes and blast susceptibility in an appropriate population derived from resistant and susceptible varieties. In our method, when cosegregation of a tested pseudogene with the susceptible phenotype is observed, the intact allele can be regarded as the candidate blast R gene. This strategy has two advantages. First, only a small number of susceptible individual plants need to be analyzed with the landing markers. Second, scoring susceptible plants is more precise than scoring resistant plants. In this study, we demonstrated the effectiveness of this strategy by identifying a previously unknown R gene, *Pid3*, from the blast-resistant variety, Digu, in contrast to our labor-intensive cloning of the *Pid2* gene from the same variety (CHEN *et al.* 2006). In fact, we also tried to identify the blast R gene other than *Pid3* revealed in the genetic analysis of the F$_2$ population by using the same strategy, but we could not find any other pseudogene marker associated with the susceptible individuals. This implies that not all of the susceptible alleles of blast R genes are pseudogenes. For example, one amino acid change resulted in the loss of resistance of *Pita* (BRYAN *et al.* 2000) and *Pi36* (LIU *et al.* 2007).

LITERATURE CITED

ABRAMOVITCH, R. B., J. C. ANDERSON and G. B. MARTIN, 2006 Bacterial elicitation and evasion of plant innate immunity. Nat. Rev. Mol. Cell. Biol. **7:** 601–611.

ASHIKAWA, I., N. HAYASHI, H. YAMANE, H. KANAMORI, J. WU *et al.*, 2008 Two adjacent nucleotide-binding site-leucine-rich repeat class genes are required to confer pikm-specific rice blast resistance. Genetics **180:** 2267–2276.

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

BALLINI, E., J. B. MOREL, G. DROC, A. PRICE, B. COURTOIS *et al.*, 2008 A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. Mol. Plant Microbe Interact. **21:** 859–868.

BATEMAN, A., E. BIRNEY, L. CERRUTI, R. DURBIN, L. ETWILLER *et al.*, 2002 The Pfam protein families database. Nucleic Acids Res. **30:** 276–280.

BIRNEY, E., M. CLAMP and R. DURBIN, 2004 Genewise and genome-wise. Genome Res. **14:** 988–995.

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

CHAUHAN, R. S., M. L. FARMAN, H. B. ZHANG and S. A. LEONG, 2002 Genetic and physical mapping of a rice blast resistance locus, Pi-CO39(t), that corresponds to the avirulence gene AVR1–CO39 of Magnaporthe grisea. Mol. Genet. Genomics **267:** 603–612.

CHEN, X., J. SHANG, D. CHEN, C. LEI, Y. ZOU *et al.*, 2006 A B-lectin receptor kinase gene conferring rice blast resistance. Plant J. **46:** 794–804.

CHEN, X. W., S. G. LI, J. C. XU, W. X. ZHAI, Z. Z. LING *et al.*, 2004 Identification of two blast resistance genes in a rice variety, Digu. J. Phytopathol. **152:** 77–85.

CHENNA, R., H. SUGAWARA, T. KOIKE, R. LOPEZ, T. J. GIBSON *et al.*, 2003 Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. **31:** 3497–3500.

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

EDDY, S. R., 1998 Profile hidden Markov models. Bioinformatics **14:** 755–763.

FAIGON-SOVERNA, A., F. G. HARMON, L. STORANI, E. KARAYEKOV, R. J. STANELONI *et al.*, 2006 A constitutive shade-avoidance mutant implicates TIR-NBS-LRR proteins in Arabidopsis photomorphogenic development. Plant Cell **18:** 2919–2928.

FLOR, H., 1956 The complementary genic systems in flax and flax rust. Adv. Genet. **8:** 29–54.

GARRIS, A. J., T. H. TAI, J. COBURN, S. KRESOVICH and S. McCOUCH, 2005 Genetic structure and diversity in Oryza sativa L. Genetics **169:** 1631–1638.

GOFF, S. A., D. RICKE, T. H. LAN, G. PRESTING, R. WANG *et al.*, 2002 A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science **296:** 92–100.

HARRISON, P. M., H. HEGYI, S. BALASUBRAMANIAN, N. M. LUSCOMBE, P. BERTONE *et al.*, 2002 Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. Genome Res. **12:** 272–280.

HIEI, Y., S. OHTA, T. KOMARI and T. KUMASHIRO, 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.

IMBE, T., S. OBA, M. J. T. YANORIA and H. TSUNEMATSU, 1997 A new gene for blast resistance in rice cultivar, IR24. Rice Genet. Newsl. **14:** 60–62.

IRGSP, 2005 The map-based sequence of the rice genome. Nature **436:** 793–800.

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

KIKUCHI, S., K. SATOH, T. NAGATA, N. KAWAGASHIRA, K. DOI *et al.*, 2003 Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science **301:** 376–379.

KIYOSAWA, S., 1984 Establishment of differential varieties for pathogenicity test of rice blast fungus. Rice Genet. Newsl. **1:** 95–97.

LEE, S. Y., J. S. SEO, M. RODRIGUEZ-LANETTY and D. H. LEE, 2003 Comparative analysis of superfamilies of NBS-encoding disease resistance gene analogs in cultivated and wild apple species. Mol. Genet. Genomics **269:** 101–108.

LEE, S. K., M. Y. SONG, Y. S. SEO, H. K. KIM, S. KO *et al.*, 2009 Rice Pi5-mediated resistance to *Magnaporthe oryzae* requires the presence of two CC-NB-LRR Genes. Genetics **181:** 1627–1638.

Lin, F., S. Chen, Z. Que, L. Wang, X. Liu et al., 2007 The blast resistance gene Pi37 encodes an NBS-LRR protein and is a member of a resistance gene cluster on rice chromosome 1. Genetics 177: 1871–1880.

Ling, Z., 1984 Blast resistance classification of some rice varieties in China (in Chinese). Scientia Agricultura Sinica 2: 19–28.

Ling, Z., T. Mew, J. Wang, C. Lei and N. Huang, 2000 Development of Chinese near isogenic lines of rice and their differential ability to pathogenic races of Pyricularia grisea (in Chinese). Scientia Agricultura Sinica 33: 1–8.

Liu, X., F. Lin, L. Wang and Q. Pan, 2007 The in silico map-based cloning of Pi36, a rice coiled-coil nucleotide-binding site leucine-rich repeat gene that confers race-specific resistance to the blast fungus. Genetics 176: 2541–2549.

Mackill, D. J., and J. M. Bonman, 1992 Inheritance of blast resistance in near-isogenic lines of rice. Phytopathology 82: 746–749.

Mackill, D. J., J. M. Bonman, H. S. Suh and R. Srilingam, 1985 Genes for resistance to Philippine isolates of the rice blast pathogen. Rice Genet. Newsl. 2: 80–81.

Mago, R., S. Nair and M. Mohan, 1999 Resistance gene analogues from rice: cloning, sequencing and mapping. Theor. Appl. Genet. 99: 50–57.

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

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

Monosi, B., R. J. Wisser, L. Pennill and S. H. Hulbert, 2004 Full-genome analysis of resistance gene homologues in rice. Theor. Appl. Genet. 109: 1434–1447.

Nelson, D. R., 2004 'Frankenstein genes', or the Mad Magazine version of the human pseudogenome. Hum. Genomics 1: 310–316.

Nimchuk, Z., T. Eulgem, B. F. Holt, 3rd and J. L. Dangl, 2003 Recognition and response in the plant immune system. Annu. Rev. Genet. 37: 579–609.

Normile, D., 1997 Archaeology: Yangtze seen as earliest rice site. Science 275: 309–310.

Pan, Q., L. Wang, H. Ikehashi and T. Tanisaka, 1996 Identification of a new blast resistance gene in the indica rice cultivar kasalath using Japanese differential cultivars and isozyme markers. Phytopathology 86: 1071–1075.

Qu, S., G. Liu, B. Zhou, M. Bellizzi, L. Zeng et al., 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.

Sallaud, C., M. Lorieux, E. Roumen, D. Tharreau, R. Berruyer et al., 2003 Identification of five new blast resistance genes in the highly blast-resistant rice variety IR64 using a QTL mapping strategy. Theor. Appl. Genet. 106: 794–803.

Swofford, D., 2000 PAUP*: Phylogenetic Analysis Using Parsimony. Sinauer Associates, Sunderland, MA.

Tsunematsu, H., M. J. T. Yanoria, L. A. Ebron, N. Hayashi, I. Ando et al., 2000 Development of monogenic lines of rice for blast resistance. Breed. Sci. 50: 229–234.

Wang, Z. X., M. Yano, U. Yamanouchi, M. Iwamoto, L. Monna et al., 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.

Yang, S., Z. Feng, X. Zhang, K. Jiang, X. Jin et al., 2006 Genome-wide investigation on the genetic variations of rice disease resistance genes. Plant Mol. Biol. 62: 181–193.

Yu, J., S. Hu, J. Wang, G. K. Wong, S. Li et al., 2002 A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296: 79–92.

Yu, J., J. Wang, W. Lin, S. Li, H. Li et al., 2005 The genomes of Oryza sativa: a history of duplications. PLoS Biol. 3: e38.

Yu, Z. H., D. J. Mackill and J. M. Bonman, 1987 Inheritance of resistance to blast in some traditional and improved rice cultivars. Phytopathology 77: 323–326.

Yuan, Q., S. Ouyang, A. Wang, W. Zhu, R. Maiti et al., 2005 The institute for genomic research Osa1 rice genome annotation database. Plant Physiol. 138: 18–26.

Zhang, Z., and M. Gerstein, 2003a The human genome has 49 cytochrome c pseudogenes, including a relic of a primordial gene that still functions in mouse. Gene 312: 61–72.

Zhang, Z., and M. Gerstein, 2003b Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. Nucleic Acids Res. 31: 5338–5348.

Zhang, Z., and M. Gerstein, 2004 Large-scale analysis of pseudogenes in the human genome. Curr. Opin. Genet. Dev. 14: 328–335.

Zhang, Z., N. Carriero and M. Gerstein, 2004 Comparative analysis of processed pseudogenes in the mouse and human genomes. Trends Genet. 20: 62–67.

Zhang, Z., N. Carriero, D. Zheng, J. Karro, P. M. Harrison et al., 2006 PseudoPipe: an automated pseudogene identification pipeline. Bioinformatics 22: 1437–1439.

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

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

Zhu, L., Y. Chen, Y. Xu, J. Xu, H. Cai et al., 1993 Construction of a molecular map of rice and gene mapping using a double haploid population of a cross between Indica and Japonica varieties. Rice Genet. Newsl. 10: 132–135.

Zhu, X., M. Gerstein and M. Snyder, 2006 ProCAT: a data analysis approach for protein microarrays. Genome Biol. 7: R110.

Zhuang, J.-Y., W.-B. Ma, J.-L. Wu, R.-Y. Chai, J. Lu et al., 2002 Mapping of leaf and neck blast resistance genes with resistance gene analog, RAPD and RFLP in rice. Euphytica 128: 363–370.

Zong, Y., Z. Chen, J. B. Innes, C. Chen, Z. Wang et al., 2007 Fire and flood management of coastal swamp enabled first rice paddy cultivation in east China. Nature 449: 459–462.

GENETICS

Identification of a New Rice Blast Resistance Gene, *Pid3*, by Genomewide Comparison of Paired Nucleotide-Binding Site-Leucine-Rich Repeat Genes and Their Pseudogene Alleles Between the Two Sequenced Rice Genomes

Junjun Shang, Yong Tao , Xuewei Chen, Yan Zou, Cailin Lei, Jing Wang, Xiaobing Li, Xianfeng Zhao, Meijun Zhang, Zhike Lu, Jichen Xu, Zhukuan Cheng, Jianmin Wan and Lihuang Zhu
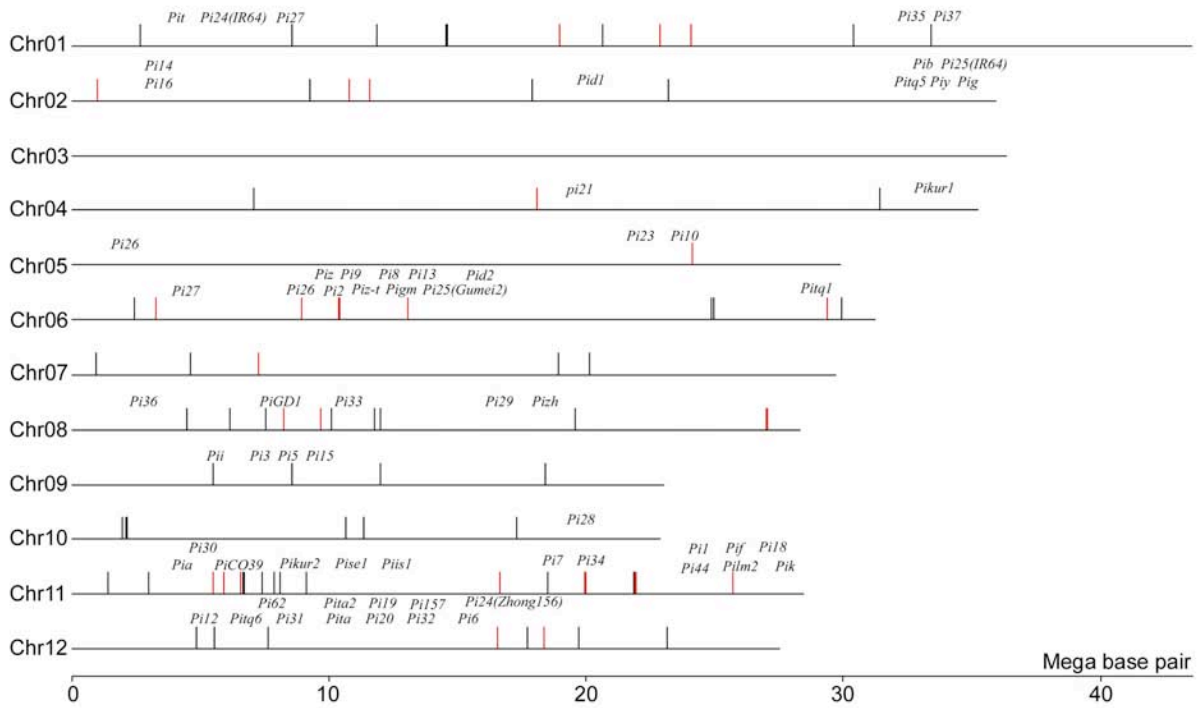
FIGURE S1.—The positions of the 94 allelic NBS-LRR gene pairs each with a pseudogene allele in 93-11 or Nipponbare and the mapped blast *R* genes on the physical map of Nipponbare chromosomes. The horizontal bars represent the physical maps of the Nipponbare chromosomes. The red and black vertical bars represent the allelic pairs each with one pseudogene at the 93-11(59) and Nipponbare locus (35), respectively.

FIGURE S2.—Structure analyses on the 35 pairs of NBS-LRR allelic genes that encoded pseudogenes in Nipponbare. Broad line represents exon, thin line represents intron. "X" means premature stop codon mutation site, "!" means frameshift mutation site.
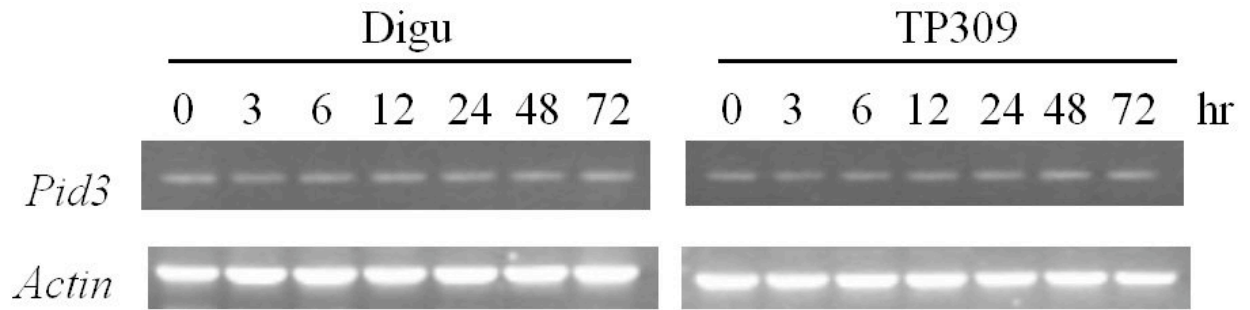
FIGURE S3.—The semi-quantitative RT-PCR analysis of *Pid3* with *Actin* gene as control. The result indicated that *Pid3* was constitutively expressed in both the resistant *indica* variety Digu and the susceptible *japonica* variety TP309 at seven different time points (0, 3, 6, 12, 24, 48, and 72 hours) after inoculation with *M. oryzae* Zhong-10-8-14.
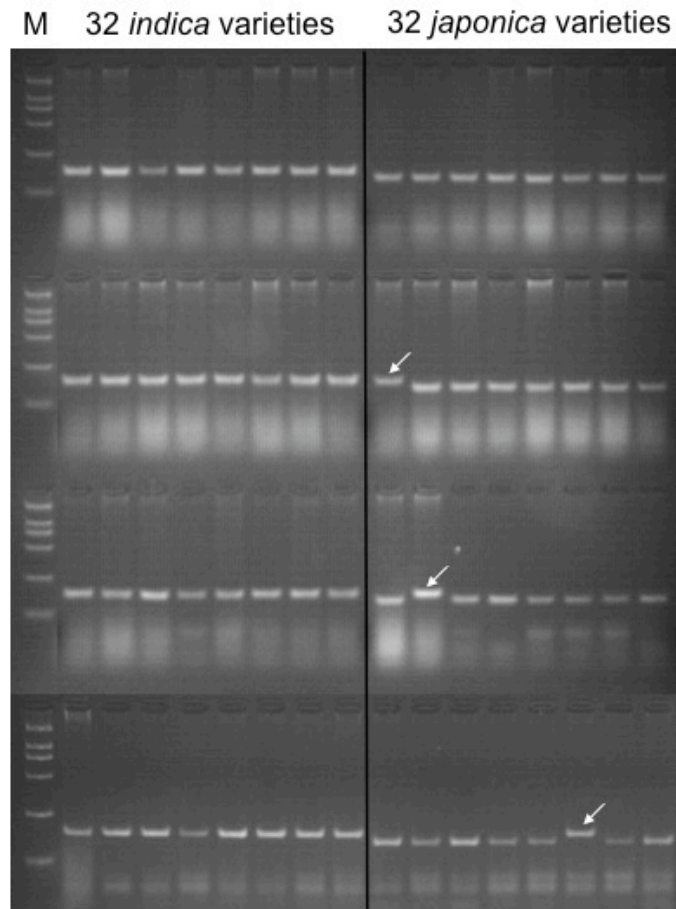
FIGURE S4.— Analysis of the premature stop mutation of *Pid3* alleles with dCAPS marker (See the primers and restriction enzyme used in the dCAPS markers in supplemental file 6) in 32 *indica* varieties and 32 *japonica* varieties. Names of the varieties are listed in supplemental file 1. "M" means molecular marker. Only three of the 32 tested *japonica* varieties do not contain the null T allele, as indicated by arrows, while none of the 32 tested *indica* varieties contain the null T allele.

J. Shang *et al.*

FIGURE S5.— The protein sequence alignment of Pid3 alleles in three *indica* varieties, 93-11, Digu, and ZYQ8, and the three *japonica* varieties, Nipponbare, TP309, and LTH, and the *O. rufipogon* wild rice line A4.

TABLE S1

Rice *indica* and *japonica* varieties used in the study

No.	varieties	subspecies
1	93-11	indica
2	Digu	indica
3	Zhaiyeqing8	indica
4	Minghui63	indica
5	Teqing	indica
6	Gumei2	indica
7	Zhongxian3037	indica
8	Jingang30	indica
9	Xinguangmei	indica
10	Zhefu802	indica
11	Zhenlong13	indica
12	Sifeng43	indica
13	Zhongsi2	indica
14	Zhongsi3	indica
15	Zhongzao9	indica
16	36fu-2	indica
17	XinxianB	indica
18	Yanhui559	indica
19	Tetep	indica
20	Minghui78	indica
21	Minghui86	indica
22	IR24	indica
23	IR36	indica
24	IR64	indica
25	CO39	indica
26	Gui630	indica
27	Xinxian97	indica
28	Hudali	indica
29	Nanjing6	indica
30	Yizhila	indica
31	Yanhui559	indica
32	Xiguxianxing	indica
33	Nipponbare	japonica
34	Lijiangxintuanheigu	japonica
35	Jingxi17	japonica
36	TP309	japonica
37	Lansheng	japonica

38	Jiangnanxiangnuo	japonica
39	Hejiang18	japonica
40	Guandong51	japonica
41	Ningjing16	japonica
42	Zhongzuo37	japonica
43	Longjing8	japonica
44	Donglong363	japonica
45	Hejiang19	japonica
46	Wuyujing3	japonica
47	Wuyujing7	japonica
48	Kongyu131	japonica
49	Bayuenuo	japonica
50	Xiaotiansanbaibang	japonica
51	Xiushui11	japonica
52	Zhi7	japonica
53	Shennong1033	japonica
54	KL908	japonica
55	Xindao11	japonica
56	Balila	japonica
57	Zhongdan2	japonica
58	2428	japonica
59	8706	japonica
60	C418	japonica
61	Nai527	japonica
62	Zhonghua9	japonica
63	Zhonghua15	japonica
64	Lemont	japonica

TABLE S2

The thirty-six lines of wild rice and four varieties of African cultivated rice used in the study

Accession	IRGC No.	Species	Geographic origin
A3	-	*O. rufipogon*	China Jiangxi Dongxiang
A4	-	*O. rufipogon*	China Jiangxi Dongxiang
25	105404	*O. rufipogon*	China
28-2	105616	*O. rufipogon*	China
30	105618	*O. rufipogon*	China
30-2	-	*O. rufipogon*	China
38	100920	*O. rufipogon*	Malaysia
40	106046	*O. rufipogon*	India
W1	W0107	*O. rufipogon*	-
C15	104489	*O. rufipogon*	-
23	103305	*O. rufipogon*	Philippine
A8	101508	*O. nivara*	India
12	101980	*O. nivara*	India
14	105343	*O. nivara*	India
17	105706	*O. nivara*	Nepal
18	105717	*O. nivara*	Kampuchea
19	105879	*O. nivara*	Bengal
C2	100593	*O. nivara*	-
C31	105706	*O. nivara*	-
28-1	-	*O. nivara*	China
15	105428	*O. nivara*	Sri Lanka
W6	W0049	*O. barthii*	-
3	104119	*O. barthii*	Lake Chad
C12	103912	*O. barthii*	-
C34	105991	*O. barthii*	-
A9	101776	*O. longistaminata*	Senegal
C22	105260	*O. longistaminata*	-
C27	105563	*O. longistaminata*	-
WR9	101230	*O. longistaminata*	Sierra Leone
C24	105288	*O. meridionalis*	-
C25	105303	*O. meridionalis*	-
C11	103812	*O. glumaepatula*	-
C29	105662	*O. glumaepatula*	-
C30	105668	*O. glumaepatula*	-
8	100970	*O. glumaepatula*	Brazil
10	105561	*O. glumaepatula*	Columbia
A10	101853	*O. glaberrima*	Burkina Faso

J. Shang *et al.*

5	100127	*O. glaberrima*	Gambia
6	103544	*O. glaberrima*	Mali
7	104161	*O. glaberrima*	Guinea

TABLE S3

The annotated CDS, protein sequence, mutation sites, chromosome locations of all the NBS-LRR genes in

9311 and Nipponbare genomes

Table S3 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.109.102871/DC1.

J. Shang *et al.*

TABLE S4

The primers and restriction enzymes used in the PCR-based markers to identify NBS-LRR gene and pseudogene pairs between *indica* and *japonica*

Gene_ID	Chromosome	Forward Primer / Reverse Primer	Enzyme	9311	Nipp
N016A01T246N05	1	GCTACTGCGTATCGCCAGGAGGATTGCATCG / TTGATGAACCTCCATGTGTA	TaqI	129bp	99bp
N013A02N162P09	1	CGCAATCGCAACAATGAGACG / CCTCCCATTCCATGCAGCGT	ScrFI	229bp, 220bp	449b
N017A01N235P09	1	GAACATCTTGGCAACATAAG / CATGCTGCTTGCTGTAGCCA	MfeI	95bp	73bp
					147b
N035B01T215P09	2	GATCAAGGATCTCTGTCATA / ATGTTTCCTATTTCTTCTGG	BseMII	299bp	152b
N027B02T357P13	2	TTAGAGGTCCTGTGTAAAGATCGA / CATCTGCTCCTGTCCTTATC	ClaI	146bp	122b
N036B01T143N09	2	GGAGGGTCAGCCTTATGGGTGACCATAT / CGGCATAATGTAAGGGTCCA	NdeI	142bp	170b
N066D01T364N11	4	CTATGTGCATTTGGATCTGT / CAGTCTCCTCTTGTTCTAATCGCCACAGATC	BglII	220bp	190b
N082E02T323N11	5	TGGTGTAAGAAAGCTGCCTT / ATGCCAGTGATCAGTTCAGT		156bp	130b
N093F01N317P09(Pid3)	6	TACTACTCATGGAAGCTAGTTCTC / GCAGCACTTCTTGACTACTGTCTGT	XbaI	178bp	153b
N093F01N317P09(Pid3)	6	TACTACTCATGGAAGCTAGTTCTC / ACGTCACAAATCATTCGCTC	BamHI	510bp,148bp	658b
N090F02T340N10	6	TATATCAAAGATGAGCTAAAAAC / TGGCTTTCAATATGGACTTTAAATTCAC	ScrFI	160bp	132b
N102F01N161P09	6	TTGTCTTGGATGATGTATGGGATAAAGATGTTCG / GTATGATAGAAGGCTTCCGA	TaqI	314bp	283b
N098F05N361P12	6	AGTTCAACATCGGTCTATAC / GCTCTTGAACTCTTTGAAAA	Hpy178III	108bp	145b
					223b
N088F01N316P09	6	GCTACTTATATCTCCCACAA / TAGGATGTTGGAGAGCTGAT	MaeI	482bp	259b
N087F02N318P08	6	CTCGGATGCTTCCTCCTCGA / AGCAGGAGTCGTTGTCATCGGTGCAGCACG	TspRI	152bp	124b
N090F04T340N10	6	AAGGTTGCTCGAATTAATCAGTGCCAATGC / CATGCATTACATGGAAAGTT	CjePI	125bp	153b
					128b
N109G01T183N09	7	CCATCAACTTGACAACAAAGCC / CTTCCATGATCACCATCTAG	PvuII	174bp	45bp
N127H01T080N10	8	GCAACACCTTTATACAGTACA / TCTACAGACACTGAAGCTGT	FokI	177bp, 219bp	396b
N137H07T217N12	8	CGCTCGAGGAATTGTTTATT / ACAAGCTGATCATTGCTGAC	ScrFI	178bp, 95bp	273b
N137H04N227P12	8	ATACAGGGGTCTTGAAGTGT / GTCTATGTATATTAGACCTAGAAAGGTCAAGATATT	PsiI	176bp	143b
N125H01T029N09	8	AAATGTCGTGGACTAGAATT / GAAGCATGGCTTGAGATAAG	DdeI	192bp	139b
N150J13T354N05	10	GCGAATGATCTGGTAAATGACTCATCTTCCTAGC / TCCATACTGAATTTCTTCTTTG	AluI	107bp	78bp
N150J17T355N04	10	TCTCCTAAGCCCTATGACAC / GTGAAAGTTTCACACTCCAT	Hpy178III	128bp, 138bp	268b

N150J14T354N05	10	ACATTTGGAAGCGCAAACCCTATGGACCA / CGGATCCCTTTCAATCTGGA	BccI	136bp	163b
N188K02T261N10	11	GCTGATCCATCTTGAACCTGA / GAGAATCAACCTGCCAACGA		343bp	no pı
N177K01T065N09	11	GTATGACATCGAGGACAGCGT / CTTTTATACAAGGTTGATAGA		273bp	no pı
N163K03N089P13	11	GACAGATGACATGTGCTGGA / TGTTTGTTGAAGGAAGCACA	BcgI	280bp, 186bp	466b
N164K04T086N10	11	TCGTGCTCATCTGCCTCAGA / CCCATGGAAAGAACGGCCGT		300bp	no pı
N181K06T212N09	11	GTTGCCTCAAGCATCTTGGA / TACAACTCGAAGGTCCAGGTPCR	DdeI	275bp, 147bp, 114bp	542b
N181K05T000P08	11	GTTGTTGGATTTGGAGGATT / CATCAACAACAATCAAGTACC	MaeI	157bp, 61bp	245b
N177K03T191N09	11	AGCATCATGGGGTGAAGCACAACTTATCAG / TGGCAACATCAAGAATCTGA	Hpy188I	166bp	136b
N162K01T258N04	11	CATCCTCTCCAGATGGCTGT / CCAGGAACCCTACAGGTAAG	EcoRV	235bp, 194bp	429b
N174K01N203P13	11	TGATTTTATACCCATCCTGA / TCATCACCAGTGGTTATGAT	XbaI	193bp, 182bp	375b
N164K08T369P09	11	GAAGAAGGTTTCTTGAAGAA / AGCTTTATAACTTGCTGGAT	MseI	87bp	74bp
N203L02N309P08	12	TATTCTTCTTCAGGGTTCCATACTGGATACAC / CAGGAACCTCGATTGCTTGCAAGTC	ScrFI	122bp	153b
N207L01T156P09	12	CCCGGTCACAGAAAGTTGCT / CGGCTTAGAAGCTCTTCAAA	BsrI	240bp, 256bp	496b