Comparative Analysis of the Genomes of Two Field Isolates of the Rice Blast Fungus Magnaporthe oryzae

Minfeng Xue1,*, Jun Yang1,*, Zhigang Li1, Songnian Hu2, Nan Yao3, Ralph A. Dean4, Wensheng Zhao1, Mi Shen1, Haiwang Zhang1, Chao Li1, Liyan Liu1, Lei Cao1, Xiaowen Xu1, Yunfei Xing1, Tom Hsiang5, Ziding Zhang6, Jin-Rong Xu7, You-Liang Peng1,*

1 State Key Laboratory of Agrobiotechnology and Ministry of Agriculture Key Laboratory of Plant Pathology, China Agricultural University, Beijing, China, 2 Beijing Genomics Institute, Chinese Academy of Sciences, Beijing, China, 3 State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, China, 4 Fungal Genomics Laboratory, Center for Integrated Fungal Research, North Carolina State University, Raleigh, North Carolina, United States of America, 5 School of Environmental Sciences, University of Guelph, Guelph, Canada, 6 State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China, 7 Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana, United States of America

Abstract

Rice blast caused by Magnaporthe oryzae is one of the most destructive diseases of rice worldwide. The fungal pathogen is notorious for its ability to overcome host resistance. To better understand its genetic variation in nature, we sequenced the genomes of two field isolates, Y34 and P131. In comparison with the previously sequenced laboratory strain 70-15, both field isolates had a similar genome size but slightly more genes. Sequences from the field isolates were used to improve genome assembly and gene prediction of 70-15. Although the overall genome structure is similar, a number of gene families that are likely involved in plant-fungal interactions are expanded in the field isolates. Genome-wide analysis on synonymous to non-synonymous nucleotide substitution rates revealed that many infection-related genes underwent diversifying selection. The field isolates also have hundreds of isolate-specific genes and a number of isolate-specific gene duplication events. Functional characterization of randomly selected isolate-specific genes revealed that they play diverse roles, some of which affect virulence. Furthermore, each genome contains thousands of loci of transposon-like elements, but less than 30% of them are conserved among different isolates, suggesting active transposition events in M. oryzae. A total of approximately 200 genes were disrupted in these three strains by transposable elements. Interestingly, transposon-like elements tend to be associated with isolate-specific or duplicated sequences. Overall, our results indicate that gain or loss of unique genes, DNA duplication, gene family expansion, and frequent translocation of transposon-like elements are important factors in genome variation of the rice blast fungus.

Introduction

Rice blast caused by the heterothallic ascomycete Magnaporthe oryzae (also known as Pyricularia oryzae) is one of the most destructive diseases of rice, which is a staple for over half of the world’s population. This pathogen also infects wheat and other small grains, and poses major threats to global food security [1,2]. In the past two decades, rice blast has been developed as a model system to study fungal-plant interactions. M. oryzae was the first plant pathogenic fungus to have its genome sequenced and made available to the public [3].

In most parts of the world, rice blast is controlled mainly with resistant cultivars. However, M. oryzae is notorious for its ability to overcome resistance based on race-specific R genes [4-6]. New cultivars often lose their resistance within a few years of introduction. Genetic variations in populations of the pathogen have been well-documented in many parts of the world [7,8]. M. oryzae isolates are also known to lose virulence and female fertility during laboratory manipulations [1] and large chunks of genomic DNA can be lost spontaneously during cultivation on artificial media, such as the deletion of over a 40 kb region containing the BUF1 locus [9]. The laboratory strain 70-15 of M. oryzae was generated by backcrossing a progeny from a cross between a rice isolate and a weeping love grass (Eragrostis curvula) isolate with the rice isolate Guy11 from French Guyana [10,11]. It has been used in many laboratories and was selected for genome sequencing [3]. Although most of the 70-15 genome should be from the rice pathogen after backcrossing with Guy11 several times, some weeping love grass pathogen sequences are likely retained. In comparison with Guy11, 70-15 is reduced in female fertility, conidiation, and virulence [12].

To determine the extent of genetic variation among isolates of M. oryzae, we sequenced two field isolates Y34 and P131. Y34 was isolated from Japonica rice in 1982 in Yunnan province, China, while both Indica and Japonica rice cultivars are cultivated [13,14]. Due to rich genetic diversity in rice cultivars and centuries
of rice cultivation, highly diverse rice blast pathogen populations exist in Yunnan [15], and hence Y34 was chosen as a representative from this region for sequencing. The other field isolate, P131, originated from Japan where Japonica rice cultivars are dominant [16,17]. The isolates P131, Y34, and 70-15 differ in some cultural characteristics (Figure S1). These three isolates also carry different avirulence genes and vary in aggressiveness toward different rice cultivars (Table S1). In comparison with 70-15, both Y34 and P131 have slightly larger genomes. The two Asian field isolates share a higher degree of similarity and contain over 200 genes that are absent in 70-15. Many pathogenesis-related genes showed evidence of exposure to diversifying selection when comparing either field isolate (P131 or Y34) to the laboratory strain (70-15). Functional characterization of randomly selected genes specific to the field isolates revealed that they play diverse roles, some of which affect virulence and others important for asexual development. The three strains differ noticeably in the distribution of transposon-like elements. Many of the transposable elements tend to be associated with isolate-specific or duplicated sequences. This study revealed genetic factors involved in genome variation of the rice blast fungus.

### Results

#### Genome sequencing and assembly

The genomes of P131 and Y34 were sequenced with the Sanger (2-fold) and 454 sequencing technologies (18-fold). The combined sequence reads for P131 and Y34 were 793.94 Mb and 843.92 Mb, representing about 20- and 21-fold genome sequence coverage, respectively (Table 1). The 454 sequence reads were assembled into contigs and placed into scaffolds by the Newbler assembler with paired-end information from the Sanger reads. The assembled scaffolds of P131 and Y34 were aligned with the assembled genome sequences (Figure 1). Blocks of such sequences can be found at chromosomal ends. In contrast, P131 and Y34 have 1.69 Mb and 2.56 Mb isolate-specific sequences in 70-15 are absent from the two field isolates. In this analysis, it remains possible that the complete genomes of these three isolates vary in abundance of repetitive sequences and actually have greater size differences.

Scaffolds of P131 and Y34 were aligned with the assembled genome of 70-15 (Figure 1). Overall, most of the 70-15 genome (96%) is also conserved in two field isolates. Only 0.43 Mb of sequence in 70-15 are absent from the two field isolates. In contrast, P131 and Y34 have 1.69 Mb and 2.56 Mb isolate-specific sequences, respectively. In general, isolate-specific sequences were dispersed throughout the genomes. For individual chromosomes, there are regions enriched for isolate-specific sequences (Figure 1). Blocks of such sequences can be found at both ends of chromosome IV and at single ends of chromosomes I, II, III, V, and VI. In M. oryzae, genetic variation and avirulence genes are known to be enriched near the telomeres [18,19]. Comparative analysis of the genomes of these three M. oryzae isolates revealed that genes responsible for variations in virulence and adaptation to the environment may be concentrated at the chromosomal ends.

To locate and verify isolate-specific sequences in the field isolates, we used clamped homogenous electric fields (CHEF) gel electrophoresis to separate the chromosomes. Chromosome size polymorphisms were observed among these three isolates.
Figure 1. Genome organization and gene distribution in P131, Y34, and 70-15. The peripheral circle represents seven chromosomes (numbered I–VII) of 70-15 with their sizes marked in Mb. The second circle of color bands shows the distribution of predicted genes categorized by gene ontology along the chromosomes: red, DNA metabolism; blue, cellular component organization; yellow, carbohydrate metabolism; orange, amino acid and lipid metabolism; violet, transcription; black, signal transduction; green, transport. The third circle of color lines shows the distribution of repetitive DNA elements (percentages of repetitive sequences in 100-kb window) in 70-15 (red), Y34 (blue), and P131 (green). The fourth to ninth circles show the percentage of isolate-specific sequences from pair-wise comparisons in 50-kb windows: 4th, 70-15 sequences absent in P131; 5th, 70-15 sequences absent in Y34; 6th, P131 sequences absent in 70-15; 7th, P131 sequences absent in Y34; 8th, Y34 sequences absent in 70-15; 9th, Y34 sequences absent in P131. Genes unique to 70-15, P131, and Y34 are displayed on the tenth to twelfth circles, respectively, with the same color code as the second circle.

doi:10.1371/journal.pgen.1002869.g001

Figure 2. Electrokaryotypes of P131, Y34, and 70-15 and Southern blot analyses. (A) Chromosomes of P131, Y34, and 70-15 were separated by CHEF electrophoresis. Molecular weights of the markers (chromosomes of Hansenula wingei and Schizosaccharomyces pombe) are labeled on the sides (in Mb). (B) Southern blots of CHEF gels hybridized with the P131- and Y34-unique sequences as the probes. When hybridized with P131_scaffold00006_11, the largest chromosome band was detected only in P131 (left panel). When probed with Y34_scaffold00824_1665, only the smallest chromosome of Y34 was detected (right panel).

doi:10.1371/journal.pgen.1002869.g002

Table 2. Gaps in the genome of isolate 70-15 filled with sequences from the field isolates P131 and/or Y34.

| 70-15 chromosome | P131 | Y34 | P131+Y34 |
|------------------|------|-----|---------|
| I                | 1    | 1   | 6       |
| II               | -    | 3   | 11      |
| III              | 2    | 3   | 4       |
| IV               | -    | 3   | 8       |
| V                | -    | -   | 1       |
| VI               | -    | 1   | -       |
| VII              | 1    | -   | 1       |
| unknown          | 4    | 1   | 4       |
| Total            | 8    | 12  | 35      |

doi:10.1371/journal.pgen.1002869.t002
Among them, 135 genes in 70-15 had incorrect intron annotations. Genes meeting these criteria were then manually annotated. A total of 340 amino acids or 20% longer or shorter in 70-15. A total of 81 candidate genes were identified in the 70-15 genome (Figure 3A), 11.3% had no orthologous sequences in other organisms. Moreover, approximately 10.1% of these *M. oryzae*-specific genes were predicted to encode secreted proteins.

To improve gene annotation in 70-15, we identified the genes that were common to all three isolates and had similar sizes (difference less than 1%) between Y34 and P131 but were 50% longer or shorter in 70-15. A total of 740 genes meeting these criteria were then manually annotated. Among them, 135 genes in 70-15 had incorrect intron annotations. The number of genes with inaccurate start or stop codon predictions was 259 or 15, respectively (Table S2).

The number of genes shared only by the two field isolates (198 from P131 and 220 from Y34) was approximately twice that of those shared by either P131 or Y34 with 70-15 (Figure 3A), implying that the two Asian field isolates share a higher degree of similarity and with about 200 genes that are absent in 70-15. For isolate-specific genes, we found that 51, 136, and 71 genes were unique to P131, Y34, and 70-15, respectively (Figure 3A). All the genes randomly selected for verification were confirmed by PCR to be either shared by two isolates or unique to one specific isolate (Figure S4). As found in 70-15, isolates P131 and Y34 also had various copies of DNA helicase Q genes and LTR elements towards the chromosomal ends [3].

For the genes common in Y34 and P131 but absent in the automated annotation of 70-15, we used their amino acid sequences to search the 70-15 scaffolds. The resulting homologous sequences of 70-15 were then used to search against *M. oryzae* ESTs deposited in GenBank. A total of 81 candidate genes were identified in the 70-15 genome and ESTs (Table S3). Seventy-six of them encoded hypothetical proteins with no known homologs in GenBank. Some of these *M. oryzae* specific genes may be important for the virulence or fitness of the pathogen because all three isolates have these genes. The other five genes had orthologous sequences of unknown functions in Sordariomycetes but were absent in lower fungi, such as Zygomycetes and Saccharomycetales.

To further analyze genetic relatedness of these three isolates, the 10,074 clusters containing one protein from one isolate were selected and the resulting individual protein sequences from each isolate were combined for distance analysis with PHYLIP. As shown in Figure 3B, the two field isolates have a closer relationship to each other than with the laboratory strain 70-15.

**Isolate-specific genes**

Based on analyses of gene content, 51, 136, and 71 genes, respectively, were unique to P131, Y34, and 70-15. Overall, 13% of these isolate-specific genes encoded secreted proteins and 46% of them had no significant homolog in GenBank (Table S4). RT-PCR analyses were performed with 10 and 14 randomly selected P131- and Y34-specific genes, respectively. All the selected genes were confirmed to be expressed in mycelia (Figure S5). While most of the isolate-specific genes were dispersed through the genome, some were located near the chromosomal ends (within 500 kb), which is consistent with the distribution tendency of isolate-specific sequences (Figure 1).

To determine the biological function of these isolate-specific genes, nine Y34-specific genes and three P131-specific genes were selected for functional characterization. For majority of them, the resulting gene deletion mutants had no obvious changes in colony growth, conidiation, or virulence (Figure S6). Their functions in plant infection may be redundant or too minor to be detected under laboratory conditions. However, deletion of one P131 unique gene, *P131_scaffold00208-2*, resulted in a reduction in virulence in infection assays with seedlings of a susceptible rice cultivar (Figure 4A). Deletion of another P131 unique gene, *P131_scaffold01777-7*, resulted in approximately 10% growth reduction on oatmeal tomato agar plates (Figure 4B). Proteins...
encoded by these two P131-unique genes were predicted to be localized in the nucleus. Homologous sequences of these two genes were not found in other sequenced fungal species. Moreover, deletion of one Y34 unique gene, Y34_scaffold00875-3, resulted in approximately 36% reduction in conidiation (Figure 4C). Interestingly, deletion of one Y34-unique gene encoding a putative G protein-coupled receptor (GPCR)-like integral membrane protein with six transmembrane domains resulted in changes in pathogenicity on a rice cultivar carrying the Pi-7 R gene, suggesting that this Y34-unique gene might be the potential AVR Pi-7 gene (data not shown).

Among the genes shared by both field isolates P131 and Y34 but absent in 70-15, 19% had signal peptides for secretion and 12% had transmembrane domains (Figure 3; Table S5). About 70% of these genes had no functional annotation. Strain 70-15 may have lost these genes during the initial genetic cross or after generations had no functional annotation. Strain 70-15 may have lost these genes during the initial genetic cross or after generations. Strain 70-15 may have lost these genes during the initial genetic cross or after generations. Strain 70-15 may have lost these genes during the initial genetic cross or after generations.

**Duplicated genomic sequences**

Duplication is one of the major mechanisms for evolutionary innovation. The total duplicated genomic DNA fragments (longer than 500 bp and greater than 90% identity) were 289 kb, 385 kb, and 825 kb in P131, Y34, and 70-15, respectively. A total of 16, 20, and 155 predicted genes in P131, Y34, and 70-15, respectively, were located in these duplicated sequences (Table S6).

Although duplicated DNA sequences were detected genome-wide in all three isolates, in general chromosomes II, IV, V, and VII had more duplicated DNA sequences than other chromosomes (Figure 5A). For individual chromosomes, the end regions tend to contain more duplicated DNA sequences than the central region. Comparative analysis indicated that P131, Y34, and 70-15 all contained isolate-specific duplicated regions (Figure 5A). However, the laboratory strain 70-15 had significantly more duplicated genes, including the AVR gene PWL2 [22] (Table S6). Other duplicated genes with known functions include LPS glycosyltransferases, MFS transporters, sugar transporters, and carboxypeptidases. Both intra- and inter-chromosomal duplications were observed, but more inter-chromosomal duplications were apparent, and only a small portion of duplication events were conserved in all three isolates (Figure 5A).

**Gene families**

To identify gene families, the entire set of the predicted proteins from all three isolates were clustered with the OrthoMCL program. A total of 38,016 proteins were grouped into 14,189 clusters with each cluster representing a group of putative orthologs. Among these clusters, 195 gene families were identified with more than one member in at least one isolate (Figure 5B), suggesting that 1.37% of the M. oryzae genes may have been evolved by gene family expansion. Among 45 clustered loci duplicated equally in each isolate, 38, 6, and 1 gene loci were duplicated between two, three or four times, respectively, per isolate (Table S7). These gene families might have existed before the divergence of the three isolates. The majority of these gene families were predicted to be involved in synthesis and transport of nutrition and secondary metabolites, suggesting that they may be related to plant infection (Table S7). There were 87 clustered loci duplicated at different frequencies in three isolates (Table S8). Most of these gene families (61 out of 87) contained duplicated genes in only one isolate, and 17 gene families contained gene loci duplicated at least three times in one or more isolates (Table S8), suggesting that they have been expanded or contracted in different strains, possibly during environmental adaptations. For example, one putative calcium P-type ATPase gene was duplicated three times in P131 and Y34, and twice in 70-15. Members of this gene family have been demonstrated to be required for disease development and induction of host resistance [23,24].

For loci duplicated in two isolates but absent in the third one, there were eight in P131 and Y34, five in P131 and 70-15, and twelve in Y34 and 70-15 (Figure 5B; Table S9). Most of these expanded gene families had unknown functions. To confirm the duplication events that were unique to the two field isolates, three genes were selected by Southern blot analysis. All of them were confirmed to be specifically duplicated in P131 and Y34 but not in 70-15 (Figure S7). There were seven, thirteen, and eighteen gene families specifically expanded in P131, Y34, and 70-15, respectively (Figure 5B; Table S10). Most of these isolate-specific gene families contained two or three duplicated members that had unknown functions or no known homologs in GenBank.
Asynchronous to synonymous nucleotide substitution rate (Ka/Ks) analysis

To analyze asynchronous and synonymous nucleotide substitutions, we first identified and removed orthologous genes with large deletions or insertions in any of the isolates from the list of common genes. In total, 9,184 highly conserved orthologs were used to identify nucleotide substitution events. Among them, 7,569 genes had neither synonymous nor asynchronous nucleotide substitution in pair-wise comparisons, indicating that most of the genes were well-conserved among different isolates. Only 428 genes had nucleotide substitutions between P131 and Y34, and 1,651 genes had nucleotide substitutions between 70-15 and P131 or Y34, further indicating that the field isolates had closer relationship with each other than with the laboratory strain. Genes with substitutions in the 70-15 versus P131/Y34 comparison could be categorized into four groups: 414 genes only with synonymous nucleotide substitutions, 697 genes only with asynchronous nucleotide substitutions, 124 genes with Ka/Ks <1, and 6 genes with Ka/Ks >1.

Overall, similar numbers of genes identical between Y34 and P131 but with nucleotide variations in 70-15 were thought to have undergone diversifying versus purifying selections. However, several functional categories of genes, such as those involved in cellular responses to stimuli and organophosphate metabolisms, had more members exhibiting diversifying selection in the two field isolates (Table S11). Several of the genes underwent diversifying selection in the 70-15 versus P131/Y34 comparison (Table S12), including ATG4, HEX1, MCK1, MoSNF1, PTH2, and RGS1, which are known virulence factors in M. oryzae [25–30]. Three of them encode putative CFEM-domain receptors that may be involved in recognizing different environmental and plant signals (Table S12).

Repetitive sequences and transposable elements

Repetitive sequences were masked by Newbler for assembling 454 sequence data of P131 and Y34. To compare repetitive sequences of these two isolates, we assembled the Sanger reads of P131 and Y34 (approximately 2-fold genome coverage) and found that 10.8%, 10.3%, and 10.6% of the 70-15, P131, and Y34 genomes, respectively, were repetitive sequences, indicating that the abundance of repetitive sequences is similar among these three isolates. Transposable elements (TE) and their insertion sites (flanking sequences) were identified by RepeatMasker. Although the exact copy numbers vary, both field isolates contained all classes of transposable elements identified in 70-15 (Table 3). In general, 70-15 has more members of the LINE, Maggy, and

| Table 3. Repetitive and transposable elements identified in isolates P131, Y34, and 70-15. |
|-------------------------------------|---------|---------|---------|--------|---------|
| DNA transposon:                     | 70-15   | Y34     | P131    | 70-15/P131* | 70-15/Y34 |
| Pot2/Pot4                           | 277     | 405     | 341     | 109     | 131     | 135     |
| Occan                               | 77      | 93      | 65      | 44      | 61      | 35      |
| Pot3                                | 68      | 85      | 79      | 29      | 25      | 27      |
| LTR retrotransposon:                |         |         |         |         |         |         |
| Maggy                               | 267     | 60      | 64      | 20      | 23      | 20      |
| MGLR3                               | 81      | 74      | 50      | 30      | 35      | 27      |
| Pyret                               | 232     | 471     | 425     | 160     | 188     | 128     |
| RETRO5                              | 329     | 103     | 86      | 21      | 21      | 37      |
| RETRO6                              | 107     | 108     | 85      | 47      | 40      | 42      |
| RETRO7                              | 119     | 103     | 102     | 51      | 42      | 47      |
| Grasshopper                         | 13      | 42      | 34      | 14      | 10      | 16      |
| LINE:                               |         |         |         |         |         |         |
| MGL                                 | 187     | 141     | 127     | 58      | 53      | 52      |
| SINE:                               |         |         |         |         |         |         |
| Mg-SINE/Mg-MINE/Ch-SINE              | 176     | 195     | 179     | 64      | 58      | 57      |
| New repetitive elements:            |         |         |         |         |         |         |
| cluster1 (JQ929644)<sup>b</sup>    | 25      | 54      | 52      | 33      | 32      | 31      |
| cluster2 (JQ929665)                 | 36      | 23      | 24      | 22      | 21      | 16      |
| cluster3 (JQ929666)                 | 78      | 86      | 78      | 58      | 60      | 45      |
| cluster4 (JQ929667)                 | 17      | 34      | 27      | 7       | 13      | 8       |
| cluster5 (JQ929668)                 | 47      | 58      | 63      | 47      | 43      | 47      |
| cluster6 (JQ929669)                 | 5       | 25      | 13      | 4       | 4       | 8       |
| cluster7 (JQ929670)                 | 22      | 43      | 30      | 14      | 17      | 15      |
| cluster8 (JQ929671)                 | 118     | 101     | 116     | 98      | 88      | 79      |
| cluster9 (JQ929672)                 | 16      | 18      | 15      | 14      | 15      | 12      |

<sup>a</sup>The number of conserved repetitive and transposable elements in the 70-15 and P131 comparison.

<sup>b</sup>GenBank accession number of new repetitive elements. The copy number of repetitive and transposable elements was calculated by RepeatMasker. LTR, long terminal repeat; LINE, long interspersed repeat element; SINE, short interspersed repeat element. Novel repetitive elements were identified by RepeatScout.

doi:10.1371/journal.pgen.1002869.t003
The Pot2/Pot4 DNA transposons and the Pyret and Grasshopper LTR retrotransposons were more abundant in P131 and Y34. In addition, nine new clusters of repetitive sequences were identified by analysis with RepeatScout (Table 3). However, none of them was unique to the field isolates. While clusters 1, 4, 5, 6, and 7 were much more abundant in the field isolates, 70-15 had more copies of the cluster 2 repetitive elements (Table 3).

In comparison with 70-15, the two field isolates were more similar in the distribution pattern of repetitive sequences (Figure 1 and 6A). While chromosomal ends tend to have more repetitive sequences, all three isolates had much reduced numbers of TEs in

**Figure 6. Transposable elements (TEs).** (A) Distribution of TEs on seven chromosomes. The y-axis represents the density probability of TEs along the chromosome. (B) Venn diagram showing the number of TEs with conserved genomic positions in 70-15, Y34, and P131. (C) The percentage of each TE element inserted into gene coding regions of TE-disrupted genes. Red, 70-15; blue, Y34; green, P131.
doi:10.1371/journal.pgen.1002869.g006
the gene-rich regions of chromosomes III, V, and VI (Figure 6A). For the TEs that could be assembled into the genome sequences, approximately 27% of them had the same locations in all three isolates by comparison of their flanking sequences (Figure 6B). Y34 had more TEs with unique chromosomal positions (1,061) than P131 (930) or 70-15 (976). In addition to the 609 locations of TEs conserved among the three strains, Y34 and P131 also shared 201 TEs with the same chromosomal locations, which was fewer than the 277 between 70-15 and Y34 or the 341 between 70-15 and P131 (Figure 6B). While over two-thirds of the members of some TEs, including Occan, had conserved genomic locations, TEs such as Retro5 and Maggy differed significantly in their chromosomal positions between Y34 and 70-15. Similar results were obtained with the P131 and 70-15 comparison (Table 3). A total of 41.1% and 46.0% of TEs in 70-15 and P131, respectively, had conserved genomic locations. The Po3c, Maggy, Retro5, and Retro7 elements had the highest variation in chromosomal positions between 70-15 and P131.

We also analyzed the impact of TEs on the genome evolution by comparing two-fold coverage Sanger data of P131 and Y34 with the 70-15 assembly. A total of 35, 38, and 116 genes were disrupted by the insertion of TEs in P131, Y34, and 70-15, respectively (Table S13, S14, S15). Over 50% of the gene disruptions were caused by TEs belonging to MGL, Mg-SINE, or Pot2/Pot4. Strain 70-15 had a number of genes disrupted by cluster 7, cluster 9, Occan, and RETR05 elements, which were not observed in P131 or Y34 (Figure 6C). Some of these genes may have been disrupted by transposition events occurring during generations of cultivation under laboratory conditions, and these genes may play roles in plant infection or survival in the field isolates but were not required for the laboratory isolate. In comparison with 70-15, the field isolates P131 and Y34 had more genes disrupted by SINE (Figure 6C), which may indicate that these SINE elements were more active in these two field isolates.

Among all the genes disrupted by TEs in three isolates, only approximately one third of them have known functions based on their orthologs in GenBank, and most of them are involved in protein metabolism, transportation, transcription, or lipid metabolism. The majority of the TE-disrupted genes encode hypothetical proteins with unknown functions. Interestingly, 23.8% of them contained putative signal peptide sequences, which is significantly higher than the average percentage of predicted extracellular proteins in the genomes of these three strains (Table S13, S14, S15). Some of them may function as effectors involved in fungal-plant interactions, such as AVR Pi-ta1 in 70-15 (Table S13). In addition, 14.7%, 14.2% and 15.8% of the TE-disrupted genes in 70-15, P131, and Y34, respectively, encoded proteins with putative nuclear localization sequences.

Intriguingly, the regions containing isolate-specific sequences or duplicated genes families were often near areas with high frequency of TEs (Figure S8). In 70-15, several TEs were found within 1.0 kb from 23 duplicated genes families, including the avirulence gene PWL2 (Table S16) although many of these duplicated sequences were not closely linked or located on different chromosomes. Taken together, it is likely that the transposition events of TEs might be related to translocation of duplicated DNA fragments and presence of isolate-unique sequences in these three strains.

Discussion

In a number of eukaryotic organisms, comparative analysis of multiple genomes of the same species has been used to improve assembly and annotation and to identify genome variations [31–34]. The rice blast fungus is well-known for its natural genetic variation [1,2]. In this study, we sequenced two field isolates of M. oryzae from Asia. Genome analysis indicated that these two field isolates are more closely related to each other than to 70-15, which is a laboratory strain derived from three backcrosses of rice pathogen Guy11 with a progeny of a cross involving a weeping love grass pathogen, and maintained for many years under laboratory conditions. The overall genome content and composition are similar among these three isolates, but the genomes of P131 and Y34 with only A/C/T/G and no N's were slightly larger than that of 70-15.

Although the 70-15 genome has been updated several times, it still has many gaps (www.broadinstitute.org/annotation/genome/magnaoptore_grisea). In this study, a total of 55 gaps of the 70-15 genome (version 6) were filled in with sequences from P131 and Y34, and the results were validated by PCR analyses of 70-15. This number of putative filled gaps with sequences from two isolates may seem low, but because of the short read length, the threshold set may have been too stringent. For 35 gaps, they were filled with consensus sequences found in both field isolates. For the gaps with sequences only available in either Y34 or P131, the filling sequence for 70-15 was less certain, but of high probability because the overall nucleotide sequence identity between 70-15 with P131 or Y34 was over 98%. Besides improving the genome assembly, the sequences of P131 and Y34 were used to improve the annotation of 70-15. We identified 81 genes that were not predicted in the automated annotation of the 70-15 genome sequence, and none of them were related to the sequence gaps. In addition, we identified potential annotation errors in 340 predicted genes of 70-15. Most of them were related to the problems with the prediction of the boundaries of introns and start or stop codons.

Our study revealed that each M. oryzae isolate had some unique genomic DNA sequences. Because genome sequences of P131 and Y34 were aligned with that of 70-15, it was impossible to locate most of the sequences unique to Y34 and P131 onto specific chromosomes or chromosomal regions. However, sequences unique to 70-15 were distributed over all seven chromosomes. Because 70-15 was derived from three backcrosses of rice pathogen Guy11 with a progeny of a cross involving a weeping love grass pathogen, we expected that a small portion of its genome was from the weeping love grass pathogen. The isolates Y34, P131, and 70-15 had 136, 51, and 71 unique genes, respectively. Therefore, less than 1% of the predicted genes were unique to each isolate and these genes play diverse roles, some of which might possibly contribute to the specificity of individual isolates. Some of the isolate-specific genes were clustered, suggesting that isolate-specific DNA fragments might be gained or lost during evolution. The P131-specific gene P131_scaffold00208-2 encoded a hypothetical protein without known homologs in other fungi. Deletion of this gene resulted in reduced virulence toward rice plants. Because it might be involved in plant infection, P131_scaffold00208-2 may play an isolate-specific role in suppressing or overcoming plant defense responses. These results suggest that some of the field isolate-specific genes may play important roles in plant infection.

In all three M. oryzae isolates, most of the duplicated genes are functionally unknown. Duplicated sequences are distributed all over seven chromosomes and appear to be enriched in the telomeric regions. For the duplicated genes with known functions, many of them are predicted to be involved in primary and secondary metabolism and interactions with the host (such as cutinases and Avr proteins), which is consistent with earlier
observations with 70-15 [3]. Interestingly, several gene families involved in synthesis and transport of nutrients and secondary metabolites were expanded with different frequencies in these three isolates. Some of these duplicated genes may contribute to the adaption of *M. oryzae* to different environmental conditions.

Among the genes that had undergone diversifying selection in Y34 and P131 in comparison with 70-15, a number of them are known to be important for virulence, suggesting that such genes may have been under strong selection pressure in their natural field environments. There were six genes under positive selection in the two field isolates compared to 70-15. Two of them encoded two hypothetical proteins, a serine/threonine protein kinase, an acyltransferase, a putative catalytic domain of diacylglycerol kinase, and an aspartic-type endopeptidase. Three of them are located on chromosome I. In contrast, there were no genes showing positive selection in the comparison between the field isolates.

Because sexual reproduction has not been observed in the field, it is possible that translocations of the repetitive sequences may be one of the major sources for genome variation and rapid adaption to different host and environmental conditions. Consistent with this hypothesis, over 10% of the genome sequences were found to be repetitive sequences. In addition to TEs that have been identified in previous studies [3], nine new clusters of repetitive sequences were identified in all three *M. oryzae* strains in this study. Most of these TEs have different copy numbers in different isolates. Strikingly, among thousands of TE loci, less than 30% of them were conserved among these isolates, suggesting active transposition of these TEs in *M. oryzae*. Moreover, approximately 200 genes were totally disrupted by TEs in these three strains, and approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation. Interestingly, TEs tended to be found near isolate-specific variations in host-microbe interactions and transcriptional regulation. Approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation. Interestingly, TEs tended to be found near isolate-specific variations in host-microbe interactions and transcriptional regulation. Approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation. Interestingly, TEs tended to be found near isolate-specific variations in host-microbe interactions and transcriptional regulation. Approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation. Approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation. Approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation. Approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation. Approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation. Approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation. Approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation. Approximately 40% of them encoded extracellular or nuclear proteins, suggesting that transpositions of TEs may contribute to variations in host-microbe interactions and transcriptional regulation.
condiation assessed [17]. Mycelia collected from two-day-old cultures in complete media (CM) shaken at 150 rpm were used for extraction of fungal DNA and protoplasts. Media were supplemented with 250 μg/ml hygromycin B (Roche, USA) or 400 μg/ml neomycin (Amresco, USA) to select hygromycin-resistant or neomycin-resistant transformants. Four-week-old seedlings of monogenic rice cultivars (Table S1) and eight-day-old seedlings of barley cultivar ‘E8’ were inoculated as previously described [17,54]. Lesion development was examined 5–7 days after inoculation.

CHEF electrophoresis

Chromosome-size DNA were prepared with protoplasts isolated from vegetative hyphae as previously described [55,56], and separated on 0.65% Megabase agarose (Bio-Rad, USA) gels with a Bio-Rad DR III system with switching intervals of 60 min for 48 h, 55 min for 72 h, 45 min for 72 h, and 35 min for 72 h at 1.5 V/cm. Chromosomal DNA of Schizasaccharomyces pombe and Hansenula Wingei (Bio-Rad, USA) were used as the molecular weight markers.

Generation of the gene replacement constructs and mutants

To generate the P131_scaffold00208-2 gene replacement vector pKOPS208-2, its 0.97 kb upstream and 0.82 kb downstream fragments were amplified with primer pairs P131_scaffold00208-2KO_LBf plus P131_scaffold00208-2KO_RBf, and P131_scaffold00208-2KO_RBf plus P131_scaffold00208-2KO_LBf, respectively. The resulting PCR products were cloned into the pKOPS208-2 KO_RBf plus P131_scaffold00208-2KO_RBf, and P131_scaffold00208-2KO_RBf plus P131_scaffold00208-2KO_LBf, respectively. The resulting PCR products were cloned into the kpn I-HindIII and EcoRI-SphI sites of pKOV21 [56,57]. After linearization with NcoI, pKOPS208-2 was introduced into protoplasts of P131. Hygromycin resistant transformants were isolated and assayed for neomycin-resistance. The resulting transformants were screened by primer pairs P1/I and P2/P12. The putative deletion mutants were identified and confirmed by Southern blot analysis. The same approach was used to generate gene replacement constructs and mutants for isolate-specific genes: P131_scaffold00297-2, P131_scaffold00493-1, Y34_scaffold00875-1, Y34_scaffold00875-3, Y34_scaffold00875-6, Y34_scaffold01193-2, Y34_scaffold00005-1, Y34_scaffold01048-2, Y34_scaffold00105-3, Y34_scaffold00105-2, and Y34_scaffold00855-11. The primer pairs used for generating the gene replacement constructs and for mutant screening are listed in Table S17.

Accession number

The genome sequence data of Y34 and P131 were deposited in the NCBI Genome Database (www.ncbi.nlm.nih.gov/genome) under the accession numbers AHZS00000000 and AHZT00000000, respectively. The nucleotide sequence data of the genomes of three isolates. The 8th to 13th circles represent the distribution of the TEs in 50-kb windows same as the 3rd to 8th circles shown in Figure 1. The 2nd to 4th circles represent duplicated genes families along seven chromosomes. The 5th to 7th circles represent the distribution of the TEs in 50-kb windows same as the 3rd to 8th circles shown in Figure 1. The 2nd to 4th circles represent duplicated genes families along seven chromosomes.

Supporting Information

Figure S1 Comparison of the isolates P131, Y34, and 70-15 on asexual development and plant infection. (A) Colonies of P131, Y34, and 70-15 on OTA plates photographed at 120 hours after inoculation (hpi). (B) Vegetative mycelia of P131, Y34, and 70-15 shaken in liquid CM, photographed at 48 hpi. (C) Seedlings of the susceptible rice cultivar ‘LTH’ sprayed with conidia of P131, Y34, and 70-15, respectively, photographed 7 days after inoculation (day). (TIF)

Figure S2 The whole mitochondrial genomes of P131 or Y34 were compared with that of 70-15. The differences in nucleotide acid substitution or deletion among the three isolates are shown. The mitochondrial genome of Y34 lacks two fragments with a combined length shorter than 350 bp. (TIF)

Figure S3 PCR validation of the gaps in the assembly of 70-15 filled with genome sequences of P131 and Y34. The genomic DNA of 70-15 was used for PCR amplification with the primer pairs listed in Table S17. (TIF)

Figure S4 PCR validation of the selected genes unique to single isolates P131 (A), Y34 (B), or 70-15 (C), and specific to two isolates P131 and Y34 (D), P131 and 70-15 (E), or P131 and Y34 (F). The genomic DNA of isolates P131, Y34, and 70-15 were used for PCR amplification with the primer pairs shown in Table S17. (TIF)

Figure S5 RT-PCR validation of the isolate-specific genes unique to P131 or Y34. Primers listed in Table S17 were used to amplify sequences unique to P131 or Y34 with cDNA synthesized from RNA isolated from vegetative hyphae. (TIF)

Figure S6 Functional analyses of 12 field isolate-specific genes. (A) Colonies of the wild-type strain Y34 and the null mutants of Y34-unique genes KOY875-1 (Y34_scaffold00875-1), KOY857-3 (Y34_scaffold00875-3), KOY857-6 (Y34_scaffold00875-6), KOY1193-2 (Y34_scaffold001193-2), KOY5-1 (Y34_scaffold00005-1), KOY1048-2 (Y34_scaffold01048-2), KOY105-1 (Y34_scaffold00105-1), KOY105-2 (Y34_scaffold00105-2), KOY855-11 (Y34_scaffold00855-11), and the wild-type strain P131 and the null mutants of P131-unique genes KOP208-2 (P131_scaffold00208-2) and KOP1784-1-2-3 (P131_scaffold01784-1-2-3). Representative photographs were taken on OTA plates 3 dai. (B) Barley seedlings sprayed with conidia of the wild-type strains P131 and Y34, and with null mutants of all 12 genes photographed 5 dai. (TIF)

Figure S7 Verification of genes duplicated specifically to the field isolates. Genomic DNA of P131, Y34, and 70-15 were digested by two restriction enzymes. Amplified fragments of the gene P131_scaffold01531-1, P131_scaffold01428-1, and Y34_scaffold00846-5 were used as probes. M, λ-HindIII ladder. (TIF)

Figure S8 Co-distribution of the TEs with the duplicated genes families and isolate-specific sequences in the genomes of P131, Y34, and 70-15. The peripheral circle represents seven chromosomes (numbered I–VII) of 70-15 with their sizes marked in Mb. The 2nd to 4th circles represent duplicated genes families along seven chromosomes. The 5th to 7th circles represent the distribution of the TEs in the genomes of three isolates. The 8th to 13th circles represent the percentage of isolate-specific sequences from pair-wise comparisons in 50-kb windows same as the 3rd to 8th circles shown in Figure 1. Red, 70-15; blue, Y34; green, P131. (TIF)
Table S1 Pathotypes of P131, Y34, and 70-15 based on their infectivity towards different monogenic rice cultivars generated by the International Rice Research Institute.

Table S2 Genes in 70-15 with potential annotation errors adjusted with data from the assembled genomes of P131 and Y34.

Table S3 Genes that were absent in 70-15 version 6 but identified by comparative analysis with the genomes of P131 and Y34, and compared against GenBank NR.

Table S4 Isolate-unique genes in P131, Y34, and 70-15.

Table S5 Genes specific to the field isolates P131 and Y34.

Table S6 Genes predicted within duplicated genomic fragments of P131, Y34, and 70-15.

Table S7 Gene families with the same number of members in all three isolates.

Table S8 Gene families with different numbers of members in each of the three isolates.

Table S9 Gene families specific to two of the three isolates.

Table S10 Isolate-specific gene families.

Table S11 The number of the genes identical between Y34 and P131 but with nucleotide variations in 70-15 and which showed evidence of exposure to diversifying and purifying selection pressures based on GO classification.

Table S12 Genes with only synonymous nucleotide substitutions in field isolates P131 and Y34 compared with the laboratory strain 70-15.

Table S13 Genes of isolates P131 and Y34 mapped against chromosomal assembly of 70-15 and found to be disrupted by TE.

Table S14 Genes of isolates Y34 and 70-15 mapped against chromosomal assembly of P131 and found to be disrupted by TE.

Table S15 Genes of isolates P131 and 70-15 mapped against chromosomal assembly of Y34 and found to be disrupted by TE.

Table S16 Transposable elements located within 1.0 kb of members of the duplicated gene families in 70-15.

Table S17 PCR primers used in this study.

Acknowledgments

We thank Jiye Yan and Xinghong Li at Beijing Academy of Agriculture & Forestry Science for help with CHEF electrophoresis.

Author Contributions

Conceived and designed the experiments: Y-LP J-RX. Performed the experiments: JY MX ZL. Analyzed the data: Y-LP J-RX MX YL. Contributed reagents/materials/analysis tools: Y-LP. Wrote the paper: Y-LP J-RX JY MX ZL RAD TH ZZ.

References

1. Valent B, Chumley FG (1991) Molecular genetic analysis of the rice blast fungus, Magnaporthe grisea. Annu Rev Phytopathol 29: 443–467.
2. Talbot NJ (2003) On the trail of a cereal killer: Exploring the biology of Magnaporthe grisea. Annu Rev Microbiol 57: 177–202.
3. Dean RA, Talbot NJ, Ebbole DJ, Faran ML, Mitchell TK, et al. (2005) The genome sequence of the rice blast fungus Magnaporthe grisea. Nature 434: 980–986.
4. Kiyosawa S (1982) Genetics and epidemiological modeling of breakdown of plant disease resistance. Annu Rev Phytopathol 20: 93–117.
5. Zeigler RS, Leong SA, Feng PS (1994) Rice blast disease. International Rice Research Institute, Wallingford, Oxon (United Kingdom).
6. Orbach MJ, Farrall I, Sveigard JA, Chumley FG, Valent B (2000) A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. Plant Cell 12: 2019–2032.
7. Levy M, Correavictoria FJ, Zeigler RS, Xu SZ, Hammer JE (1993) Genetic diversity of the rice blast fungus in a disease nursery in Colombia. Phytopathology 83: 1427–1433.
8. Kumar J, Nelson RJ, Zeigler RS (1999) Population structure and dynamics of Magnaporthe grisea in the Indian Himalayas. Genetics 152: 971–984.
9. Farman ML (2002) Meiotic deletion at the BUF1 locus of the fungus Magnaporthe grisea is controlled by interaction with the homologous chromosome. Genetics 160: 137–148.
10. Leung H, Borromeo ES, Bernardo MA, Norteghem JL (1988) Genetic analysis of virulence in the rice blast fungus Magnaporthe grisea. Phytopathology 78: 1227–1233.
11. Chao C-CT, Ellingsoe AH (1991) Selection for mating competence in Magnaporthe grisea pathogenic to rice. Can J Bot 69: 2130–2134.
12. Xu JK, Zhao X, Dean RA (2007) From genes to genomes: a new paradigm for studying fungal pathogenesis in Magnaporthe oryzae. Adv Genet 57: 175–210.
13. Zheng F, Yang Q, Zhao Z, Li J (1998) Variability of pathogenicity of Pyricularia oryzae. J Yunnan Agric Univ 13: 20–24.
14. Zeng Y, Li Z, Yang Z, Wang X, Shen S, et al. (2001) Ecological and genetic diversity of rice gynophyl in Yunnan, China. PGR Newslet 125: 24–28.
15. Chen QH, Wang YC, Zheng XB (2006) Genetic diversity of Magnaporthe grisea in China as revealed by DNA fingerprint haplotypes and pathotypes. J Phytopathol 154: 361–369.
16. Yamada M, Kiyosawa S, Yamaguchi T, Hirano T, Kobayashi T, et al. (1976) Proposal of a new method for differentiating races of Pyricularia oryzae in Japan. Ann Phytopathol Society Japan 42: 216–219.
17. Peng YL, Sindayiwa J (1986) Temporal sequence of cytological events in rice leaves infected with Pyricularia oryzae. Can J Bot 66: 730–735.
18. Farman ML, Kim YS (2005) Telomere hypervariability in Magnaporthe grisea. Mol Plant Pathol 6: 287–298.
19. Rehneyer CI, Li W, Kasaba M, Kim YS, Brown D, et al. (2006) Organization of chromosome ends in the rice blast fungus, Magnaporthe oryzae. Nucleic Acids Res 34: 4605–4610.
20. Kulkarni RD, Kelkar HS, Dean RA (2003) An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. Trends Biochem Sci 28: 118–121.
21. Miki S, Matsui K, Kito H, Otsuka K, Ashizawa T, et al. (2009) Molecular cloning and characterization of the AFR-Pa locus from a Japanese field isolate of Magnaporthe oryzae. Mol Plant Pathol 10: 361–374.
22. Sveigard JA, Carroll AM, Kang S, Farrall L, Chumley FG, et al. (1995) Identification, cloning, and characterization of PWL2, a gene for host speciﬁcity in the rice blast fungus, Magnaporthe grisea. Mol Plant Microbe Interact 8: 118–121.
23. Balhade`re PV, Talbot NJ (2001) PDE1 encodes a Ptype ATPase involved in appressorium-mediated plant infection by the rice blast fungus Magnaporthe grisea. Plant Cell 13:1221–2004.
24. Gilbert MJ, Thornton CR, Walkey TJ, Talbot NJ (2006) A Pyr1e ATPase required for rice blast disease and induction of host resistance. Nature 440:555–539.
25. Bhambra GK, Wang ZY, Soanes DM, Walkey GE, Talbot NJ (2006) Peroxosomal carnitine acetyl transferase is required for elaboration of penetration hyphae during plant infection by Magnaporthe grisea. Mol Microbiol 61: 46–60.
26. Jeon J, Goh J, Yoo S, Chi MH, Choi J, et al. (2008) A putative MAP kinase kinase, MCK1, is required for cell wall integrity and pathogenicity of the rice blast fungus, Magnaporthe oryzae. Mol Plant Microbe Interact 21: 525–534.
27. Liu H, Suess A, Willard FS, Siferovska DP, Lu S, et al. (2007) Rgp1 regulates multiple Gα subunits in Magnaporthe pathogenesis, asexual growth and thigmotropism. EMBO J 26: 690–700.
28. Soundararajan S, Jedd G, Li X, Ramos-Pampolona M, Chua NH, et al. (2004) Woronin body function in Magnaporthe grisea is essential for efficient pathogenesis and for survival during nitrogen starvation stress. Plant Cell 16: 1564–1574.

29. Kershaw MJ, Talbot NJ (2009) Genome-wide functional analysis reveals that infection-associated fungal autophagy is necessary for rice blast disease. Proc Natl Acad Sci USA 106: 15967–15972.

30. Yi M, Park JH, Ahn JH, Lee YH (2008) MAvPII regulates sporulation and pathogenicity in the rice blast fungus Magnaporthe oryzae. Fungal Genet Biol 45: 1172–1181.

31. Kells M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 429: 241–254.

32. Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature 428: 617–624.

33. Novo M, Bigey F, Beyne E, Galeote V, Gavory F, et al. (2009) Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1105. Proc Natl Acad Sci USA 106: 16335–16338.

34. Andersen MR, Salazar MP, Schaap PJ, van de Vondervoort PJ, Culley D, et al. (2011) Comparative genomics of citric-acid-producing Aspergillus niger ATCC 1015 versus enzyme-producing CBS 513.88. Genome Res 21: 885–897.

35. Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Dubois MJ, et al. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature 464: 367–373.

36. Tyler BM, Tripathy S, Zhang XM, Dehal P, Jiang RHY, et al. (2006) Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313: 1261–1266.

37. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380.

38. Kurtz S, Phillippy A, Delius M, Bader JS, et al. (2004) Versatile and open software for comparing large genomes. Genome Biol 5: R12.

39. Salamov AA, Solovyev VV (2000) Ab initio gene finding in Drosophila genomic DNA. Genome Res 10: 516–522.

40. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955–964.

41. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, et al. (2008) The Pfam protein families database. Nucleic Acids Res 36: D281–D288.

42. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. (2002) InterPro: an integrated documentation resource for protein families, domains and functional sites. Brief Bioinform 3: 225–235.

43. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6: 175–182.

44. Bendures,JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP3.0. J Mol Biol 340: 783–795.

45. Horton P, Park KJ, Ohayashi T, Fujita N, Harada H, et al. (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res 35: W583–W587.

46. Alsulaim SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.

47. Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178–2189.

48. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302: 205–217.

49. Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6.

50. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.

51. Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular Biology and Evolution 17: 32–43.

52. Wang J, Wong GKS, Ni PX, Han YJ, Huang XG, et al. (2002) RePS: A sequence assembler that masks exact repeats identified from the shotgun data. Genome Res 12: 824–831.

53. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27: 573–580.

54. Park G, Bruno KS, Staiger CJ, Talbot NJ, Xu JR (2004) Independent genetic mechanisms mediate turgor generation and penetration peg formation during plant infection in the rice blast fungus. Mol Microbiol 53: 1695–1707.

55. Luo CX, Yin LF, Koyanagi S, Farman ML, Kusaba M, et al. (2005) Genetic mapping and chromosomal assignment of Magnaporthe oryzae avirulence genes AvrPd, AvrPpi, and AvrPiz1 controlling cultivar specificity on rice. Phytopathology 95: 640–647.

56. Yang J, Zhao X, Sun J, Kang Z, Ding S, et al. (2010) A novel protein Com1 is required for normal conidiom morphology and full virulence in Magnaporthe oryzae. Mol Plant Microbe Interact 23: 112–123.

57. Kong L, Yang J, Li G, Qi L, Zhang Y, et al. (2012) Different chitin synthase genes are required for various developmental and plant infection processes in the rice blast fungus Magnaporthe oryzae. PLoS Pathog 8: e1002526.