



# Genetic Variation Bias toward Noncoding Regions and Secreted Proteins in the Rice Blast Fungus *Magnaporthe oryzae*

Zhenhui Zhong,<sup>a</sup> Meilian Chen,<sup>a,b</sup> Lianyu Lin,<sup>a,c</sup> Ruiqi Chen,<sup>a</sup> Dan Liu,<sup>a</sup> Justice Norvienyeku,<sup>a,c</sup> Huakun Zheng,<sup>a,c</sup> Zonghua Wang<sup>a,b,c</sup>

<sup>a</sup>State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou, China

<sup>b</sup>Institute of Oceanography, Minjiang University, Fuzhou, China

<sup>c</sup>College of Life Science, Fujian Agriculture and Forestry University, Fuzhou, China

**ABSTRACT** The genomes of plant pathogens are highly variable and plastic. Pathogen gene repertoires change quickly with the plant environment, which results in a rapid loss of plant resistance shortly after the pathogen emerges in the field. Extensive studies have evaluated natural pathogen populations to understand their evolutionary effects; however, the number of studies that have examined the dynamic processes of the mutation and adaptation of plant pathogens to host plants remains limited. Here, we applied experimental evolution and high-throughput pool sequencing to *Magnaporthe oryzae*, a fungal pathogen that causes massive losses in rice production, to observe the evolution of genome variation. We found that mutations, including single-nucleotide variants (SNVs), insertions and deletions (indels), and transposable element (TE) insertions, accumulated very rapidly throughout the genome of *M. oryzae* during sequential plant inoculation and preferentially in noncoding regions, while such mutations were not frequently found in coding regions. However, we also observed that new TE insertions accumulated with time and preferentially accumulated at the proximal region of secreted protein (SP) coding genes in *M. oryzae* populations. Taken together, these results revealed a bias in genetic variation toward noncoding regions and SP genes in *M. oryzae* and may contribute to the rapid adaptive evolution of the blast fungal effectors under host selection.

**IMPORTANCE** Plants “lose” resistance toward pathogens shortly after their widespread emergence in the field because plant pathogens mutate and adapt rapidly under resistance selection. Thus, the rapid evolution of pathogens is a serious threat to plant health. Extensive studies have evaluated natural pathogen populations to understand their evolutionary effects; however, the study of the dynamic processes of the mutation and adaptation of plant pathogens to host plants remains limited. Here, by performing an experimental evolution study, we found a bias in genetic variation toward noncoding regions and SPs in the rice blast fungus *Magnaporthe oryzae*, which explains the ability of the rice blast fungus to maintain high virulence variation to overcome rice resistance in the field.

**KEYWORDS** *Magnaporthe*, experimental evolution, genetic variant

Genome plasticity, mediated in part by deletions or insertions promoted by transposable elements (TEs), contributes to the adaptation of fungal pathogens to their hosts (1, 2), which renders fungal pathogens great threats to human health and food security. As with the genomes of other pathogens, the genomes of filamentous fungi encode an arsenal of effectors that play critical roles in the interaction between fungi and their hosts. Fungal effectors are initially secreted to surmount their host’s defenses,

**Citation** Zhong Z, Chen M, Lin L, Chen R, Liu D, Norvienyeku J, Zheng H, Wang Z. 2020. Genetic variation bias toward noncoding regions and secreted proteins in the rice blast fungus *Magnaporthe oryzae*. mSystems 5:e00346-20. <https://doi.org/10.1128/mSystems.00346-20>.

**Editor** Michelle Heck, USDA-Agricultural Research Service, Boyce Thompson Institute, Cornell University

**Copyright** © 2020 Zhong et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Huakun Zheng, huakunzheng@163.com, or Zonghua Wang, wangzh@fafu.edu.cn.

**Received** 18 April 2020

**Accepted** 11 June 2020

**Published** 30 June 2020

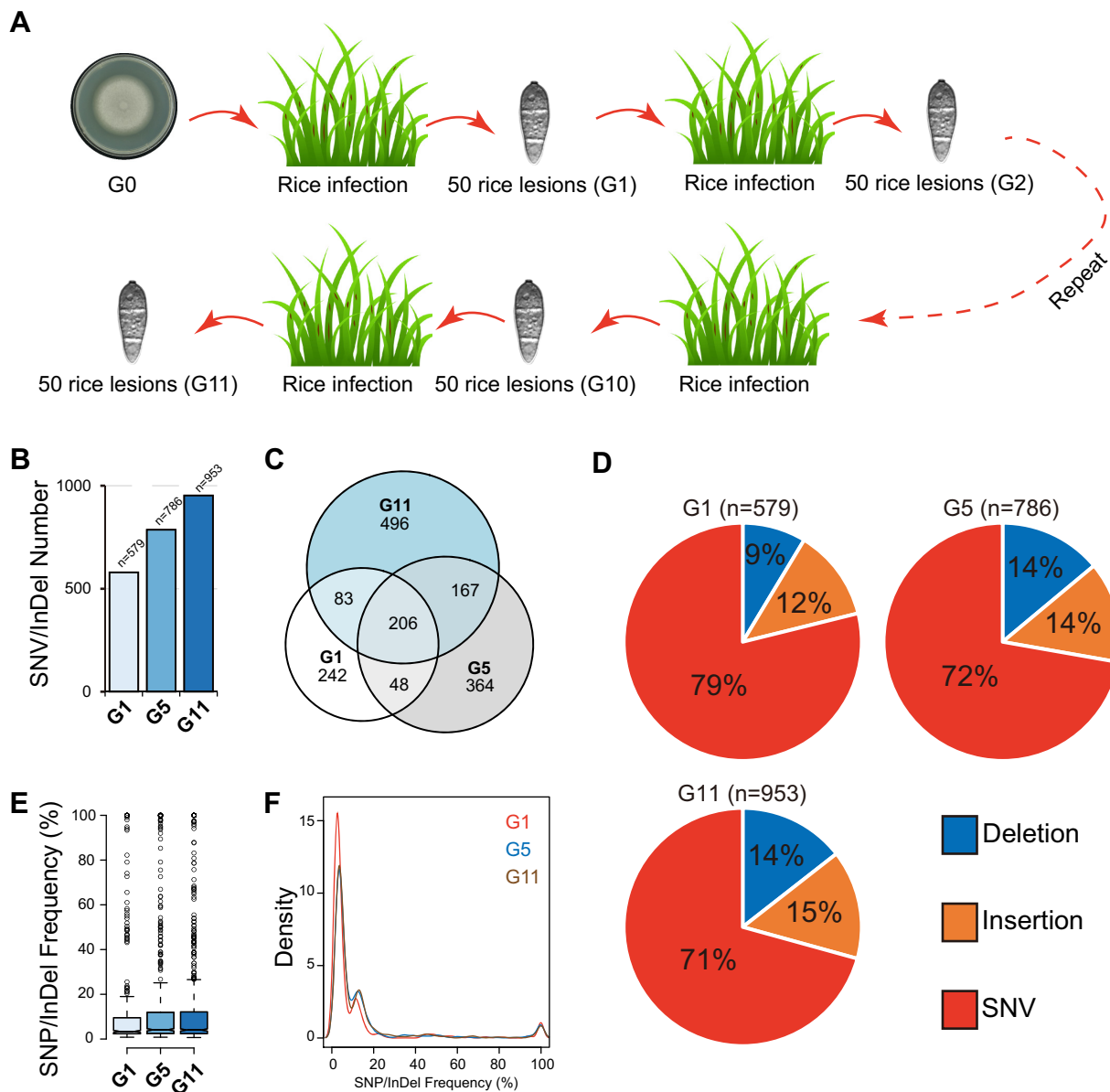
while some of them (called avirulence [*Avr*] genes) can be recognized by host R proteins and cause an avirulence phenotype (3). Most of the effectors identified so far are small secreted proteins (SPs) (4). Surveys of the sequenced genomes of filamentous fungi have indicated that fungal effector genes are not evenly distributed throughout the genome but are enriched in gene-sparse and repeat-rich two-speed genome regions where the genes undergo variation more rapidly than the genes in gene-rich compartments (2, 5, 6).

In the evolution of pathogens, mutations accumulate throughout the genome, while beneficial mutations gradually become fixed and dominant in the population. The fixation of beneficial mutations is highly dramatic and rapid in pathogens under constant selection, for instance, in a highly uniform agroecosystem (7, 8). However, the mechanisms of the evolution of pathogens in the coevolution process and the fixation of beneficial mutations in a population under the constant selection of host plants are still largely unknown. Unraveling the dynamic variations in a plant pathogen's population while interacting with its host can enhance our understanding of pathogen evolution in nature, especially in agroecosystems, and thus provide efficient and environmentally friendly disease management strategies (9).

Many efforts have been made to uncover the dynamic variation in plant pathogen populations; among these efforts, experimental evolution has been well established in microbial population research and has become a powerful tool that is complementary to modern genetic and pathogen epidemiology research (10–14). Compared with population genetics, which ultimately provides information regarding evolution in the context of many undetectable environmental factors, experimental evolution methods can simulate the evolutionary process of organisms under controlled conditions and retain retraceable samples of different stages. The combination of well-developed and inexpensive sequencing technology enables the experimental evolution method to reveal molecular mechanisms under gradual adaptation and provide a real-time perspective of evolution dynamics.

*Magnaporthe oryzae* (syn. *Pyricularia oryzae*), the causative agent of rice blast disease, leads to a 10% to 30% reduction in rice production annually (15). Investigations have shown that *Magnaporthe* species are also capable of causing blast disease in more than 50 plant species of monocot origin in addition to rice, including food crops such as wheat, millet, and barley. In addition, *Magnaporthe* species also infect wild grass hosts such as *Digitaria sanguinalis*, *Setaria viridis*, and *Eleusine indica* (16). A previous study identified a novel avirulent gene, *AvrPi9*, in *M. oryzae* via the sequential planting method, which operates under a theory similar to that of experimental evolution (17, 18). In a previous study, the experimental evolution of *M. oryzae* on artificial medium resulted in a rapid accumulation of mutations prior to the observation of a phenotype; however, the results of that study also revealed a significant reduction in the virulence of the progeny population, which may be due to the lack of an infection stage (19). The genome sequence of *M. oryzae* varied greatly when it adapted to hosts from different species and subspecies of rice (20, 21). Although sequential inoculation *in planta* was performed using a mixture of *M. oryzae* isolates (17, 18), no real-time evolution studies have been conducted *in planta* so far to investigate the role of either host selection or clonality in the rapid evolution of rice blast fungus.

Next-generation sequencing of pooled samples (Pool-Seq) has been used as a cost- and time-effective approach for studying population variability and differentiation (22). In this study, we performed an experimental evolution study by embracing the advantages of the experimental evolution method and high-throughput Pool-Seq technology to monitor the coevolution process by serially inoculating pathogens into host plants to evaluate the gradual variation in *M. oryzae* in rice. We found the rapid accumulation of low-frequency single-nucleotide variants (SNVs), insertions and deletions (indels), and TEs during infection, and interestingly, these mutations were enriched in intergenic regions and the proximal region of SP coding genes, whereas they were depleted in coding regions.



**FIG 1** Experimental evolution of *M. oryzae*. (A) Schematic map of the experimental evolution assay used in this study. The wild-type strain Guy11 was grown on rice bran medium to generate spores that were adopted as the initial generation (G<sub>0</sub>). Seedlings of the susceptible rice cultivar (TP309) were inoculated with spores obtained from G<sub>0</sub> to commence the first infection cycle. In total, 50 compatible leaf lesions from the first infection cycle were randomly collected, sterilized, and then incubated for 7 days on rice bran medium at 26°C under constant light to yield enough G<sub>1</sub> spores. The G<sub>1</sub> spores were used as the inoculum for the next infection cycle. A total of 11 infection cycles were carried out in this study to generate 11 generations of *M. oryzae* populations (G<sub>1</sub> to G<sub>11</sub>). (B) Total number of SNVs/indels identified in G<sub>1</sub>, G<sub>5</sub>, and G<sub>11</sub>. Pool-Seq data. (C) Venn diagram showing overlapping of SNVs/indels identified in G<sub>1</sub>, G<sub>5</sub>, and G<sub>11</sub>. (D) Percentage of different types (deletion, insertion, or SNV) of SNVs/indels identified in G<sub>1</sub>, G<sub>5</sub>, and G<sub>11</sub>. (E) Box plot showing frequency of all SNVs/indels identified in G<sub>1</sub>, G<sub>5</sub>, and G<sub>11</sub>. (F) Density plot showing density distribution of SNVs/indels identified in G<sub>1</sub>, G<sub>5</sub>, and G<sub>11</sub>. The red line represents the G<sub>1</sub> population, the blue line represents the G<sub>5</sub> population, and the brown line represents the G<sub>11</sub> population.

## RESULTS

**Experimental evolution via sequential infection with *M. oryzae*.** To evaluate the host-induced adaptive evolution among succeeding generations of *M. oryzae* populations, the experimental evolution assay was conducted as shown in Fig. 1A. The wild-type strain Guy11 was grown on rice bran medium to generate spores that were adopted as the initial generation (G<sub>0</sub> [G stands for generation herein]). Seedlings of a susceptible rice cultivar (TP309) were spray inoculated with spores obtained from G<sub>0</sub> to commence the first infection cycle (infection cycle refers to 7 days postinoculation). In

total, 50 compatible leaf lesions from multiple plants of the first infection cycle were randomly collected, sterilized, and then incubated for 7 days on rice bran medium at 26°C under constant light to yield enough  $G_1$  spores. The  $G_1$  spores were used as the inoculum for the next infection cycle. A total of 11 infection cycles were carried out in this study to generate 11 generations of *M. oryzae* populations ( $G_1$  to  $G_{11}$ ).

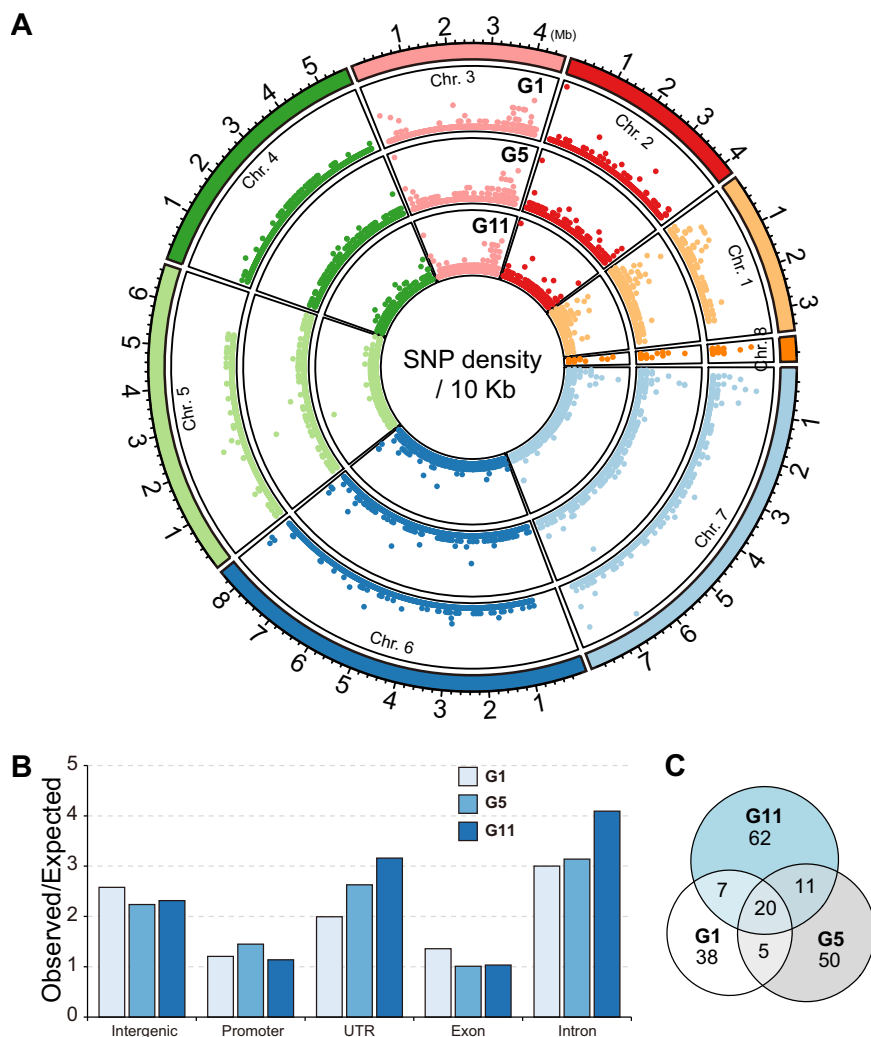
**High-throughput pool sequencing of experimental evolution populations.** To investigate the variation in *M. oryzae* that occurred under constant host selection, whole-genome sequencing was performed using Illumina paired-end sequencing. During the evolution process, mutations occurred in different sites in each individual; however, only beneficial mutations facilitating host adaptation gradually accumulated in the population. We thus selected  $G_1$ ,  $G_5$ , and  $G_{11}$  for in-depth whole-genome sequencing. To minimize the difference between the reference genome and the original isolate ( $G_0$ ) used in this study, the original isolate was also sequenced and used as a control in this study. To investigate mutations at the population level, we used pool sequencing (Pool-Seq), which is a cost-effective method that pools sequences of individual DNAs. For  $G_1$ ,  $G_5$ , and  $G_{11}$ , we randomly selected 50 individual isolates for each generation and combined their DNA in equal amounts for Pool-Seq.

The obtained reads of  $G_0$ ,  $G_1$ ,  $G_5$ , and  $G_{11}$  were mapped to the genome sequence of the *M. oryzae* 70-15 reference genome for single-nucleotide variant (SNV) and insertion and deletion (indel) analyses (see Table S1 in the supplemental material). We performed SNV/indel calling separately for all four samples, and the obtained SNVs/indels were filtered with the SNVs/indels of  $G_0$ . Overall, we obtained 579, 786, and 953 SNVs/indels for  $G_1$ ,  $G_5$ , and  $G_{11}$ , respectively (Fig. 1B). Further analysis of overlapping SNVs/indels in  $G_1$ ,  $G_5$ , and  $G_{11}$  showed that 206 SNVs/indels were shared by the three generations and 242, 364, 496 SNVs/indels were uniquely present in  $G_1$ ,  $G_5$ , and  $G_{11}$ , respectively (Fig. 1C). To evaluate variation sites at the population level, we analyzed the SNV/indel frequency of each site in  $G_1$ ,  $G_5$  and  $G_{11}$  (Fig. 1F). The results showed that  $G_5$  and  $G_{11}$  had similar average SNV/indel frequencies (Fig. 1D and E), while  $G_1$  had more low-frequency sites than  $G_5$  and  $G_{11}$ .

**Genomic distribution of SNVs/indels.** To evaluate the impact of constant selection on the pathogen's genome, we calculated the SNV/indel number in every 10-kb region and observed that the three samples exhibited significant differences in SNV/indel distribution pattern (Fig. 2A). We found the most variation in chromosome 1 (Chr 1), Chr 3, Chr 6, and Chr 7 in the  $G_1$ ,  $G_5$ , and  $G_{11}$  genomes, and the number of SNVs/indels increased consistently in some regions. Consistent with our previous findings that  $G_1$  was more heterozygous and had more low-frequency sites than  $G_5$  and  $G_{11}$ , the distribution of SNVs/indels in  $G_1$  was more diversified than that in  $G_5$  and  $G_{11}$ .

We then questioned whether these SNVs/indels had any bias in genome distribution. To answer this question, we annotated the genomic distribution of SNVs/indels by dividing the genome into five categories: intergenic, promoter (−1,000 to 0 bp), exon, intron, and 5' and 3' untranslated regions (UTRs). We calculated the expected numbers of SNVs/indels in these five categories by randomly selecting an equal number of control sites in the genome and comparing it with the SNVs/indels of  $G_1$ ,  $G_5$ , and  $G_{11}$  (observed number). Interestingly, we found that the ratios of observed number/expected number of SNVs/indels were much higher in intergenic (2.58, 2.24, and 2.32) and intron regions (3.00, 3.14, and 4.09) than in exon regions (1.36, 1.01, and 1.03) in  $G_1$ ,  $G_5$  and  $G_{11}$ , respectively, suggesting that SNVs/indels were preferentially enriched in intergenic and intron regions than in exon regions (Fig. 2B). In addition, the ratios of observed number/expected number of SNVs/indels in the promoter and exon were close to 1, suggesting no bias in distribution in these regions.

To determine the effects of SNVs/indels on genes, transcripts, protein sequences, and regulatory regions, we next investigated the effect of these SNVs/indels on gene products. Consistent with the SNV/indel distribution results, most of the SNVs/indels that caused upstream or downstream gene variations had no significant effect on gene products. Some variation types identified in this study lead to amino acid changes,



**FIG 2** Annotation of SNVs/indels identified in the experimental evolution assay. (A) Chromosomal distribution of SNVs/indels. The numbers of SNVs/indels in 10-kb windows were calculated and plotted. (B) Genomic distribution of SNVs/indels identified in  $G_1$ ,  $G_{5r}$ , and  $G_{11}$ . The number was calculated by comparing the expected number of SNVs/indels in each of five categories (intergenic, promoter, exon, intron, and UTR) with the observed number of sites in the genome. (C) Venn diagram showing overlapping SNV/indel-affected genes in  $G_1$ ,  $G_{5r}$ , and  $G_{11}$ . Only frameshift variants, in-frame deletions, in-frame insertions, missense variants, start codon losses, stop codon losses, and stop codon gains that changed amino acid sequences were analyzed.

including frameshift variants, in-frame deletions, in-frame insertions, missense variants, start codon losses, stop codon losses, and stop codon gains (Table 1). We found that SNVs/indels with these variations occurred in 193 genes, among which 20 were shared by 3 samples and 38, 50, and 62 were presented uniquely in  $G_1$ ,  $G_{5r}$ , and  $G_{11}$ , respectively (Fig. 2C). Functional annotation indicated multiple functions of these 193 genes, involved in pathways ranging from the metabolic pathway and transcription regulation to the mating signal transduction pathway (Table S2). Overall, although SNVs/indels accumulated rapidly in the genome, it appeared that coding regions were not a hot spot of variation.

**Dynamics of TEs in experimental evolution populations.** Transposable elements (TEs) play important roles in genome shaping and genome stability in *M. oryzae* (23, 24). Plant pathogens can avoid plant host immunity recognition through TE insertion-mediated silencing of some avirulent genes (25, 26). We thus set out to analyze TE dynamics in experimental evolution populations. We first annotated newly formed TE insertion events in  $G_1$ ,  $G_{5r}$ , and  $G_{11}$  by annotating all TE insertion events in the genome

**TABLE 1** Quantification of different types of effects from SNV/indel mutations

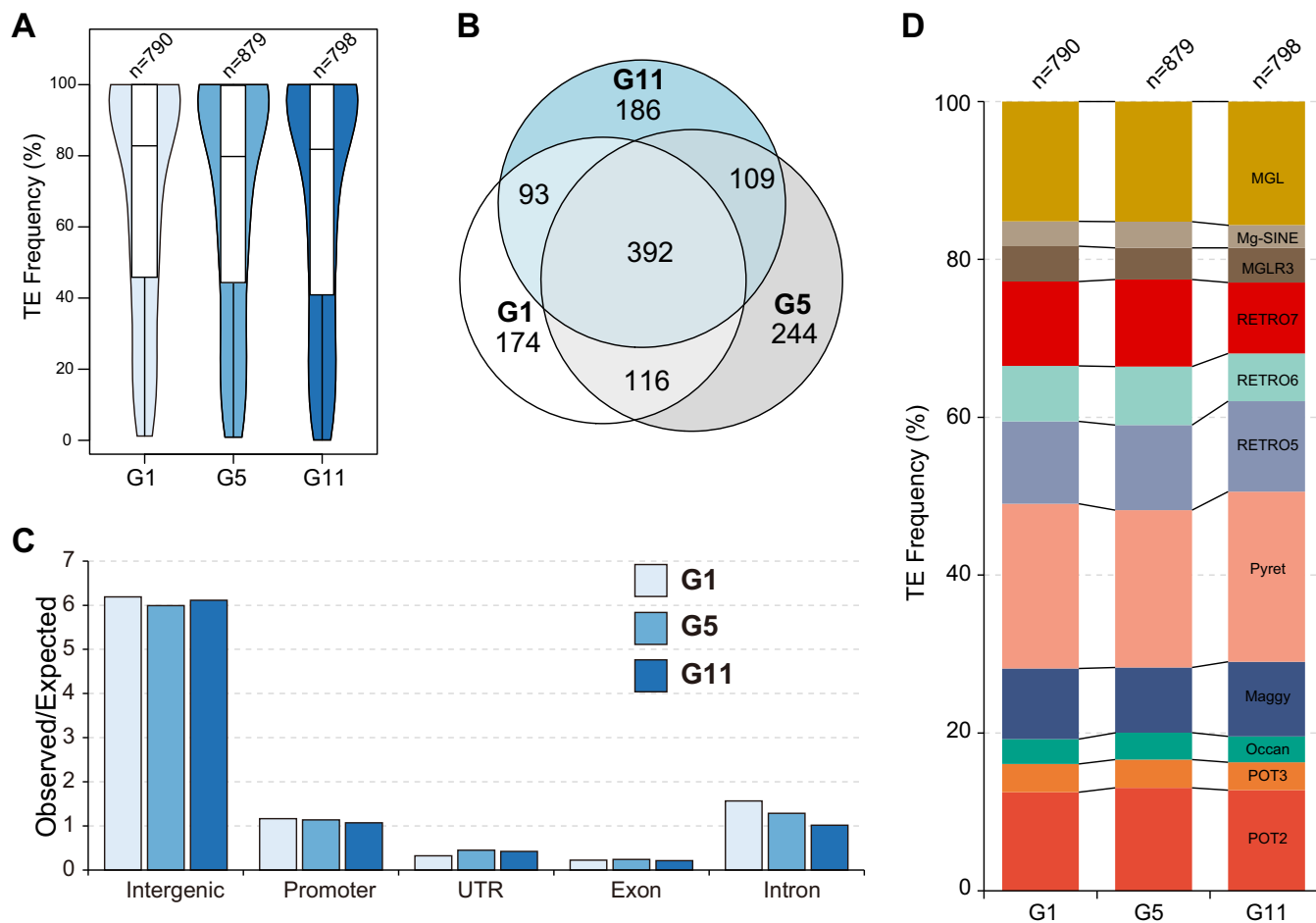
Effect type	No. of effects in:		
	G <sub>1</sub>	G <sub>5</sub>	G <sub>11</sub>
Splice acceptor variant	3	2	4
Stop gained	9	3	4
Frameshift variant	5	10	9
Start lost	0	0	1
In-frame insertion	1	1	1
In-frame deletion	2	1	3
Missense variant	83	101	116
Splice region variant	6	7	8
Synonymous variant	40	50	56
5' UTR variant	52	68	90
3' UTR variant	43	66	85
Noncoding transcript exon variant	3	4	3
Intron variant	46	83	97
Upstream gene variant	258	351	424
Downstream gene variant	28	36	51

and filtering them with G<sub>0</sub>. We identified 790, 879, and 798 newly formed TE insertion events in G<sub>1</sub>, G<sub>5</sub>, and G<sub>11</sub>, respectively (Fig. 3A). A Venn diagram showing the overlapping TEs in the three samples indicated that more than 50% of TEs were shared among these samples and that 186 TEs were uniquely present in G<sub>11</sub> (Fig. 3B). The insertion of TEs at the promoter region results in gene silencing; thus, we investigated the relationship between the TE insertion site and the coding gene. We found that, as observed in the SNV/indel analysis, new TEs were preferentially enriched in intergenic regions and depleted in exon regions (Fig. 3C). To determine the activity of these TEs, we calculated the copy number of newly formed TEs in G<sub>1</sub> ( $n = 790$ ), G<sub>5</sub> ( $n = 879$ ), and G<sub>11</sub> ( $n = 798$ ). In general, long-terminal-repeat (LTR) retrotransposons (Maggy, RETRO5, RETRO6, RETRO7, Pyret, and MGLR3) are more active than DNA transposons (POT2, POT3, and Occan). We found that among the LTR retrotransposons, Pyret was the most active TE, followed by long interspersed elements and MGL (Fig. 3D). Compared with POT3 and Occan, POT2 was the most active DNA transposon. In summary, extensive variation in TE copy number was observed in the process of successive host interaction, and TEs were preferentially enriched in intergenic regions and depleted in exon regions.

**TE insertion in genes encoding SPs.** Secreted proteins (SPs) play dual functions in plant-pathogen interactions (27, 28). They can facilitate pathogen infection by overcoming the plant defense response, such as suppressing the plant immune system or hijacking host metabolism (29). Some SP genes (*Avr* genes) can be recognized by plant *R* gene products to induce plant immunity. Previous studies indicated that this group of genes is a hot spot for SNV variation and TE insertion because these variations can help pathogens avoid plant recognition (17, 25, 26). However, so far, there has been no direct evidence to support the preferential insertion of TEs in SP genes. Therefore, we investigated the percentage of SP genes among genes that have newly formed TE insertions in the promoter, exon, intron, or transcription termination site (TTS) regions. The abundance of SP genes in the whole genome was 11.84% (1,539/12,991), which was similar to that of SNV/indel-associated genes (10.69%), the control set of SNV/indel-associated genes (11.15%), and the control set of TE-associated genes (10.33%); however, the abundance of SP genes in the whole genome was significantly ( $P < 0.05$ ) lower than that of TE-associated SP genes (16.32%), especially at promoter and intron regions (Fig. 4A; Fig. S1A and B). This result provides direct evidence to support the hypothesis that TEs tend to insert into SP genes.

## DISCUSSION

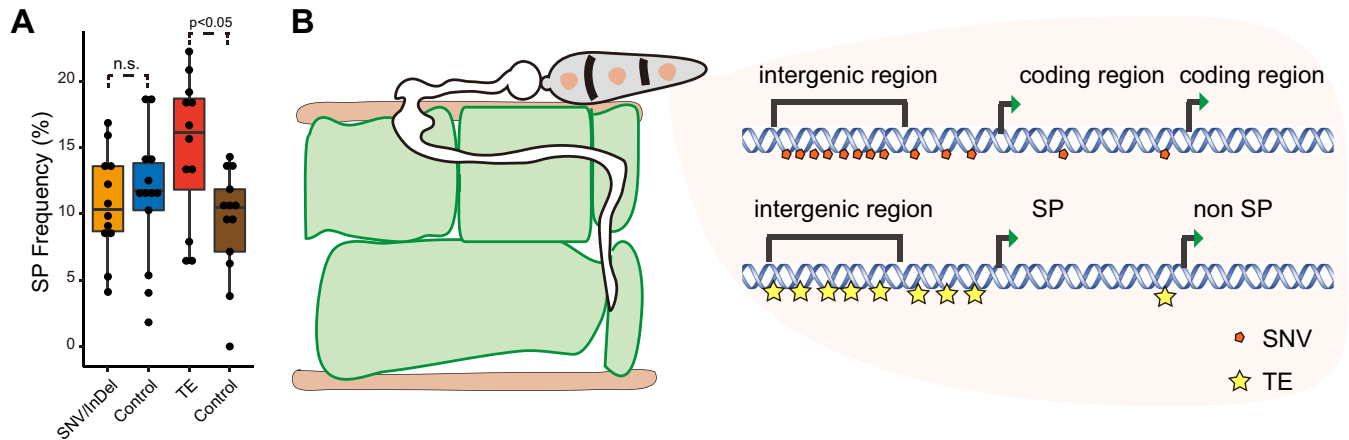
Genome plasticity allows the rapid evolution of plant pathogens and enables plant pathogens to overcome host resistance quickly, which makes the control of fungal



**FIG 3** Annotation of TEs identified in the experimental evolution assay. (A) Copy number and frequency of SNVs/indels identified in G<sub>1</sub>, G<sub>5</sub>, and G<sub>11</sub>. The *P* value was calculated using Student's *t* test. (B) Venn diagram showing overlapping TEs identified in G<sub>1</sub>, G<sub>5</sub>, and G<sub>11</sub>. (C) Genomic distribution of TEs identified in G<sub>1</sub>, G<sub>5</sub>, and G<sub>11</sub>. The number was calculated by comparing the expected number of TEs in each of five categories (intergenic, promoter, exon, intron, and UTR) with the observed number of sites in the genome. (D) Proportion of different types of TEs among newly formed TEs in G<sub>1</sub>, G<sub>5</sub>, and G<sub>11</sub>.

diseases challenging. The filamentous fungus *M. oryzae* is the causal pathogen of rice blast disease, which leads to substantial losses in rice production annually (15, 30). Our previous studies suggested that the host plant is the major force that shapes the *M. oryzae* genome (20, 21, 31). However, no real-time assays have been performed so far to investigate the occurrence and accumulation of mutations in *M. oryzae* populations in interaction with rice plants. To this end, we performed the *in planta* experimental evolution of *M. oryzae* by monitoring the genomic variation in serially subinoculated populations.

During adaptation to an alternative host, pathogens are thought to maintain a balance between generalism and specialization and finally to reach a suboptimal fitness (32, 33). In this study, by monitoring the genomic variation in a population of progeny derived from sequentially inoculating *M. oryzae*, we found that G<sub>1</sub> presented a vast number of variations and that a large number of variations were shared by all samples. These results suggested that genomic variations occurred rapidly within a very short period of time and accumulated very rapidly in the genome and that G<sub>1</sub> has a founder effect on the sequential inoculation population. Because TP309 is not the original host of the Guy11 strain, it may possess immunity to Guy11, in contrast to its original host plant. We therefore proposed that the fungal genome may have been subjected to strong selection in response to plant immunity at the beginning of the experiment and reached a bottleneck when the strain showed optimal fitness under continuous selection in this plant variety. Similar results were shown by Jeon et al. (19). However, the



**FIG 4** Bias in transposable element (TE) insertion in secreted protein (SP) coding genes. (A) Percentage of SP genes among genes with newly formed TE insertions. Dots in the box plot include promoter, exon, intron, or TTS region data. The  $P$  value was calculated using Student's  $t$  test. Detailed data for each dot are shown in Fig. S1 in the supplemental material. n.s., nonsignificant. (B) Proposed model depicting the changes in an *in planta* experimental evolution assay of *M. oryzae*. Single-nucleotide variants (SNVs) and TE insertions accumulated rapidly throughout the genome of *M. oryzae* in noncoding regions. TE insertions preferentially accumulated in the promoter regions of SP coding genes in *M. oryzae*.

host plant may impose a much more specific and efficient selection on the fungal genome than does oatmeal medium. In the future, using strains to inoculate their original host plants might further illustrate whether such bottleneck effects are common during adaptive evolution. It is also worth noting that although the medium cultivation process has been minimized and mimics the natural rice environments by using rice bran medium to exclude most of the selection stress on the isolates in this study, the medium cultivation process may still result in some stress on the population that is different from host selection, resulting in some abiotic stress variations in the *M. oryzae* genome. Development of advanced sequencing technologies, such as single-cell sequencing, may provide resolution to waive the artificial effects caused by medium culture. Taking these results together, we propose that *M. oryzae* maintained a balance between population diversity and bottleneck effects while adapting to its host plant.

Notably, genetic variation does not neutrally accumulate in the genome during adaptive evolution. We observed that increased mutations accumulated in Chr 1, Chr 3, Chr 6, and Chr 7, and this result is consistent with that of a previous experimental evolution study (19). It is possible that there are mutation hot spots in these chromosomes, or alternatively, these chromosomes might buffer more mutations and undergo stronger selection. Interestingly, more SNV/indel mutations were found in noncoding regions than in coding regions, as shown previously by Jeon et al. (19). The different mutation rates in the noncoding (intergenic) and coding (exon) regions can at least partially explain the accumulation of variations in *M. oryzae* in a short period of time despite *M. oryzae* rarely showing phenotypic variations. We proposed that this bias indicates a purifying selection on coding regions, especially some important house-keeping genes, to maintain genome stability. To search for genes that are involved in the adaptive process, we annotated the genomic consequences of these variations and found that these genes were involved in multiple pathways.

Previous studies have indicated that transposable elements (TEs) play important roles in host-pathogen interactions. TEs are major components of facultative heterochromatin regions that provide not only epigenetic regulation of the transcription of effector/secreted protein (SP) genes but also a cradle for rapid adaptive evolution (6, 34) and can induce total silencing of *Avr* genes and thus help pathogens avoid host recognition (25, 26, 35). Our results showed that  $G_1$  also has a founder effect on newly formed TE insertion, although the number of new TE insertions gradually increased in  $G_5$  and  $G_{11}$ . These results suggested that the transposability of TEs in the  $G_1$  fungal genome is the highest. In line with the bias mutation rate of SNVs/indels in the



intergenic and exon regions, TE insertions were also enriched in the intergenic region and depleted in the exon region. In addition, we found that Pyret was the most active TE. This result is supported by a previous study, which also found that Pyret was the most active TE under different stresses (36).

The pathogen effector/SP genes play pivotal roles during the interaction with their host, especially for biotrophic or hemibiotrophic pathogens (27, 28). These genes are enriched preferentially in repeat-rich regions and are thought to be evolved more rapidly than genes in other genomic compartments (17, 25, 26). Moreover, most of these genes are functionally redundant (37) and could tolerate more mutations, which provides genetic diversity for adaptive evolution on diverse hosts. In *M. oryzae*, an increasing amount of evidence has demonstrated that SP genes, especially the *Avr* genes, tend to evolve more rapidly than other genes. For instance, comparative genomic analysis revealed that the SP genes presented a high level of diversity among different strains (38–40). In addition, some of the SP genes (*Avr* genes) have been suggested to be hot spots for SNV variation and TE insertion as a tool for avoiding plant recognition (17, 25, 26, 41). Consistent with this hypothesis, we found that in addition to the intergenic region, TEs also showed an insertion bias toward genes encoding SPs. These results suggested that SP genes are very important for adaptive evolution toward diverse rice hosts (different subspecies/cultivars) and that TEs are one of the major forces fueling the rapid evolution of *M. oryzae*. The high bias toward TE insertion into SP genes indicated that TE insertion at the *Avr* gene may not be neutral.

In conclusion, we applied an experimental evolution assay to *M. oryzae* and identified the preferential accumulation of SNVs/indels and TEs in noncoding regions and SP genes (Fig. 4B). Further studies on the genetic function of the mutated genes identified in this study will provide more insights into the adaptive evolution of *M. oryzae*. We believe that the results obtained through experimental evolution may also enhance our understanding of pathogen evolution in nature and explain the mechanism of the vulnerability of rice resistance.

## MATERIALS AND METHODS

**Evolution experiment assay.** *Magnaporthe oryzae* strain Guy11, isolated from French Guyana (42), was used as the initial strain in this study. Single spores isolated from diseased rice (TP309, a cultivar susceptible to Guy11) leaves 7 days postinoculation (dpi) of each generation were grown on rice bran medium (2% rice polish and 1.5% agar at pH 6.5) at 26°C under constant light for conidiation. Small pieces of sterilized filter paper were also put on the medium and collected when it was colonized by the fungus. The fungal cultures on the dried filter paper were then stored in a –80°C refrigerator. Genomic DNA samples were extracted using the CTAB extraction method from mycelia cultured in liquid CM medium (0.6% yeast extract, 0.6% casein hydrolysate, 1% sucrose, and 1.5% agar) with 130 rpm shaking at 26°C for 3 to 4 days. Detailed steps for the experimental evolution assay are described in Results and in the legend for Fig. 1A.

**Genome sequencing.** DNA samples from individual isolates were combined in equal amounts for sequencing. DNA samples were sheared to ~350 bp in average size. Sequencing libraries were prepared using the Illumina paired-end DNA sample prep kit and sequenced on an Illumina HiSeq 2500 platform with 2 × 150-bp paired-end reads.

**Read alignment and SNV/indel calling and annotation.** SNV/indel calling was performed according to a previously described method with some modifications (43, 44). Briefly, all sequenced reads were aligned to the *M. oryzae* 70-15 reference genome with Bowtie2 with the default parameters (45). The resulting bam files were subjected to Picard MarkDuplicates function to remove PCR duplicates (<https://broadinstitute.github.io/picard>). PCR-duplicated reads were removed, and the remaining files were then subjected to variant calling. Additionally, mapping reads with a mapping quality (MAPQ) value of >40 were retained for variant calling. CRISP has been used for variant calling with a minimum of 10 reads with alternate alleles (46). The SNV/indel density was calculated by VCFtools (v0.1.15) in every 10-kb window and visualized with circlize (47, 48). The SNV/indel control data set was randomly selected from the whole genome with BEDTools (v2.21.0) (49). The genomic distribution of the SNVs/indels was annotated with ChIPseeker (v1.20.0) (50). The types of consequences of the SNVs/indels were predicted by VEP (v96) and the *Magnaporthe* 70-15 reference genome version 43 from EnsembleFungi ([https://fungi.ensembl.org/Magnaporthe\\_oryzae/Info/Index](https://fungi.ensembl.org/Magnaporthe_oryzae/Info/Index)) (51).

**TE annotation and SP prediction.** RepeatMasker (version 3.3.0; <http://www.repeatmasker.org/>) was used to search for TEs in 70-15 (52). The presence and absence of TE polymorphisms were detected by PoPoolationTE2 with the 70-15 sequence as the reference genome and at least 3 reads of support (53). TE insertions supported by at least 3 reads in  $G_0$  were maintained for TE filtering in the  $G_1$ ,  $G_5$ , and  $G_{11}$  samples. The genomic distribution of TE insertions was annotated with the annotatePeaks function in Homer2 (v4.8.3) (54). Peaks within 50 bp between two samples were merged into the same insertion by

BEDTools (v2.21.0) (49). The TE control data set was randomly selected from the whole genome with BEDTools (v2.21.0) (49). SPs were defined as proteins containing a signal peptide cleavage site, no transmembrane domain after the signal peptide cleavage site, and an amino acid sequence length smaller than 400 amino acids. SignalP 5.0 was used to predict signal peptides, and TMHMM 2.0 was used to predict transmembrane domains (55, 56).

**Statistical analysis.** All statistical analyses were performed using Student's *t* test function in R (57).

**Data availability.** All genomic sequencing data were deposited in the NCBI Sequence BioProject database under accession number [PRJNA577277](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA577277).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, PDF file, 0.4 MB.

**TABLE S1**, XLSX file, 0.01 MB.

**TABLE S2**, XLSX file, 0.02 MB.

## ACKNOWLEDGMENTS

We thank the members of the Z.W. laboratory for useful discussions.

This work was supported by grants from the Natural Science Foundations of China (U1805232, 31770156) and the National Key R&D Program of China (2016YFD0300700) to Z.W.

Z.W. and Z.Z. conceived the work, designed the experiments, and wrote the manuscript. H.Z. and J.N. contributed to manuscript writing. M.C. and D.L. provided experiment assistance. L.L. and R.C. provided support for bioinformatics analyses. All authors read and approved the manuscript.

## REFERENCES

- Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GC, Wittenberg AH, Thomma BP. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res* 26:1091–1100. <https://doi.org/10.1101/gr.204974.116>.
- Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol* 10:417–430. <https://doi.org/10.1038/nrmicro2790>.
- Bent AF, Mackey D. 2007. Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annu Rev Phytopathol* 45:399–436. <https://doi.org/10.1146/annurev.phyto.45.062806.094427>.
- Giraldo MC, Valent B. 2013. Filamentous plant pathogen effectors in action. *Nat Rev Microbiol* 11:800–814. <https://doi.org/10.1038/nrmicro3119>.
- Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog* 8:e1002608. <https://doi.org/10.1371/journal.ppat.1002608>.
- Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev* 35:57–65. <https://doi.org/10.1016/j.gde.2015.09.001>.
- McDonald BA, McDermott JM. 1993. Population genetics of plant pathogenic fungi. *Bioscience* 43:311–319. <https://doi.org/10.2307/1312063>.
- McDonald BA, Linde C. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu Rev Phytopathol* 40:349–379. <https://doi.org/10.1146/annurev.phyto.40.120501.101443>.
- Stukenbrock EH, Bataillon T. 2012. A population genomics perspective on the emergence and adaptation of new plant pathogens in agroecosystems. *PLoS Pathog* 8:e1002893. <https://doi.org/10.1371/journal.ppat.1002893>.
- Guan SH, Gris C, Cruveiller S, Pouzet C, Tasse L, Leru A, Maillard A, Médigue C, Batut J, Masson-Boivin C, Capela D. 2013. Experimental evolution of nodule intracellular infection in legume symbionts. *ISME J* 7:1367–1377. <https://doi.org/10.1038/ismej.2013.24>.
- Guidot A, Jiang W, Ferdy J-B, Thébaud C, Barberis P, Gouzy J, Genin S. 2014. Multihost experimental evolution of the pathogen *Ralstonia solanacearum* unveils genes involved in adaptation to plants. *Mol Biol Evol* 31:2913–2928. <https://doi.org/10.1093/molbev/msu229>.
- Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489:513–518. <https://doi.org/10.1038/nature11514>.
- Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461:1243–1247. <https://doi.org/10.1038/nature08480>.
- Marchetti M, Capela D, Glew M, Cruveiller S, Chane-Woon-Ming B, Gris C, Timmers T, Poinot V, Gilbert LB, Heeb P, Médigue C, Batut J, Masson-Boivin C. 2010. Experimental evolution of a plant pathogen into a legume symbiont. *PLoS Biol* 8:e1000280. <https://doi.org/10.1371/journal.pbio.1000280>.
- Dean R, Van Kan JA, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, Foster GD. 2012. The top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol* 13:414–430. <https://doi.org/10.1111/j.1364-3703.2011.00783.x>.
- Ou SH. 1985. Rice diseases, 2nd ed. IRRI, Laguna, Philippines.
- Wu J, Kou Y, Bao J, Li Y, Tang M, Zhu X, Ponaya A, Xiao G, Li C, Song MY, Cumagun CJ, Deng Q, Lu G, Jeon JS, Naqvi NI, Zhou B. 2015. Comparative genomics identifies the Magnaporthe oryzae avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice. *New Phytol* 206:1463–1475. <https://doi.org/10.1111/nph.13310>.
- Kim B, Ahn S, Roh J, Choi S, Han S. 2004. Durability of rice cultivars to blast in Korea by sequential planting method. *Korean J Breed* 36:350–356.
- Jeon J, Choi J, Lee GW, Dean RA, Lee YH. 2013. Experimental evolution reveals genome-wide spectrum and dynamics of mutations in the rice blast fungus, *Magnaporthe oryzae*. *PLoS One* 8:e65416. <https://doi.org/10.1371/journal.pone.0065416>.
- Zhong Z, Norvienyeku J, Chen M, Bao J, Lin L, Chen L, Lin Y, Wu X, Cai Z, Zhang Q, Lin X, Hong Y, Huang J, Xu L, Zhang H, Chen L, Tang W, Zheng H, Chen X, Wang Y, Lian B, Zhang L, Tang H, Lu G, Ebbolle DJ, Wang B, Wang Z. 2016. Directional selection from host plants is a major force driving host specificity in *Magnaporthe* species. *Sci Rep* 6:25591. <https://doi.org/10.1038/srep25591>.
- Zhong Z, Chen M, Lin L, Han Y, Bao J, Tang W, Lin L, Lin Y, Somai R, Lu L, Zhang W, Chen J, Hong Y, Chen X, Wang B, Shen WC, Lu G, Norvienyeku J, Ebbolle DJ, Wang Z. 2018. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. *ISME J* 12:1867–1878. <https://doi.org/10.1038/s41396-018-0100-6>.
- Futschik A, Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186:207–218. <https://doi.org/10.1534/genetics.110.114397>.
- Bao J, Chen M, Zhong Z, Tang W, Lin L, Zhang X, Jiang H, Zhang D, Miao C, Tang H, Zhang J, Lu G, Ming R, Norvienyeku J, Wang B, Wang Z. 2017. PacBio sequencing reveals transposable elements as a key contributor to

- genomic plasticity and virulence variation in *Magnaporthe oryzae*. *Mol Plant* 10:1465–1468. <https://doi.org/10.1016/j.molp.2017.08.008>.
24. Zhong Z, Lin L, Zheng H, Bao J, Chen M, Zhang L, Tang W, Ebbolle DJ, Wang Z. 2020. Emergence of a hybrid PKS-NRPS secondary metabolite cluster in a clonal population of the rice blast fungus *Magnaporthe oryzae*. *Environ Microbiol* <https://doi.org/10.1111/1462-2920.14994>.
  25. Li W, Wang B, Wu J, Lu G, Hu Y, Zhang X, Zhang Z, Zhao Q, Feng Q, Zhang H, Wang Z, Wang G, Han B, Wang Z, Zhou B. 2009. The *Magnaporthe oryzae* avirulence gene *AvrPiz-t* encodes a predicted secreted protein that triggers the immunity in rice mediated by the blast resistance gene *Piz-t*. *Mol Plant Microbe Interact* 22:411–420. <https://doi.org/10.1094/MPMI-22-4-0411>.
  26. Zhang S, Wang L, Wu W, He L, Yang X, Pan Q. 2015. Function and evolution of *Magnaporthe oryzae* avirulence gene *AvrPib* responding to the rice blast resistance gene *Pib*. *Sci Rep* 5:11642. <https://doi.org/10.1038/srep11642>.
  27. Liu W, Liu J, Ning Y, Ding B, Wang X, Wang Z, Wang GL. 2013. Recent progress in understanding PAMP- and effector-triggered immunity against the rice blast fungus *Magnaporthe oryzae*. *Mol Plant* 6:605–620. <https://doi.org/10.1093/mp/sst015>.
  28. Liu W, Liu J, Triplett L, Leach JE, Wang G-L. 2014. Novel insights into rice innate immunity against bacterial and fungal pathogens. *Annu Rev Phytopathol* 52:213–241. <https://doi.org/10.1146/annurev-phyto-102313-045926>.
  29. Rodriguez-Moreno L, Ebert MK, Bolton MD, Thomma B. 2018. Tools of the crook-infection strategies of fungal plant pathogens. *Plant J* 93:664–674. <https://doi.org/10.1111/tpj.13810>.
  30. Ebbolle DJ. 2007. *Magnaporthe* as a model for understanding host-pathogen interactions. *Annu Rev Phytopathol* 45:437–456. <https://doi.org/10.1146/annurev.phyto.45.062806.094346>.
  31. Zheng H, Zhong Z, Shi M, Zhang L, Lin L, Hong Y, Fang T, Zhu Y, Guo J, Zhang L, Fang J, Lin H, Norvienyeku J, Chen X, Lu G, Hu H, Wang Z. 2018. Comparative genomic analysis revealed rapid differentiation in the pathogenicity-related gene repertoires between *Pyricularia oryzae* and *Pyricularia penniseti* isolated from a Pennisetum grass. *BMC Genomics* 19:927. <https://doi.org/10.1186/s12864-018-5222-8>.
  32. Benmayer R, Hodgson DJ, Perron GG, Buckling A. 2009. Host mixing and disease emergence. *Curr Biol* 19:764–767. <https://doi.org/10.1016/j.cub.2009.03.023>.
  33. van Baalen M, Sabelis MW. 1995. The dynamics of multiple infection and the evolution of virulence. *Am Nat* 146:881–910. <https://doi.org/10.1086/285830>.
  34. Jeon J, Choi J, Lee G-W, Park S-Y, Huh A, Dean RA, Lee Y-H. 2015. Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae*. *Sci Rep* 5:8567. <https://doi.org/10.1038/srep08567>.
  35. Chen C, Chen M, Hu J, Zhang W, Zhong Z, Jia Y, Allaux L, Fournier E, Tharreau D, Wang G-L, Wang A, Shen W-C, Lu G, Wang B, Mitchell TK. 2014. Sequence variation and recognition specificity of the avirulence gene *AvrPiz-t* in *Magnaporthe oryzae* field populations. *Fungal Genom Biol* 4:113.
  36. Chadha S, Sharma M. 2014. Transposable elements as stress adaptive capacitors induce genomic instability in fungal pathogen *Magnaporthe oryzae*. *PLoS One* 9:e94415. <https://doi.org/10.1371/journal.pone.0094415>.
  37. Saitoh H, Fujisawa S, Mitsuoka C, Ito A, Hirabuchi A, Ikeda K, Irieda H, Yoshino K, Yoshida K, Matsumura H, Tosa Y, Win J, Kamoun S, Takano Y, Terauchi R. 2012. Large-scale gene disruption in *Magnaporthe oryzae* identifies MC69, a secreted protein required for infection by monocot and dicot fungal pathogens. *PLoS Pathog* 8:e1002711. <https://doi.org/10.1371/journal.ppat.1002711>.
  38. Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, Tosa Y, Chuma I, Takano Y, Win J, Kamoun S, Terauchi R. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* 21:1573–1591. <https://doi.org/10.1105/tpc.109.066324>.
  39. Xue M, Yang J, Li Z, Hu S, Yao N, Dean RA, Zhao W, Shen M, Zhang H, Li C, Liu L, Cao L, Xu X, Xing Y, Hsiang T, Zhang Z, Xu JR, Peng YL. 2012. Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. *PLoS Genet* 8:e1002869. <https://doi.org/10.1371/journal.pgen.1002869>.
  40. Chen C, Lian B, Hu J, Zhai H, Wang X, Venu RC, Liu E, Wang Z, Chen M, Wang B, Wang GL, Wang Z, Mitchell TK. 2013. Genome comparison of two *Magnaporthe oryzae* field isolates reveals genome variations and potential virulence effectors. *BMC Genomics* 14:887. <https://doi.org/10.1186/1471-2164-14-887>.
  41. Huang J, Si W, Deng Q, Li P, Yang S. 2014. Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. *BMC Genet* 15:45. <https://doi.org/10.1186/1471-2156-15-45>.
  42. Leung H, Borromeo ES, Bernardo MA, Nottoghem JL. 1988. Genetic analysis of virulence in the rice blast fungus *Magnaporthe grisea*. *Phytopathology* 78:1227–1233. <https://doi.org/10.1094/Phyto-78-1227>.
  43. Anand S, Mangano E, Barizzone N, Bordoni R, Sorosina M, Clarelli F, Corrado L, Martinelli Boneschi F, D'Alfonso S, De Bellis G. 2016. Next generation sequencing of pooled samples: guideline for variants' filtering. *Sci Rep* 6:33735. <https://doi.org/10.1038/srep33735>.
  44. Cutler DJ, Jensen JD. 2010. To pool, or not to pool? *Genetics* 186:41–43. <https://doi.org/10.1534/genetics.110.121012>.
  45. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
  46. Bansal V. 2010. A statistical method for the detection of variants from next-generation resequencing of DNA pools. *Bioinformatics* 26:i318–i324. <https://doi.org/10.1093/bioinformatics/btq214>.
  47. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Grp G, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
  48. Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. circlize implements and enhances circular visualization in R. *Bioinformatics* 30:2811–2812. <https://doi.org/10.1093/bioinformatics/btu393>.
  49. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
  50. Yu G, Wang L-G, He Q-Y. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31:2382–2383. <https://doi.org/10.1093/bioinformatics/btv145>.
  51. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The ensembl variant effect predictor. *Genome Biol* 17:122. <https://doi.org/10.1186/s13059-016-0974-4>.
  52. Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protoc Bioinformatics* 25:Unit 4.10. <https://doi.org/10.1002/0471250953.bi0410s25>.
  53. Kofler R, Gómez-Sánchez D, Schlötterer C. 2016. PoPoolationTE2: comparative population genomics of transposable elements using Pool-Seq. *Mol Biol Evol* 33:2759–2764. <https://doi.org/10.1093/molbev/msw137>.
  54. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.
  55. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. <https://doi.org/10.1038/s41587-019-0036-z>.
  56. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>.
  57. R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.