Rapidly Evolving R Genes in Diverse Grass Species Confer Resistance to Rice Blast Disease

Major Classification: BIOLOGICAL SCIENCES; Minor Classification: Evolution

Rapidly evolving *R*-genes in diverse grass species confer resistance to rice blast disease

Sihai Yang [a,1], Jing Li [a,1], Xiaohui Zhang [a,1], Qijun Zhang [b], Ju Huang [a], Jian-Qun Chen[a], Daniel L. Hartl [c,2], Dacheng Tian [a, 2]

[a] State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, China; [b] Jiangsu Academy of Agricultural Sciences, Nanjing, China; [c] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA

[1] S.Y., J.L., and X.Z. contributed equally to this work.

[2] To whom correspondence should be addressed.

E-mail: dhartl@oeb.harvard.edu (D.H.) and dtian@nju.edu.cn (D. T.)

Short title (50 character limit): Trans-specific plant pathogen resistance

Abstract

We show that the genomes of maize, sorghum, and brachypodium contain genes that, when transformed into rice, confer resistance to rice blast disease. The genes are *R*-genes (resistance genes) that encode proteins with nucleotide-binding site (NBS) and leucine-rich repeat (LRR) domains (NBS-LRR proteins). Using criteria associated with rapid molecular evolution, we identified three rapidly evolving *R*-gene families in these species as well as in rice, and transformed a randomly chosen subset of these genes into rice strains known to be sensitive to rice blast disease caused by the fungus *Magnaporthe oryzae*. The transformed strains were then tested for sensitivity or resistance to 12 diverse strains of *M. oryzae*. A total of 15 functional blast *R*-genes were identified among 60 *NBS-LRR* genes cloned from maize, sorghum, and brachypodium; and 13 blast *R*-genes were obtained from 20 *NBS-LRR* paralogs in rice. These results show that abundant blast *R*-genes occur not only within species but also among species, and that the *R*-genes in the same rapidly evolving gene family can exhibit an effector response that confers resistance to rapidly evolving fungal pathogens. Neither conventional evolutionary conservation nor conventional evolutionary convergence supplies a satisfactory explanation of our findings. We suggest a new mechanism termed constrained divergence, in which both *R*-genes and pathogen effectors can follow only limited evolutionary pathways to increase

fitness. Our results open new avenues for *R*-gene discovery that will help to elucidate *R*-gene versus effector mechanisms and may yield new sources of durable pathogen resistance.

Introduction

Ascomycete fungi of the *Magnaporthe oryzae* species complex are ancient pathogens of grasses first described more than a century ago (1). The agent of rice blast disease, *M. oryzae,* is the most devastating pathogen of rice, and it can also infect other important crops including wheat and barley. The pathogen evolves rapidly and exhibits a multitude of species-specific and cultivar-specific races (2-5). Species in the *M. oryzae* complex have therefore become a leading model for the study of interactions between plants and their pathogens (3).

Plants have several layers of defense against pathogens. The first consists of passive structural barriers including cell walls and a waxy cuticle. The second line of defense is an active response triggered by transmembrane pathogen-associated molecular pattern (PAMP) recognition receptors, which act against slowly evolving protein motifs that may be shared among multiple different types of pathogens (6, 7). The third level of response defends against specific pathogen races. In this process, pathogen effectors activate host resistance genes (*R*-genes). The main class of *R*-genes encodes proteins containing a nucleotide-binding site (NBS) along with leucine-rich repeats (LRRs). The NBS domains contain conserved motifs that have been demonstrated to bind and hydrolyze ATP and

GTP, and the LRR motif is typically involved in protein-protein interactions. Such NBS-LRR proteins play an important role in the recognition and resistance to diverse pathogens ranging from viruses, bacteria, and fungi to insects and nematodes (8).

NBS-LRR proteins can act through variety of mechanisms. One mechanism acts via direct interaction with pathogen effectors to render them ineffective (the gene-for-gene hypothesis) (9). Another mechanism features indirect protection of host plant molecules targeted by the pathogen effectors, wherein effector modification of the protected protein activates the *R*-gene response. Indirect protection may secure host plant molecules that are essential (the guard hypothesis) (8) or host plant molecules that mimic those that are essential (the decoy hypothesis) (10).

R-genes encoding NBS-LRR proteins are among the most highly amplified gene families in plants. *Arabidopsis thaliana* has about 150 *NBS-LRR* genes, and *Oryza sativa* has over 400 (8). Amplification of *R*-genes often involves tandem or segmental duplication, and hence the genes tend to be clustered. The rate of evolution of *R*-genes varies according to subfamily (6, 8). Some evolve relatively slowly, while others exhibit typical features of rapid evolution including multiple and variable copy number, short branches in the phylogenetic gene tree, notably low divergence between paralogs within a genome, a high ratio of nonsynonymous to synonymous substitutions (K_a/K_s ratio), and high levels of within-species polymorphism (11, 12).

It has been suggested that the allele frequencies of rapidly evolving, highly polymorphic pathogen effectors and those of their corresponding *R*-genes might undergo

cyclic, out-of -phase changes (6). A caricature of this process is as follows: When the *effector(+)* allele is at high frequency, the corresponding *R(+)* allele is favored and increases in frequency, whereupon the *effector(–)* allele becomes favored and increases in frequency at the expense of *effector(+)*; this favors an intrinsically more fit *R(–)* allele at the expense of *R(+)*; and round and round they go (6).

We reasoned that, if this cyclical model were correct, then the cycles may persist for many generations, and perhaps even through speciation of both the host plant and the pathogen (barring severe population bottlenecks). Were this the case, then one might discover that rapidly evolving *R*-gene families in any plant species might include alleles that confer resistance to an effector present in some races of a pathogen that affects a different but related species.

To test this prediction, we characterized *R*-gene families encoding NBS-LRR proteins in the genomes of maize, sorghum, and brachypodium, and identified those subfamilies undergoing rapid evolution. We cloned a subset of genes from these subfamilies and transferred them into rice lines susceptible to rice blast disease. We discovered that 25% of the cloned genes conferred resistance to at least one of 12 independent isolates of *M. oryzae*. These results demonstrate that a diverse repertoire of *R*-genes for rapidly evolving pathogens occurs not only within species but also among species and that there is functional similarity among NBS-LRR proteins undergoing rapid evolution in different species. Exploiting this diversity could be an important source of *R*-genes for rice blast disease and perhaps other plant pathogens.

Results

Identification of rapidly-evolving NBS-LRR families in distantly related species. Four well-sequenced genomes — maize (*Zea mays*), sorghum (*Sorghum bicolor*), brachypodium (*Brachypodium distachyon*), and rice (*Oryza sativa*) — differ extensively in the number of genes encoding NBS-LRR proteins. The respective numbers of such genes previously identified are 95, 184, 212, and 458 (12). To identify rapidly evolving gene families for the cloning of candidate *R*-genes, a gene tree based on maximum likelihood was constructed using the amino acid sequences encoded in 949 genes as well as 17 functional *R*-genes previously cloned in these species (Fig. S1). A total of 16 large gene families were identified in this gene tree (Table 1; Fig. S1) based on the following criteria: four or more copies of genes on average per species, two or more recently duplicated paralogs (nucleotide divergence < 5%), and branches that are adjacent to the root of the tree well supported by high bootstrap values (≥ 70% in Fig. S1). These features are characteristic of rapidly evolving gene families.

For each of the 16 large gene families, we calculated the number of paralogs having < 5% divergence with the other paralogs from the same genome, the average copy number per species, and the ratio K_a/K_s. By these criteria, three rapidly evolving gene families stand out (Table 1, Table S1):

• Family 2 (Fig. S1), which we call the *Rp1*/*Pi37* gene family because includes the maize *Rp1-D* gene for rust resistance and rice *Pi37* gene for blast resistance (Fig. 1a,b;

Table 1), has the maximum number of recently duplicated paralogs that have a divergence < 5% within the same genome, indicating a large number of young duplicates in these species. (In three out of the four species, this is the only family in which young duplicates are observed.) In addition, a high level of K_a/K_s (1.37) was also observed between paralogs in this family (Table 1).

• Family 3 (Fig. S1), designated *AC134922* (Table 1), has the largest average copy number (8.7) per species and the largest number of paralogs with low divergence in rice (6 in total). In addition, the family's copy number and paralog divergence differ dramatically among cultivars, and the K_a/K_s ratio is high in comparison with rice *NBS-LRR* genes (Table 1) (13). Members of this gene family represent the most diversified locus in rice genomes based on current annotations.

• Family 10 (Fig. S1), called the *Rp3/Pc* gene family because it includes maize *Rp3* gene for rust resistance and sorghum *Pc-B* gene for resistance to root rot caused by *Periconia circinata* (Fig. 1c), has the highest level of K_a/K_s (3.22) between paralogs (Table 1; Table S1) and the second greatest number of young duplicates (Table 1). Both the *Rp1/Pi37* and the *Rp3/Pc* gene families are well known as typical rapidly evolving *R*-gene families (11, 12). Based on these signatures of rapid evolution, we focused on these three families as possible sources of rice blast resistance.

Cloning and screening blast *R*-genes in the three families. Based on the identification of these gene families and the convenience of transformation in rice, we set

out to test: (1) whether the paralogous genes of the *Rp1*/*Pi37* gene family from maize, sorghum, and/or brachypodium could confer resistance to rice blast disease; (2) whether the rice paralogs of the *Rp3*/*Pc* gene family include *R*-genes against blast disease; and (3) whether the rapidly evolving family *AC134922* includes *R*-genes against blast disease.

To carry out these tests, we cloned genes with their native promoters and terminators from the *Rp1*/*Pi37*, *Rp3*/*Pc*, and *AC134922* gene families and transformed them into rice. All cloned genes have been sequenced and annotated; all have complete and normal NBS-LRR coding regions with no signs of pseudogenization. Twelve blast isolates from 85 strains collected throughout China were chosen based on their high level of nucleotide diversities among six avrulence (*Avr*) genes, their geographical distribution, and their ability to produce large numbers of spores (Table S2). Phylogenetic analysis shows that each strain is distinct from the others (Fig. S2). The transformed lines of rice were exposed to spores from each of these 12 strains of rice blast disease and classified for their sensitivity (S) or resistance (R) to each blast strain (Fig. 1e).

Cloning of paralogs of the *Rp1*/*Pi37* family was carried out with DNA fragments prepared by long PCR from 9 maize, 6 sorghum, and 1 brachypodium lines selected at random (Table S3). In addition, 3 highly resistant rice lines were used (Table S3). To clone a complete set of genes in this family, 14 primer pairs were designed for PCR of the gene groups A to G (Fig. 1a and Table S4). In total, 62 paralogs were successfully cloned and sequenced to confirm their complete agreement with the NBS-LRR coding regions and the natural promoters and terminators (Fig. 1b and Table 2). All intact genes were then

transferred into two rice cultivars (TP309 and Shin2) highly susceptible to blast disease. For each of the cloned genes, we obtained sufficient seeds for the subsequent testing stage from first-generation (T1) or second-generation (T2) plants.

Repeated tests and screening of the transformants showed that many of genes from all the selected species consistently conferred resistance to rice blast disease (Fig. 1b, e). For example, two genes from the G-group, when transferred into either of two susceptible lines, confer resistance to 2–3 blast strains (Table S5). In total, in clones from 6 of the 7 gene groups in Fig. 1b, 12–75% of transgenic lines were resistant. (No resistant lines were obtained from group D). Altogether, from the *Rp1*/*Pi37* gene family we identified and confirmed 17 new functional blast *R*-genes. Furthermore, in almost all resistant lines, the ratio of resistant to susceptible plants in approximately 30 T2 plants was not significantly different from 3 : 1, indicating a single copy of the transgene in most of the resistant lines.

To test the *Rp3*/*Pc* family, we carried out transformation with 7 rice homologs cloned from 3 resistant lines representing all major groups in the family (Fig. 1c). In these experiments, however, we were unable to clone the complete gene and regulatory sequences from species other than rice. We therefore designed a vector with the promoter and terminator of the rice gene *Pi9*. After much effort, three maize genes were successfully cloned (Fig. 1c). The 7 rice and 3 maize genes were then subjected to the test procedures described above. Screening of 9 blast strains each with 3 replicates yielded 5 functional blast *R*-genes in the *Rp3*/*Pc* gene family from rice and maize. In total, 20–80% of transgenic lines in the groups of family 10 conferred resistance. Finally, we cloned,

transformed, and tested 8 rice homologs from the family *AC134922* (Fig. 1d), and found that 6 of the 8 conferred resistance to 1–3 strains of rice blast.

The resistance of the *R*-genes we observed might result merely from anomalously high expression levels of the transgenes. To test this possibility, two genes in the *Rp1*/*Pi37* family and one gene in the *Rp3*/*Pc* family were sampled to measure their gene expression level (Fig. S3a). When the two genes in the *Rp1*/*Pi37* family, cloned from sorghum and maize, were transformed to the susceptible lines of rice (Shin2), we found that the expression levels of the transgenic *NBS-LRR* genes were respectively 35.7 and 47.1 fold-lower than their endogenous controls, an indication that the genes are not highly expressed. In addition, the intra-species transformation of rice from GM2 to Shin2 from the *Rp3*/*Pc* family reveals no significant difference in expression level. After 48 hours of pathogen inoculation, the expression levels of these three genes in the transgenic individuals had, respectively, 17.2-fold, 10.4-fold and 7.4-fold increases relative to the transgenic plants prior to pathogen inoculation (Fig. S3b). The expression levels of the genes in these experiments are similar to those in previous reports (14-16).

Discussion

Using conventional criteria for assessing the rate of molecular evolution, we identified three families of *R*-genes encoding NBS-LRR proteins in maize, sorghum, and brachypodium undergoing rapid evolution. Individual members of these gene families were transformed into sensitive strains of rice and the transformants tested for resistance

to each of 12 diverse strains of *M. oryzae* causing rice blast disease. Among 60 *R*-genes tested from maize, sorghum, and brachypodium, a total of 15 conferred resistance to at least one of the rice blast strains. We also tested members of the corresponding *R*-gene families in the rice genome, and among 20 *R*-genes not previously known to be associated with blast resistance, 13 conferred resistance to at least one pathogenic strain. Measurements of gene expression among a subset of the resistant transformants indicate that resistance to blast disease conferred by the *R*-gene cannot be attributed to overexpression of the transgenes. The main implications of these findings are discussed below.

Conservation, convergence, or constrained divergence. The molecular mechanisms of the cross-species *R*-gene interaction with rice blast effectors remain to be determined. Quite possibly, the 15 trans-specific and 13 intraspecific *R*-genes discovered in this study will turn out to exhibit a diversity of different mechanisms of resistance. Some *R*-genes might exhibit direct gene-for-gene interaction (9), and others might act according to the guard hypothesis (8) or to the decoy hypothesis (10). There may well be novel mechanisms among these *R*-genes that do not fit any of these models.

The evolutionary mechanisms of the cross-species *R*-gene resistance to rice blast disease also remain to be studied. There are precedents for cross-species resistance. In one example, a maize *NBS-LRR* gene, *Rxo1*, confers resistance to maize bacterial stripe disease caused by *Burkholderia andropogonis* as well as resistance to rice bacterial streak disease

caused by the unrelated pathogen *Xanthomonas oryzae pv. oryzicola*; in this case, however, *Rxo1* does not exhibit signatures of rapid evolution, and the *X. oryzae* pathogen evolves slowly (17). In another example, the pepper *NBS-LRR* gene, *Bs2*, confers resistance to bacterial spot disease caused by *Xanthomonas campestris pv. vesicatoria* expressing the bacterial avirulence gene *avrBs2*; the same gene confers resistance to tomato spot disease, and so, in this case, the pathovars are the same (18). The situation we describe is different from previous findings in that the *R*-gene families we studied undergo rapid evolution, and so does the rice blast pathogen.

Classical convergent evolution affords a possible mechanism for our findings provided that the effectors in the pathogens that elicit the *R*-gene response in maize, sorghum, and brachypodium are identical to, or molecular mimics of, those in the rice blast strains. On the other hand, it seems likely that significant molecular variation occurs among these effectors, because trans-specific *R*-gene resistance to rice blast occurs only in specific subsets of rice blast strains, differing according to the particular *R*-gene. The variation in resistance among rice blast strains argues against conventional convergent evolution and suggests an evolutionary process whose outcome resembles convergence but whose molecular mechanism is different.

We propose a model based on the observation that, in the evolution of antibiotic resistance, the target protein of an antibiotic can follow only limited evolutionary pathways to greater fitness (19, 20). If this were also the case for both rapidly evolving pathogen effectors and their rapidly evolving *R*-genes, then the evolutionary trajectories of pathogen

effectors would be expected to inosculate or intertwine, as would the evolutionary trajectories of their corresponding *R*-genes. In this way, the rapidly evolving effectors in related pathogens, and their interacting *R*-genes, would have evolutionary landscapes whose realizable trajectories would sometimes coincide, and sometimes not. Rapidly evolving *NBS-LRR* genes are notably under positive selection and contain rich trans-specific polymorphisms (21), suggesting that genetic diversity in these genes is selectively maintained, which increases the likelihood of recognizing rapidly evolving pathogens.

This model of protein coevolution is best understood by analogy with the paper-and-pencil game of tic-tac-toe. Different games may yield a final pattern of noughts and crosses that matches exactly, even though the trajectory of moves in each game was different owing to differences in the opponent's individual strategy. We are not suggesting that molecular evolution is as simple as tic-tac-toe, but we do suggest that the analogy may be apt if molecular coevolution admits of changes in a restricted subset of amino acid residues in the interacting molecules. There is, as far as we know, no name for the intertwining of molecular interactions owing to limited accessible evolutionary pathways, although the term *constrained divergence* seems to capture its essential features.

From a process of constrained divergence, one might expect to find *R*-genes that evolved in response to pathogen effectors in one species to sometimes also respond to effectors in some, but not all, strains of a pathogen that affects a related species. This spotty congruence -- some *R*-genes from one species conferring resistance to some but not all strains of a pathogen in a related species -- is exactly the pattern that we see. The

pattern is consistent with constrained divergence, as distinct from classical convergent evolution, however we put this mechanism forward only as a hypothesis to be tested by additional observations and experiments.

Constrained divergence as an evolutionary mechanism for trans-specific pathogen resistance is supported by the observation that two of the *R*-gene subfamilies we studied confer resistance to quite different plant diseases. In particular, the Rp1/Pi37 family includes *R*-genes for maize rust resistance at the *Rp1-D* locus (22) as well as for rice blast resistance at the *Pi37* locus (15); and the Rp3/Pc family includes maize *R*-genes against rust resistance at the *Rp3* locus (23) as well as sorghum *R*-genes against root rot caused by *Periconia circinata* at the *Pc-B* locus (16). It is remarkable that the same *R*-gene families can confer resistance to both blast and rust diseases. Both types of pathogens are known to evolve rapidly (4) and have been suggested to share similar mechanisms of delivering effector proteins (24).

Novel source of *R*-genes against rapidly evolving pathogens. The large number of blast *R*-genes identified in this study demonstrates that an efficient approach to identifying new *R*-genes is to seek them among paralogs in gene families carefully chosen based on evolutionary analysis. This approach might be as effective or more effective than map-based cloning or other methods commonly employed. The number of rice blast *R*-genes identified previously using other methods is on the order of 19 (4), whereas the number of new rice blast R-genes identified in this single study is 28.

Equally as important, the high efficiency of *R*-gene identification in our experiments suggests that the *NBS-LRR* genes in rapidly evolving families are most likely to confer resistance to fast-evolving pathogens, and that the *NBS-LRR* genes from the same ancestor retain a similar function between species. The 28 newly identified rice blast *R*-genes (15 from maize, sorghum, and brachypodium as well as 13 from rice) distribute in all three families: 17 are from the Rp1/Pi37 family, 5 from the Rp3/Pc family, and 6 from the Ac134922 family. The abundance of functional blast *R*-genes in these families confirms the principle of efficient cloning based on evolutionary characteristics. The topology of the gene trees of the three families (Fig.1) clearly shows that the newly identified *R*-genes are phylogenetically widely distributed among groups with high bootstrap support, indicating that they derived from the same common ancestor, possibly even prior to speciation. While this approach to *R*-gene discovery opens new opportunities for research on *R*-genes and their interactions with pathogen effectors, whether this approach will yield *R*-genes that are agriculturally useful and durable has yet to be determined.

Materials and Methods

Plant materials and rice blast sources. Seeds of the lines used were from various resources (Table S3). Genomic DNA was extracted from fresh leaves using the CTAB method. The primers were designed based on the sequenced genomes (Table S4). The products of long PCR were inserted into the binary vector pCAMBIAI3000. All clones, validated by sequencing, were transferred into blast-susceptible rice cultivars using

Agrobacterium strain EHA105 (25). All transgenic lines (T0) were reproduced to obtain enough seeds (T1 or T2). The presence of the transgenic DNA fragments in the T0 plants was confirmed by PCR.

Blast strains were collected and isolated from different areas of China in 2008-2009 (Table S2). The phylogenetic tree of the 12 selected rice blast strains constructed using nucleotide sequences from six *Avr* genes (*Avr-pita, Avr-pia, Avr-pik, Avr-pii, Avr*-pwl2 and *Avr-Ace-1*) (15, 22, 23, 26) by maximum likelihood and bootstrapped 1000 times with a Kimura two-parameter model using MEGA v5.02 (Fig. S2) (27). The T1 or T2 generations of transgenic lines were used to evaluate blast resistance (28). Seeds from the T1 or T2 plants were tested in three independent replicates of a plate consisting of a 4 × 8 grid of pots, in which each pot contained about 10 seedlings of a given transformed line. Each plate also included two negative (sensitive) and two positive (resistant) controls. When the seedlings were 3 weeks old, the plates were inoculated by spraying blast spore suspension ($2.5–7.5 \times 10^5$ spores/ml) and then placed in a dew inoculation incubator at 26 °C and 100% humidity in the dark for 24 h. In total, each transformed line was screened with whichever 9 of the 12 blast strains yielded sufficient spores for 3 replicate inoculations. The inoculated plants were then transferred to a greenhouse with a 12/12h light/dark photoperiod at 90% relative humidity for 7 days before the disease reaction was examined.

Blast resistance assays. Each transformed line was tested in 3 independent replicates, with approximately 10 plants of each transformed line in each replicate, and rigorous criteria were used to classify any transformed strain as having the R (resistance) phenotype

(Fig 1e). First, in any replicate, two R control cultivars (TTP and GM2) were required to be clearly resistant, and four S (sensitive) controls from rice cultivars (TP309 and Shin2) and from null-gene vector transgenic lines of TP309 and Shin2 were required to be clearly sensitive. Second, the transformed line was required to exhibit a consistent R phenotype against at least one of the 12 blast strains across all 3 replicates. Third, all lines classified as R were confirmed in additional tests with 3 additional independent replicates including additional controls consisting of a sensitive line rendered resistant by transformation of with *Pi9* or another line rendered resistant by transformation with *Pi37* (both *Pi9* and *Pi37* are well known blast *R*-genes). Fourth, any line classified as R was required to show an equal or more strongly resistant phenotype than these *Pi9* and *Pi37* controls in the corresponding blast strains.

Gene expression assays. Total RNA was isolated from leaf using the RNAiso Plus (Takara) according to the manufacturer's instructions. First-strand cDNA synthesis was carried out using a PrimeScript 1st Strand cDNA Synthesis Kit (Takara) according to the manufacturer's instructions. Real-time qPCR was carried out using a CFX96 (BioRad) real-time PCR system. Gene-specific primers were designed using Oligo 6 software. All quantitative reverse transcription-PCR was carried out in triplicate. PCR was performed using the following cycling parameters: 95 ºC for 2 min, 40 cycles of 95 ºC for 15s, and 60 ºC for 30s. The expressions of genes were quantified using the ΔΔCT method in comparison with the endogenous control (actin) gene in this study.

Phylogenetic analysis. To select the loci for this study, protein sequences of *NBS-LRR* genes were initially aligned by MUSCLE (29). Then a phylogenic tree was constructed by maximum likelihood using amino acid sequences identified previously (11, 12) and bootstrapped 1000 times with a Kimura two-parameter model using MEGA v5.02 (27). Based on the gene tree, we chose the clades which included four or more genes per species, two or more recent duplicated paralogs (nucleotide divergence < 5% between paralogs), and were well supported by high bootstrap values (≥70 % in Fig. S1) with the branches adjacent to the root. The nucleotide divergence and ratio of the non-synonymous (K_a) to synonymous (K_s) nucleotide substitution (K_a/K_s) of each family were calculated by MEGA v5.02 (27). The average copy number of each selected family was calculated as the total number of genes in this family divided by the number of species observed in this family based on Figure S1. Because the xxLxLxx (L = Leu or other aliphatic amino acid; x = any amino acid) motif of the LRR domain, which often exhibit strikingly fast rates of evolution, is assumed to be a determinant of recognition specificity for *Avr* factors (30), the K_a/K_s were calculated in these motifs among paralogs in each family (Table S1).

Author contributions: D.T., D.L.H., S.Y. and J.Q.C. designed the experiments and analyses. S.Y. organized all aspects of the project and analyzed the data. J.L., X.Z., Q.Z. and J.H. did the gene cloning, transformation, screening and sequencing for cloned genes. D.L.H. and D.T. wrote the paper.

The authors declare no conflicts of interest.

Acknowledgements. We are very indebted to Dr. Kirsten Bomblies for her critical review and comments on an earlier version of this paper, and to Drs. Elizabeth A. Kellogg and Robert E. Pruitt for their many helpful suggestions. This work was supported by NSFC of Jiangsu province (BK2011015), National Natural Science Foundation of China (30930049, 30930008 & 31071062), and PCSIRT (IRT1020) to D.T. or J.Q.C.

This article contains Supporting Information online.

References

1. Ou SH (1985) *Rice Diseases, 2E* (Commonwealth Mycological Institute, Kew, Surrey) p 411.
2. Skamnioti P & Gurr SJ (2009) Against the grain: safeguarding rice from rice blast disease. *Trends Biotechnol.* 27:141-150.
3. Thakur S, Jha S, Roy-Barman S, & Chattoo B (2009) Genomic resources of *Magnaporthe oryzae* (GROMO): A comprehensive and integrated database on rice blast fungus. *BMC Genomics* 10:316.
4. Liu J*, et al.* (2010) Recent progress and understanding of the molecular mechanisms of the rice-*Magnaporthe oryzae* interaction. *Mol. Plant Pathol.* 11(3):419-427.

5. Xue M*, et al.* (2012) Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. *PLoS Genet.* 8:e1002869.
6. Jones JDG & Dangl JL (2006) The plant immune system. *Nature* 444:323-329.
7. Bent AF & Mackey D (2007) Elicitors, effectors, and I Genes: The new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol.* 45:399–436.
8. McHale L, Tan X, Koehl, & Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 7(4).
9. Flor HH (1971) Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology* 9(1):275-296.
10. van der Hoorn RA & Kamoun S (2008) From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell* 20:2009-2017.
11. Yang S*, et al.* (2006) Genome-wide investigation on the genetic variations of rice disease resistance genes. *Plant Mol. Biol.* 62(1-2):181-193.
12. Li J*, et al.* (2010) Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol. Genet. Genomics* 283(5):427-438.
13. Yang S*, et al.* (2008) Genetic variation of NBS-LRR class resistance genes in rice lines. *Theor. Appl. Genet.* 116(2):165-177.
14. Hulbert SH, Webb CA, Smith SM, & Sun Q (2001) Resistance gene complexes: evolution and utilization. *Annual review of phytopathology* 39:285-312.

15. Lin F*, et al.* (2007) The blast resistance gene *Pi37* encodes a nucleotide binding site leucine-rich repeat protein and is a member of a resistance gene cluster on rice chromosome 1. *Genetics* 177(3):1871-1880.

16. Nagy ED & Bennetzen JL (2008) Pathogen corruption and site-directed recombination at a plant disease resistance gene cluster. *Genome Res.* 18(12):1918-1923.

17. Zhao B*, et al.* (2005) A maize resistance gene functions against bacterial streak disease in rice. *Proc. Natl. Acad. Sci. U.S.A.* 102(43):15383-15388.

18. Tai TH*, et al.* (1999) Expression of the *Bs2* pepper gene confers resistance to bacterial spot disease in tomato. *Proc. Natl. Acad. Sci. U.S.A.* 96(24):14153-14158.

19. Lozovsky ER*, et al.* (2009) Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl. Acad. Sci. U. S. A.* 106:12025-12030.

20. Weinreich DM, Delaney NF, DePristo MA, & Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111-114.

21. Wang J, Zhang L, Li J, Lawton-Rauh A, & Tian D (2011) Unusual signatures of highly adaptable *R*-loci in closely-related Arabidopsis species. *Gene* 482(1-2):24-33.

22. Collins N*, et al.* (1999) Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. *Plant Cell* 11(7):1365-1376.

23. Webb CA*, et al.* (2002) Genetic and molecular characterization of the maize *rp3* rust resistance locus. *Genetics* 162(1):381-394.

24. Panstruga R & Dodds PN (2009) Terrific protein traffic: The mystery of effector protein delivery by filamentous plant pathogens. *Science* 324(5928):748-750.

25. Hiei Y, Ohta S, Komari T, & 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(2):271-282.

26. Jia Y, Bryan GT, Farrall L, & Valent B (2003) Natural variation at the *Pi-ta* rice blast resistance locus. *Phytopathology* 93(11):1452-1459.

27. Tamura K*, et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28(10):2731-2739.

28. Hayashi K, Hashimoto N, Daigen M, & Ashikawa I (2004) Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus. *Theor. Appl. Genet.* 108(7):1212-1220.

29. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32(5):1792-1797.

30. Ellis JG, Lawrence GJ, Luck JE, & Dodds PN (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell* 11(3):495-506.

Table 1. The evolutionary features of the three cloned and the other large *NBS-LRR* gene families identified from the phylogenetic tree of the four distantly related species of grasses (Fig. S1).

Large families		Number of paralogs with low divergence	Average copy number in each species	K_a/K_s
Cloned families in this study	2 *Rp1*/*Pi37* [a]	12 (2R, 2S, 8M)	6.0	1.37
	3 *AC134922* [b]	6 (R)	8.7	1.16
	10 *Rp3*/*Pc* [c]	8 (5M, 3S)	8.0	3.22
Other families	1	4 (2 B, 2 S)	4.0	1.51
	4	8 (S)	6.3	1.07
	5	4 (B)	6.0	2.86
	6	2 (M)	6.0	0.84
	7	4 (B)	7.0	1.14
	8 (*Pd3*)	2 (S)	4.3	1.03
	9	6 (4R,2S)	6.5	0.63
	11	2 (S)	8.0	0.47
	12	4 (R)	6.8	0.45
	13 (*Pi2/9*)	2 (R)	5.0	1.32
	14 (*Pik*)	2 (S)	4.3	0.56
	15	2 (R)	4.0	0.86
	16	2(R)	6.7	0.74

The number of paralogs with low divergence was counted only for these paralogs that have a divergence <5% with one or more paralogs from the same genome. This number reflects how many recent duplicates are in a gene family. S, B, R and M represent sorghum, *Brachypodium*, rice and maize, respectively. [a] The *Rp1*/*Pi37* gene family has the maximum number of recent duplicated paralogs (12). [b]The *AC134922*e gene family has the largest average copy number (8.7) per species. [c] The *Rp3*/*Pc* gene family has the highest level of K_a/K_s (3.22) between paralogs.

Table 2. Cloned *R*-genes, their transgenic lines, and the evaluation of rice blast resistance among them.

Loci	Primers groups	Species	Number of clones	Transgenic lines screened	Number of *R*-lines	Percentage of *R*-lines
Family 2 *Rp1*/*Pi37*	A	*Z. mays*	24	29	9	31.0%
	B	*Z. mays*	20	25	3	12.0%
	C	*S. bicolor*	7	11	2	18.2%
	D	*S. bicolor*	5	7	0	0.0%
	E	*B. distachyon*	1	1	1	100%
	F	*O. sativa*	3	5	1	20.0%
	G	*O. sativa*	2	4	3	75.0%
Family 10 *Rp3*/*Pc*	A	*S. bicolor*	0	—	—	—
	B	*O. sativa*	3	5	4	80.0%
	C	*O. sativa*	3	6	2	33.3%
	D	*O. sativa*	1	1	NE	NE
	E	*Z. mays*	3	5	1	20.0%
Family 3 *AC134922*	A	*O. sativa*	1	2	0	0.0%
	B	*O. sativa*	—	—	—	—
	C	*O. sativa*	4	8	4	50.0%
	D	*O. sativa*	3	4	3	75.0%
R-gene Controls	*Pi37* a	*O. sativa*	2	2	2	100%
	Pi9 b	*O. sativa*	1	2	2	100%
Null-vector Control	—	—	1	2	0	0

NE, not evaluated due to insufficient seeds for screening.

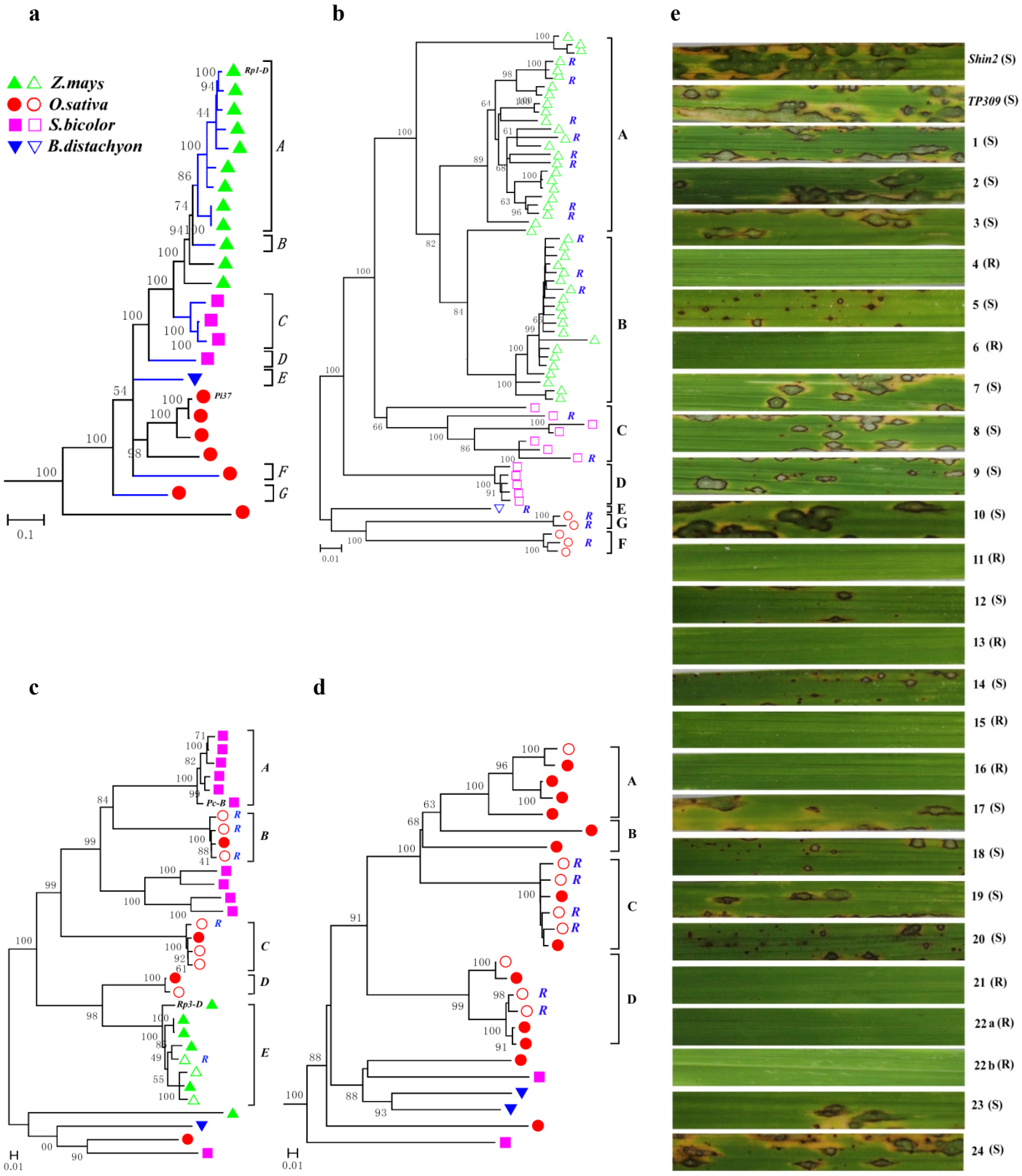
a. Two *Pi37* alleles were cloned from cultivars of Nipponbare and No. 4, respectively, which have been transformed into the highly susceptible cultivar TP309.

b. The construct *Pi9* was provided by Dr. Bo Zhou of Zhejiang University, which has been transformed into the rice cultivar TP309 and Shin2 as *R*-gene controls.

Figure Legend

Figure 1. Gene trees of selected *R*-gene subfamilies and phenotypes with respect to rice blast disease. (a) Gene tree of the Rp1/Pi37 subfamily based on genome sequences. (b) Gene tree of the Rp1/Pi37 subfamily based on cloned sequences. (c) Gene tree of the Rp3/Pc subfamily based on genome sequences (solid symbols) and cloned sequences (open symbols). (d) Gene tree of the Ac134922 subfamily based on genome sequences (solid symbols) and cloned sequences (open symbols); only the subgroup of the Ac134922 subfamily with cloned genes is shown. In (c)–(d), the letters A–F denote the groups of primers used for cloning (Table 2 and Table S4), and the symbol R in blue beside an open symbol indicates resistant to one or more blast strains while the other open symbols indicate sensitivity to blast strains. (e) Sensitive and resistant rice blast phenotypes of transformants containing maize genes in group A of the Rp1/Pi37 subfamily, numbered consecutively from top to bottom. The top two leaves are susceptible controls. The scales of each tree are labeled at the bottom; a larger scale was used for the tree in (a) because of the high divergence among these genes.

Fig.1


a
Z.mays
O.sativa
S.bicolor
B.distachyon
Rp1-D
Pi37
Rp3-D
A
B
C
D
E
F
G
0.1
b
A
B
C
D
E
G
F
0.01
c
Pc-B
Rp3-D
A
B
C
D
E
0.01
d
A
B
C
D
0.01
e
Shin2 (S)
TP309 (S)
1 (S)
2 (S)
3 (S)
4 (R)
5 (S)
6 (R)
7 (S)
8 (S)
9 (S)
10 (S)
11 (R)
12 (S)
13 (R)
14 (S)
15 (R)
16 (R)
17 (S)
18 (S)
19 (S)
20 (S)
21 (R)
22 a (R)
22 b (R)
23 (S)
24 (S)

Supporting Information:

Figure S1. Phylogenetic gene tree of *R*-genes encoding NBS-LRR proteins from maize, sorghum, brachypodium, and rice.

Figure S2. Phylogenetic tree of the 12 selected rice blast strains constructed by nucleotide sequences from six *Avr* genes.

Figure S3. Comparison of gene expression levels (a) between transgenic lines and (b) between the lines before and after pathogen inoculation.

Table S1. Ratios of nonsynonymous (K_a) to synonymous (K_s) nucleotide substitutions in groups of the selected gene families.

Table S2. Collection locations of the rice blast isolates of *M.oryzae*.

Table S3. Plant materials used to clone *R*-genes in this study.

Table S4. Primers used in this study.

Table S5. Assays of resistant (R) or sensitive (S) phenotypes for transgenic rice lines by rice blast strains.