





Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades

Zhenhui Zhong^{1,2} · Meilian Chen^{1,2} · Lianyu Lin^{1,2} · Yijuan Han^{1,2} · Jiandong Bao^{1,2}  · Wei Tang^{1,2} · Lili Lin^{1,2} · Yahong Lin^{1,2} · Rewish Somai^{1,2} · Lin Lu^{1,2} · Wenjing Zhang^{1,2} · Jian Chen^{1,2} · Yonghe Hong^{1,2} · Xiaofeng Chen^{1,2} · Baohua Wang^{1,2} · Wei-Chiang Shen³ · Guodong Lu^{1,2} · Justice Norvienyeku^{1,2} · Daniel J. Ebbole^{1,4}  · Zonghua Wang^{1,2,5}

Received: 1 September 2017 / Revised: 10 February 2018 / Accepted: 20 February 2018 / Published online: 22 March 2018
© The Author(s) 2018. This article is published with open access

Abstract

We examined the genomes of 100 isolates of *Magnaporthe oryzae* (*Pyricularia oryzae*), the causal agent of rice blast disease. We grouped current field populations of *M. oryzae* into three major globally distributed groups. A genetically diverse group, clade 1, which may represent a group of closely related lineages, contains isolates of both mating types. Two well-separated clades, clades 2 and 3, appear to have arisen as clonal lineages distinct from the genetically diverse clade. Examination of genes involved in mating pathways identified clade-specific diversification of several genes with orthologs involved in mating behavior in other fungi. All isolates within each clonal lineage are of the same mating type. Clade 2 is distinguished by a unique deletion allele of a gene encoding a small cysteine-rich protein that we determined to be a virulence factor. Clade 3 isolates have a small deletion within the *MFA2* pheromone precursor gene, and this allele is shared with an unusual group of isolates we placed within clade 1 that contain *AVRI-CO39* alleles. These markers could be used for rapid screening of isolates and suggest specific events in evolution that shaped these populations. Our findings are consistent with the view that *M. oryzae* populations in Asia generate diversity through recombination and may have served as the source of the clades 2 and 3 isolates that comprise a large fraction of the global population.

These authors contributed equally: Zhenhui Zhong, Meilian Chen

Data access: Sequencing raw reads and assembled contigs are available at NCBI under BioProject ID: PRJNA354675; BioSample: SAMN06050060–SAMN06050147; SRA: SRP132141.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41396-018-0100-6>) contains supplementary material, which is available to authorized users.

- ✉ Wei-Chiang Shen
wshen@ntu.edu.tw
- ✉ Justice Norvienyeku
jk_norvienyeku@fafu.edu.cn
- ✉ Daniel J. Ebbole
d-ebbole@tamu.edu
- ✉ Zonghua Wang
wangzh@fafu.edu.cn

¹ State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, College of Plant Protection, Fujian Agriculture and

Introduction

Magnaporthe oryzae (*Pyricularia oryzae*), the causal agent of rice blast disease, results in significant yield losses worldwide and is an important threat to food security [1, 2]. Although deployment of resistance genes is an important strategy for disease control, the genetic diversity of the pathogen has limited the effectiveness of this approach. Understanding the structure of the current population and

Forestry University, Fuzhou 350002, China

² Fujian-Taiwan Joint Center for Ecological Control of Crop Pests, Fujian Agriculture and Forestry University, Fuzhou 350002, China

³ Department of Plant Pathology and Microbiology, National Taiwan University, Taipei, Republic of China

⁴ Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX, USA

⁵ Institute of Ocean Science, Minjiang University, Fuzhou 350108, China

how diversity is generated is important for managing the disease.

Like other heterothallic ascomycetes, *M. oryzae* has two mating types determined by a single master locus carrying either Mat1-1 or Mat1-2 sequences [3, 4]. Although sexual reproduction in the field has not been observed directly, population studies indicate that sexual recombination occurs in geographical regions associated with the origin of rice domestication [5, 6]. However, there are also examples where isolates of both mating types are present but recombination appears to be absent or very rare [7–9]. In addition, *Magnaporthe* may experience parasexual recombination in the field [10–14]. Transposable elements also contribute to generating diversity through transposition to cause gene deletion or otherwise inactive genes [15, 16]. Relationships between populations adapted to different hosts have been defined using multiple gene loci and recently with population genomic analysis [16–20]. These studies are critical for understanding current population structures and the potential for invasion of new hosts, such as wheat [21].

Population studies conducted on *M. oryzae* rice isolates found that local *M. oryzae* populations are often dominated by isolates of one mating type [7, 22–26], suggesting that rice blast populations are predominately clonal lineages that are selected based on the host genotype, such that collections made from any one location may show restricted genetic diversity.

In studies using microsatellite markers [6, 27], three main genetic clusters were found in *M. oryzae* population. One genetically diverse population undergoes sexual recombination and could be separated into two subgroups, each containing both mating types. The two other genetic clusters were mainly clonal and these groups were distributed globally, but not uniformly so. A recent study also identified recombining and clonal populations in six lineages, and estimated the time of divergence of the lineages from each other approximately 1000 years before present (YBP) [28].

Understanding the complexity of these populations will require additional genome and population analyses to define the relationships within and between clades with more resolution and to help identify, from among the evolutionary events that have accompanied the expansion of particular lineages, which genes were targeted by natural selection. Here we describe high-quality sequencing, de novo assembly, and annotation for globally distributed *M. oryzae* isolates to assess mechanisms that may help explain the observed population structure. We investigated the population structure of isolates from China as well as several different regions around the world and assigned isolates into three major clades. Clade 1 contained isolates of both mating types, clade 2 contained only Mat1-2 isolates, and clade 3 contained only Mat1-1 isolates. We were unable to

detect gene flow between the two single-mating-type clades. Clade 2 contained a deletion of a gene encoding an intracellular effector that was able to suppress BAX-mediated cell death in *Nicotiana benthamiana*. The occurrence of genetically distinct clonal lineages indicates that hybrids between clades have reduced fitness and/or that mating between clades is restricted. To search for variation in the genes of the mating pathway that may contribute to poor fertility, we surveyed allelic variation in genes known to be involved in fungal mating pathways. This analysis revealed a fixed non-synonymous substitution allele of *MST20* in clade 3. Clade 3 isolates also possess a 30 bp deletion within the *MFA2* pheromone precursor gene. These unique alleles can serve as markers to assign isolates to clades. We estimate the clades diverged with each other approximately 1000 YBP. Additional genomic diversity within the clades revealed that rapid genetic divergence is occurring between the two clades undergoing asexual reproduction, which appear to have undergone expansion ~250–400 YBP.

Materials and Methods

Isolate collection, cultivation, DNA preparation, genome sequencing, and annotation

Isolates in this study are described in supplemental Table S1 [18, 19, 29–34]. To maximize genetic diversity, the isolates were selected based on distinct geographic location and year of collection. Genomic DNA was isolated as previously described [19]. Sequencing libraries were prepared using the Illumina Paired-End DNA sample Prep Kit and sequenced using an Illumina HiSeq2500. De novo sequence assembly was conducted with CLC Genomic Workbench 7.0. Gene predictions were conducted through a combination of evidence-based prediction by Exonerate (version 2.2.0) [35], and ab initio gene prediction with Fgenesh from SoftBerry (<http://linux1.softberry.com/berry.phtml>). Evidence-based and ab initio gene prediction results were combined by EVIDENCEModeler [36].

Single-nucleotide polymorphism discovery, identification of mating-related genes, and transposon element annotation

Sequenced reads were aligned to the reference genome (70-15 v8) with Bowtie2 [37, 38]. SAMtools (version 0.1.19) and Genome Analysis Toolkit (GATK, version 3.3-0), for single-nucleotide polymorphism (SNP) calling with allele/depth <0.9, and filtered for multiple alleles with other parameters set to the default [39–41]. The genome-to-genome SNP calling was performed with MUMmer, version 3.23 (-maxmatch -c 100 -p) [42]. Amino-acid

sequences of yeast mating-related genes were obtained from the *Saccharomyces* Genome Database (<http://www.yeastgenome.org/>). Yeast homologs were identified by reciprocal BLASTP ($E < 10^{-5}$) with the *M. oryzae* genome [3, 43].

Phylogenomic tree construction, population structure, and divergence time estimation

The phylogenomic tree was built using FastTree (Version 2.1.9) based on the SNP dataset with the approximately-maximum-likelihood model [44]. STRUCTURE analysis was conducted using Structure 2.3.4 with whole-genome SNPs, and K value selected by the Evanno method in STRUCTURE HARVESTER [45, 46]. Principal component analysis (PCA) based on SNPs was conducted by using Tassel 5.0 and plotted with R [47]. BEAST (v1.8.3) was used to estimate divergence time with the timescale calibrated using the collection year of samples (1978–2015) and two *Setaria viridis* isolates as outgroup [48]. We used the HKY + Gamma substitution model and selected the strict clock model based on prior test of both strict and relaxed clock results analyzed with Tracer (v1.6), and Markov chain Monte Carlo analyses of 10 million iterations were run in duplicate to check for convergence. The timescaled tree was displayed with FigTree (v1.4.3, <http://tree.bio.ed.ac.uk/software/figtree/>). Additional public sequence data (GenBank PRJNA245782) were assessed for concordance with our analysis of the main group of genomes.

Population genetics and genomics

DNA_{sp} was used for nucleotide diversity calculations [49]. KaKs_Calculator 2.0 was used to calculate Ka/Ks values with the YN model [50, 51]. Consensus genome sequences obtained with the SNP dataset were used for recombination rate (ρ) calculation with CPLDhat as well as nucleotide diversity, Tajima's D , and Theta calculation with VariScan [52, 53]. Vcftools (v0.1.12b) was used to calculate F_{st} values with a 50 kb window [54]. To calculate F_{st} value for genes, SNPs located in gene regions were used for calculations in Vcftools. Gene presence and absence variation (PAV) was detected by searching the 70–15 gene sequence against each isolate with BLASTN ($E < 10^{-10}$).

Live-cell imaging analysis and *Agrobacterium*-mediated transient expression assay for suppression of plant cell death response in *N. benthamiana*

A 2 kb region of upstream sequence along with the coding region for *MGG_17227* (Guy11) was fused with mCherry: nuclear localization signal (NLS) or green fluorescent protein (GFP) for cloning into the pKNTG vector. *M. oryzae*

Guy11-transformed strains were used for rice leaf sheath inoculation and transformants with strong localization to the biotrophic interface complex (BIC) were documented with a Nikon NiE confocal microscope [55].

Agroinfiltration was performed as described with modifications [56, 57]. The coding region of *MGG_17227* with or without the signal peptide predicted by SignalP4.1 was cloned into pCXS_N with primers listed in supplemental Table S4 [58, 59]. *MGG_17227*, AvrPiz-t, and vector control strains were infiltrated into 5-week-old *N. benthamiana* leaves. After 24 h, plants were again infiltrated with *Agrobacterium* strains carrying the BAX expression clone. The infiltrated plants were kept at 22–25 °C with a 16 h light/8 h dark period for 3–4 days until cell death symptoms induced by BAX were detected. Leaves in which both positive and negative controls responded appropriately were then assessed for suppression of BAX-induced cell death.

Results

Phylogenomic characterization of rice-infecting *M. oryzae* populations

To characterize rice-infecting populations, we sequenced 88 rice isolates from different geographical locations with 32- to 104-fold coverage (average ~60-fold). We also acquired genome sequences of 12 previously published isolates, and thereby increased our sample size to 100 (supplemental Table S1). We used a genome-to-genome SNP calling approach because the sequencing reads for 10 out of 12 previously published genomes sequences were not available. A total of 751.6K SNPs combined from the 100 genomes covering 58.4K sites of the reference genome were identified, which represents about ~0.15% of the genome. The phylogenomic tree showed that the samples could be partitioned into three clades with two clades (clade 2 and clade 3) showing well-defined lineages that diverged from a common ancestral population, whereas the third clade (clade 1) contained more diversified samples from multiple lineages (Fig. 1a). To further validate the categorization of these isolates, we conducted PCA and STRUCTURE analysis. Consistent with phylogenomic result, both PCA and STRUCTURE analysis also categorized the isolates into three clades (Fig. 1b, c). $K = 2$ to $K = 5$ summary plots for samples indicate the most consistent number of genetic populations was $K = 3$ as determined using the Evanno method (Fig. 1b). Population studies by others showed that genetic lineages observed in populations of *M. oryzae* field isolates are globally distributed and our collection, includes isolates from Asia, Africa, and South America, is consistent with this global pattern (Fig. 1d).

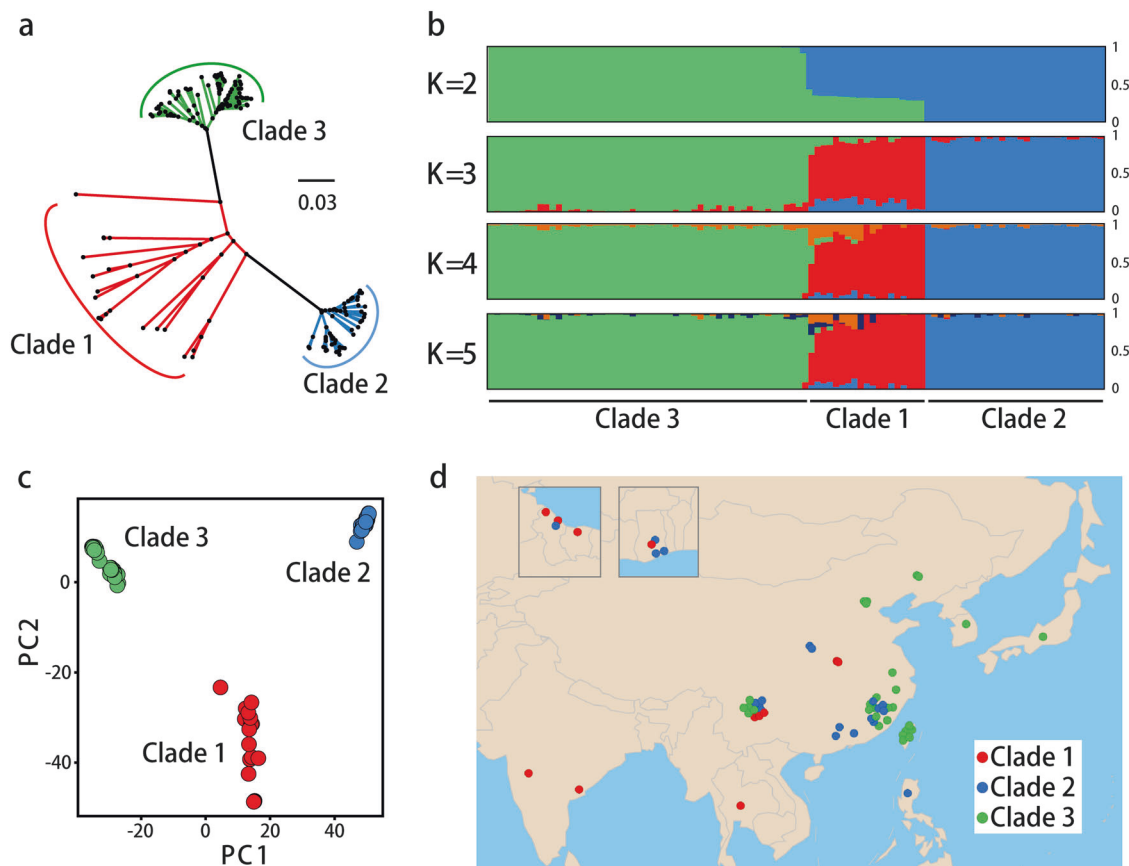


Fig. 1 Phylogenetic relationship of *M. oryzae* isolates. **a** Phylogenomic tree of *M. oryzae* isolates based on whole-genome SNP data. **b** STRUCTURE analysis of *M. oryzae* isolates. Each color in the plots represents the cluster membership coefficients. The presence of several colors in the same strain suggests admixture. **c** Principal component analysis (PCA) of *M. oryzae* isolates. **d** Origin of samples in this study.

To further test the population structure, we performed reads-to-genome SNP calling for the 90 isolates with available reads by mapping sequencing reads to reference genome and then used the corresponding results to construct a phylogenomic tree followed by PCA and STRUCTURE analysis. Reads-to-genome analyses yielded 954K SNPs covering 72.6K sites of the reference genome, and the results obtained with these SNPs are consistent with results obtained with genome alignment SNPs (Supplemental Fig. S1abc).

The divergence of *M. oryzae* field populations shows correlation with mating type

Reproduction isolation constitute major factors that contribute to population divergence, we therefore examined the reproduction machinery in our *M. oryzae* population to search for an imbalance that may provide evidence of clonality [60]. We found that the phylogenomic partitioning of isolates into three clades strongly correlated with the

Red, blue, and green dots represent isolates of clade 1, clade 2 and clade 3. Some dots represent more than one isolate with same collection site. Suriname, French Guiana (left), and Ghana (right) are presented separately in two windows, for more information see Table S1

Table 1 Mating type of *M. oryzae* in three clades

Clade	Mat1-1	Mat1-2	<i>n</i>
Clade 1	8	11	19
Clade 2	0	29	29
Clade 3	52	0	52

n: number of isolates in each clade

mating type pattern of the isolates (Table 1) and that all clade 3 isolates are Mat1-1 mating types, while all isolates in clade 2 are Mat1-2. Clade 1 consisted of both Mat1-1 and Mat1-2 isolates, including isolates from Yunnan, Hubei, Suriname, Guiana, Thailand, India and Ghana (Supplementary Table S1).

Origin and divergence of *M. oryzae* field populations

To trace back the evolutionary history of the *M. oryzae* population, we estimated divergence time of the three

clades using *S. viridis* isolates (SV, wild relative of rice isolates) as the outgroup (Fig. 2). This analysis showed that *M. oryzae* rice isolates and SV isolates diverged from each other 8000–12,000 YBP, coincident with rice domestication [20, 61]. Our analysis revealed that *M. oryzae* populations diverged at about ~1000 YBP. Unlike the genetically diverse clade 1, clades 2 and 3 appear to have expanded as clonal lineages at a similar time, just ~250–400 YBP.

To further characterize the difference between these three populations, we examined gene PAV between isolates within the three clades. We identified a total of 13,573 PAV events in all isolates and found clade 2 and clade 3 exhibit similar levels of gene deletion. Our analysis showed that the telomere regions are hot spots for gene deletion events (Supplemental Fig. S2). Further comparison of nucleotide diversity (π) and Ka/Ks value for genes in each clade yielded differential patterns with clade 1 displaying higher levels of π value with lower levels of Ka/Ks (Fig. 3a, b). Clade 2 and clade 3 isolates showed similar patterns of π and Ka/Ks values, which were distinct from clade 1.

Meanwhile, we observed the three clades of isolates showed similar patterns of whole-genome distribution of SNPs (Supplemental Fig. S3). However, intra-clade estimates of whole-genome nucleotide diversity (π), Tajima's D , recombination rate ($\rho = 2N_e r$), and inter-clade fixation index (F_{st}) exhibited significant variations between the three clades (Fig. 4 and Supplemental Table S3). Measures of nucleotide diversity and recombination rate for clade 1 ($\pi = 3.47 \times 10^{-4}$, $\rho = 4.36 \times 10^{-5}$) were significantly higher than those of clade 2 ($\pi = 1.88 \times 10^{-4}$, $\rho = 1.54 \times 10^{-5}$) and clade 3 ($\pi = 1.87 \times 10^{-4}$, $\rho = 2.65 \times 10^{-5}$). Low recombination rate is accompanied with low nucleotide diversity as well as significant negative value of Tajima's D (−0.96, −1.81, and −1.93 for clade 1, clade 2, and clade 3, respectively). The pair-wise F_{st} value in clade 2 vs. clade 3 ($F_{st} = 0.40$) is also elevated compared with that in clade 1 vs. clade 2 ($F_{st} = 0.14$) and clade 1 vs. clade 3 ($F_{st} = 0.15$), indicating gene flow is absent between clade 2 and clade 3 as expected for non-recombining lineages.

Identification of divergence-associated genes

We searched for genes with $F_{st} > 0.8$ between different clades to identify genes with differential distribution of alleles that might contribute to adaptation or differences in fertility. Such genes may represent the occurrence of divergent selection in gene regions between clades or be the result of fixation of relatively rare alleles at the time of divergence of the lineages. We found 769 genes in the analysis of clade 2 vs. clade 3, among which 516 genes are with non-synonymous substitutions (Fig. 5). Among these 516 genes, we further identified 11 genes that have $F_{st} > 0.8$ in clade 2 vs. clade 3 (Table 2) but < 0.3 in clade 1 vs. clade

2 or clade 1 vs. clade 3. We reasoned that such genes might represent alleles that were not selected against in clade 1, but could be adaptive in a clade-specific context or the result of the bottleneck associated with a clonal lineage. The biological annotation of these 11 genes showed 2 of the genes were involved in the pheromone response pathway, another 5 genes involved in metabolic or cellular pathways, and 4 genes with unknown function.

Isolates from single-mating-type lineages are generally of low female fertility. We thus examined sequence

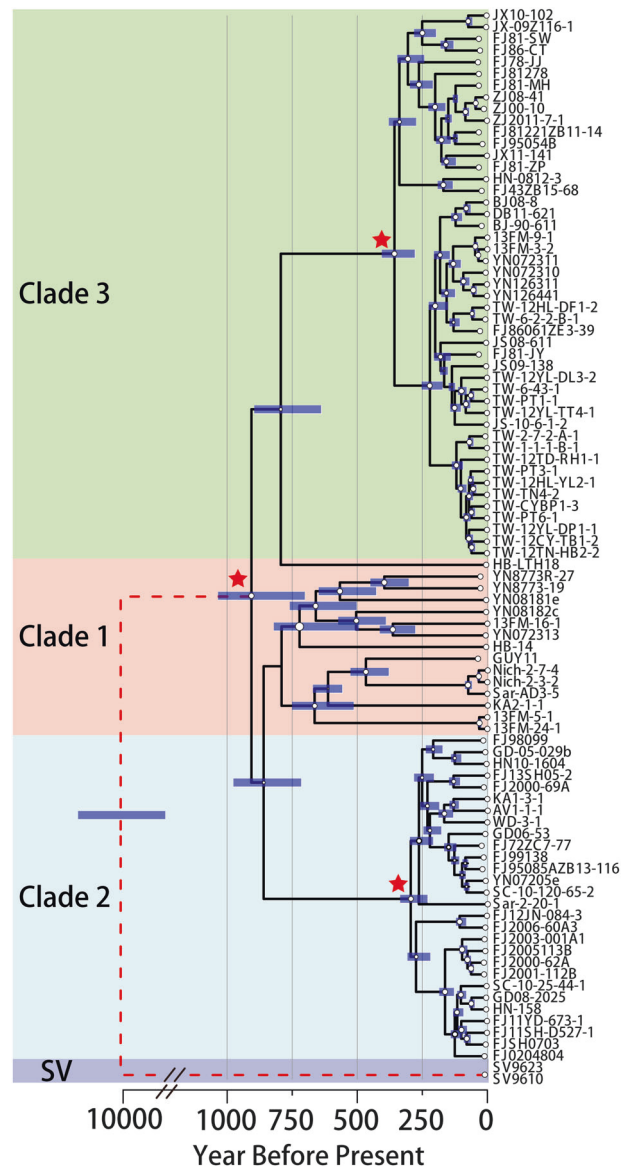
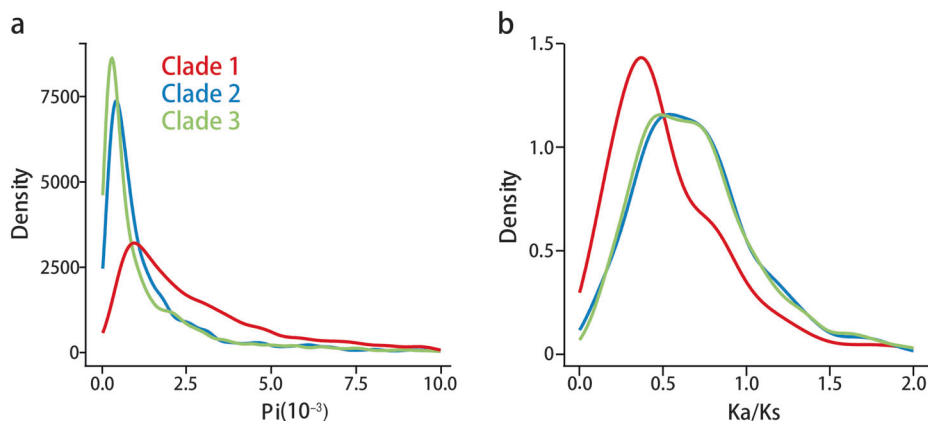


Fig. 2 Estimation of divergence time of *M. oryzae* populations. Chronogram from Bayesian phylogenetic analysis on whole-genome SNP dataset. Whole-genome SNP sequences were calibrated by collection dates. Blue bars indicate 95% highest posterior density intervals of node-age estimates. SV represent isolates collected from *S. viridis*. The red dashed line shows the break in the time axis of SV to rice population branch. Red stars marked divergence (star 2) or expansion (stars 1 and 3) time of clades

Fig. 3 Genomic divergence of *M. oryzae* field populations. **a** Distribution of Pi values against the corresponding density of genes for three clades. **b** Distribution of Ka/Ks values against the corresponding density of genes for three clades. The red, blue, and green colored lines represent the number of genes of clade 1, clade 2, and clade 3, respectively



variation of mating pathway genes in addition to the two we identified above. We observed 173 SNPs or InDels in 28 genes, and most of the mating pathway-related components were conserved or contained variation that was present across all three clades. For example, we observed a synonymous substitution in *MGG_05199* (*STE50*) for isolates in all three clades (Supplemental Table S2). However, in *MGG_12821* (*MST20*), we identified 53 SNPs/InDels and one of these is a T485A non-synonymous mutation found exclusively in all isolates of clade 3. Amino acid 485 lies between the BEM1-binding motif and the kinase domain [62]. In addition, the pheromone precursor (*MFA2*) gene, *MGG_07733*, was associated with a 30 bp deletion (hereafter full-length allele: type I; 30 bp deletion allele: type II) at position 79 bp of its open reading frame sequence in all clade 3 isolates. Interestingly, the *MFA2* type II allele also appears in a clade 1 lineage (Supplemental Fig. S4) consisting of four isolate sequences deposited at NCBI (isolates 2303.1, 1106.2, K96-07, and K98-10) (Supplemental Fig. S4b). In contrast to clade 3 Mat1-1 isolates, these four isolates were Mat1-2 and also contained a non-functional J-type allele of *AVRI-CO39* [63, 64].

Absence of a putative effector gene defines clade 2 isolates

We identified a small, cysteine-rich secreted protein (*MGG_17227*) that was absent in all clade 2 isolates. The end points of the deletion show sequence identity to the LINE retrotransposon MGL, showing a gene replacement by MGL in all clade 2 isolates replacing sequence from 3757 bp upstream to 758 bp downstream of the coding region. This 4515 bp region includes part of a second hypothetical protein of 303 amino acids (Fig. 6a). The *MGG_17227* gene was shown to be expressed *in planta* but not *in vitro* (Supplemental Fig. S5a). To assess localization *in planta*, a GFP fusion was constructed using the native promoter. The GFP fusion was localized to the BIC during infection of rice (Supplemental Fig. S5b). A second fusion

to mCherry containing a NLS was used, and localization to the BIC and to the plant nucleus confirmed that the protein was delivered into the plant cytoplasm (Supplemental Fig. S5c).

To test for a potential role for *MGG_17227* to act as an effector, we cloned the coding region with and without the signal peptide for testing in *N. benthamiana* [58]. AvrPiz-t is an effector that suppresses cell death mediated by BAX and was used as a positive control (Fig. 6b) [56]. The *MGG_17227* constructs with and without the signal peptide also displayed inhibition of BAX-mediated cell death.

Discussion

M. oryzae rice-infecting isolates from different geographic locations could be grouped into three major clades as shown in our analysis and a former study based on microsatellites markers [6]. The times of divergence could be estimated by taking into account the year of isolate collection and the SNPs between isolates. Based on our own, and a recent study with an independent set of rice isolates [28], these three main groups diverged from each other approximately 700–1000 years ago. Clade 1 genetic diversity is inferred to have expanded from a small population at that time. It was difficult by this analysis to resolve if clades 2 and 3 were derived from clade 1 or were parallel lineages that had diverged earlier from a common ancestral population that later gave rise to clade 1. Based on Supplemental Fig. S4b, the grouping we defined as clade 1 may contain as many as five lineages.

Clades 2 and 3 appear to have diversified recently, within the past ~250–400 years. This is consistent with the recent findings of Gladieux et al. [28] and they attribute these expansions to events in the history of rice cultivation. We found an effector gene, *MGG_17227*, that was absent in all clade 2 isolates. Within clade 1, the closely related isolate pair 13FM-16-1 and YN072313 share a different deletion event that eliminates *MGG_17227* as does a second closely

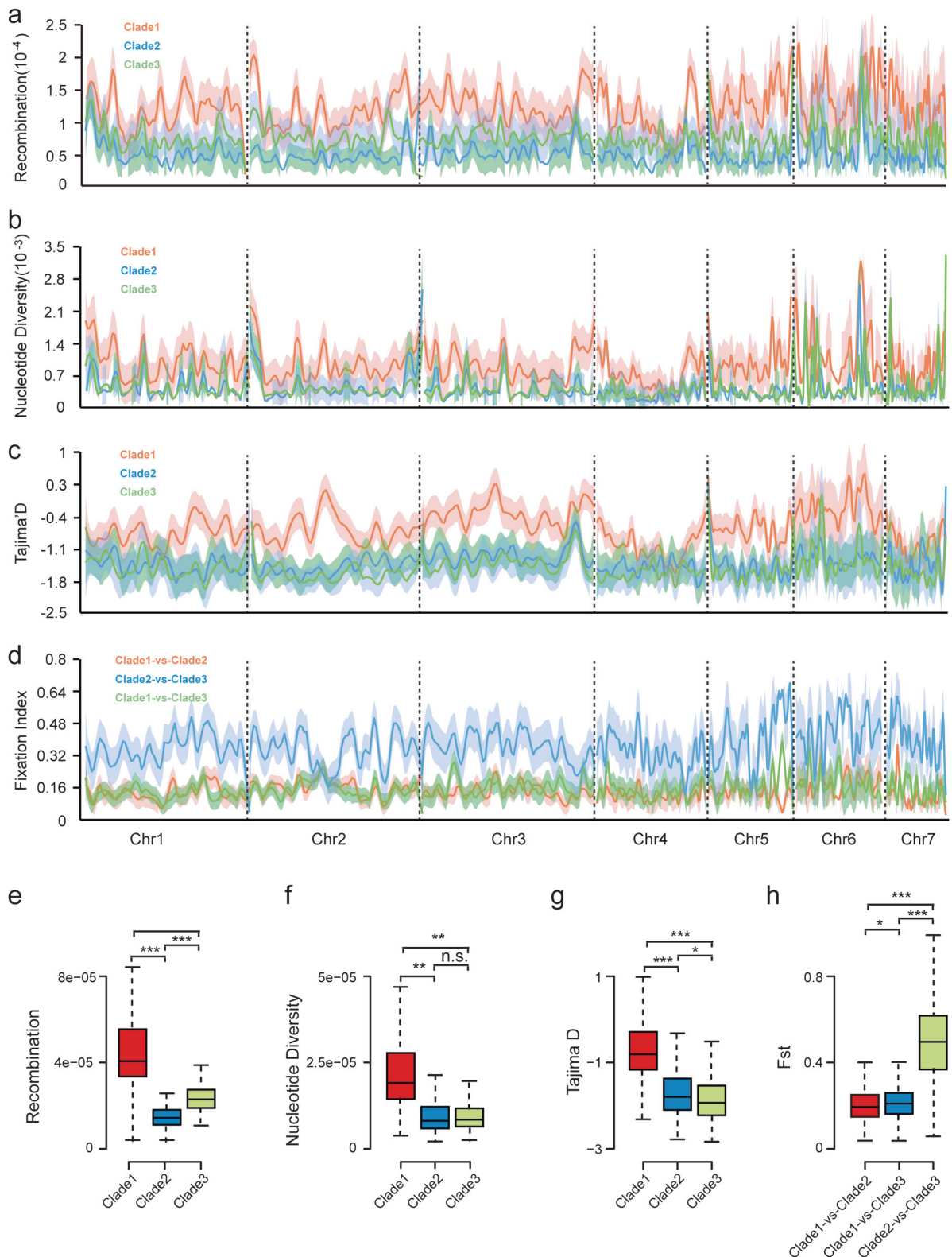


Fig. 4 Recombination rate (ρ), nucleotide diversity (π), and Tajima's D for the three clades of *M. oryzae*. **a-d** Whole-genome distribution of recombination rate (ρ), nucleotide diversity (π), Tajima's D , and F_{st} on chromosome I ~ VII of *M. oryzae* with 50 kb windows. **e-h** Boxplot of recombination rate, π , Tajima's D , and F_{st} values of three clades. The

Tukey whiskers indicate 1.5 times the interquartile range from the 25th and 75th percentiles. n.s. represents no significance, * represents p -value smaller than 0.05, ** represents p -value smaller than $1e-5$, *** represents p -value smaller than $1e-10$

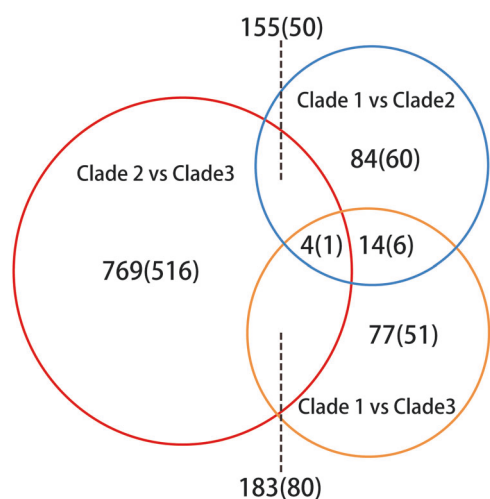


Fig. 5 Venn diagram of genes with $F_{st} > 0.8$ in the three clades. Venn diagram depicting the pair-wise comparison of genes with $F_{st} > 0.8$ between the three clades. Numbers in brackets represent the number of genes with non-synonymous substitutions

related pair, 13FM-5-1 and 13FM-24-1. This indicates that isolates lacking this effector can persist in clade 1 and clade 3 populations. For effectors that are recognized by plant resistance genes (also known as avirulence genes), gene PAV is common [65]. Conceivably, *MGG_17227* may also be recognized as an avirulence gene and deletion of the gene led to the expansion of the clade, perhaps following the movement of cultivars with a corresponding resistance gene. If so, clade 2 isolates transformed with *MGG_17227* could provide a useful tool in screening for such a resistance gene. The *MGG_17227* allele should allow for simple initial screening to identify clade 2 isolates by amplification with appropriate primers to distinguish the clade 2 replacement allele.

The view that clonal evolution is associated with clade 2 and clade 3 is supported by lack of clear recombination events within the genome sequences. To search for a potential genetic barrier to mating, we examined genes involved in fungal mating pathways. F_{st} comparison identified two mating pathway-related genes, orthologous to *Saccharomyces cerevisiae* Far1 and Smr1, containing non-synonymous substitution alleles displaying high F_{st} values in the clade 2 vs. clade 3 comparison. We observed a unique non-synonymous substitution allele of another *S. cerevisiae* mating pathway ortholog, *MST20* [62], fixed in clade 3. Clade 3 was also associated with a 30 bp deletion in the *MFA2* locus. Further study will be needed to determine whether these alleles impact mating behavior.

The *MFA2* gene is expressed in a mating-type-specific manner and should not be expressed in the Mat1-1 isolates of clade 3, but would be expressed in Mat1-2 isolates [43]. Interestingly, the *MFA2* type II allele is also present in a Mat1-2 clade 1 lineage of four isolates that also uniquely

contain non-functional *AVRI-CO39* J-type alleles [63, 64]. Couch et al. assessed populations of *M. oryzae* from different hosts and found 70 *Panicum* isolates clustered within the rice pathogen clade. The *Panicum* isolates contained *AVRI-CO39* homology by hybridization, as did the most closely related rice isolate [20], which we assume is a member of the group described here. They also proposed that a rice pathogen population invaded *Panicum* as a new host. Notably, the *Panicum* isolates in that study are not pathogenic to rice. Analysis of *Panicum* isolate *AVRI-CO39* alleles would help determine if this was simply a host shift from rice to *Panicum*, in which case one would expect to see the J-type allele in *Panicum* isolates.

The occurrence of the clade 3 *MFA2* type II allele in this group is also interesting. We speculate that either the *MFA2* type II allele existed prior to divergence of the clades, which involved recombination along the path for at least one of the lineages, since the two lineages are of opposite mating type, or that the occurrence of the clade I *MFA2* type II allele is a result of a rare recombination of clades 1 and 3. Detection of the *MFA2* allele type and mating type may serve as a simple means to initially screen for members of clade 3 and this clade 1 lineage.

Our study indicates that the divergence time of *M. oryzae* rice isolates from the closely related population infecting *S. viridis*, corresponded with the time of domestication of rice [20, 28]. As has been noted by others [27, 28], we noticed that most isolates collected from japonica rice-growing regions such as Beijing, Japan, Korea, Liaoning, Taiwan, and Fujian (in the 1980s), are grouped together in clade 3 and most isolates collected from indica rice-growing regions, such as Fujian since the 1990s are grouped in clade 2. The Fujian Province example indicates a shift from clade 3 to clade 2 isolates as the predominant members of the pathogen population was driven by a host genotype shift. Clade 1 contains isolates mostly have been sampled from regions that are associated with the cultivation of both japonica and indica, including Yunnan, India, and Thailand [66]. This observation further supports earlier research showing that japonica and indica rice cultivars exhibit differential susceptibility and resistance toward different isolates within the *M. oryzae* population [67–70].

Consistent with the view that clades 2 and 3 represent clonal lineages, our analyses found little to no evidence of recombination within clades 2 and 3, reduced allelic diversity, and lower genotypic diversity between isolates within the two clades compared to clade 1. The F_{st} value between clade 2 and clade 3 is higher than that between either clonal lineage with clade 1. This highlights the high genetic differentiation between the ancestral isolates leading to the clade 2 and 3 clonal lineages.

Although all clades are globally distributed, local populations may be composed predominantly of isolates from a

Table 2 Annotation of genes with low (<0.3) F_{st} in clade 1 vs. clade 2 and clade 1 vs. clade 3, and high (>0.8) in clade 2 vs. clade 3

ID	Yeast ortholog	F_{st}			Description	GO
		Clade 1 vs. clade 2	Clade 1 vs. clade 3	Clade 2 vs. clade 3		
MGG_00731	FAR11	0.27	0.22	1.00	Required for hyphal anastomosis; protein involved in recovery from cell cycle arrest; acts in response to pheromone	GO:1903436, regulation of mitotic cytokinetic process
MGG_01846	APL6	0.29	0.20	1.00	Transport of alkaline phosphatase to the vacuole via the alternate pathway	GO:0045184, establishment of protein localization
MGG_02874	KOG1	0.04	0.09	1.00	WD repeat-containing protein mip1; subunit of TORC1	GO:0019222, regulation of metabolic process
MGG_03254	MDR1	0.24	0.09	1.00	TBC1 domain family member 10A; cytoplasmic GTPase-activating protein; regulation of Golgi secretory function	GO:0043169, cation binding
MGG_04563	NA	0.27	0.18	1.00	NA	GO:0046872, metal ion binding
MGG_05466	NA	0.29	0.20	1.00	NA	NA
MGG_08984	OPT2	0.29	0.20	1.00	Oligopeptide transporter	GO:0051234, establishment of localization
MGG_10558	NA	0.08	0.15	0.87	NA	NA
MGG_10887	NA	0.23	0.09	0.97	NA	NA
MGG_13215	NA	0.07	0.30	1.00	NA	NA
MGG_13576	SRM1	0.27	0.24	1.00	RCC1 domain-containing protein; nucleotide exchange factor for Gsp1p; suppressor of the pheromone response pathway	GO:0008168, methyltransferase activity

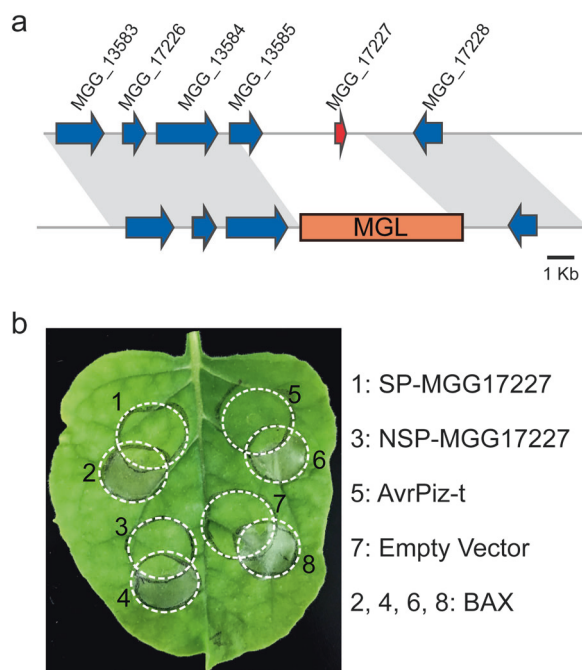


Fig. 6 Deletion of a putative effector gene defines clade 2 isolates. **a** Schematic of *MGG_17227* in all of clade 2 isolates indicating *MGG_17227* was replaced by a 5977 bp LINE retrotransposon MGL from 3757 bp upstream to 758 bp downstream of the coding region. **b** *MGG_17227* suppresses BAX-induced cell death following *Agrobacterium*-mediated transfection of tobacco leaves. Both *MGG_17227* full-length (1) and secretion signal-deleted construct (3) can suppress BAX-induced cell death (2, 4, 6, 8). AvrPiz-t (5) and empty vector (7) have been used as positive and negative controls, respectively

single clade. It has been recognized that lineage exclusion may be a strategy for deployment of resistance genes with enhanced durability [71, 72]. Thus, if lineage exclusion of clade 2 isolates could be employed in indica rice-growing areas (or clade 3 isolates in japonica rice-growing areas), the less-well-adapted clades would be able to cause disease, but severity may be reduced. However, areas where clade 1 isolates are present indicate the potential for enhanced recombination, therefore, determining the clades represented within local populations will provide useful information for breeding programs. The screening of the *MGG_17227*, *MFA2*, and *Mat* loci could help to quickly distinguish between clades (supplemental Table S5).

In the collection of isolates we analyzed, we found that a significant fraction of pathogen isolates are members of the two clonal clades that have spread globally. Clade 1 is a genetic reservoir that could generate new lineages to adapt to new host genotypes while it maintains a sufficient level of success to avoid being outcompeted by clonal lineages. It will be of interest to determine if a similar genetic structure occurs for *M. oryzae* populations parasitizing other hosts.

Acknowledgements We appreciate the scientists who have kindly shared isolates. We thank Dr. Ray Ming from the University of Illinois

at Urbana-Champaign, Dr. Haibao Tang from the University of Arizona, and Dr. Jiasui Zhan and Dr. Liangsheng Zhang from Fujian Agriculture and Forestry University for helpful suggestions.

Funding This work was supported by grants from the Natural Science Foundations of China (U1305211, 31770156, 91231121, and 31401692), the National Key Research and Development Program of China (2016YFD0300700), the Fujian Province 100 Talent Program, USDA NIFA Hatch project 1013944, China Scholarship Council, and the Scientific Research Foundation of the Graduate School of Fujian Agriculture and Forestry University.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Dean R, Van Kan JA, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, et al. The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol*. 2012;13:414–30.
- Ebbole DJ. Magnaporthe as a model for understanding host-pathogen interactions. *Annu Rev Phytopathol*. 2007;45:437–56.
- Bardwell L. A walk-through of the yeast mating pheromone response pathway. *Peptides*. 2005;26:339–50.
- Kang S, Chumley FG, Valent B. Isolation of the mating-type genes of the phytopathogenic fungus *Magnaporthe grisea* using genomic subtraction. *Genetics*. 1994;138:289–96.
- Kumar J, Nelson R, Zeigler R. Population structure and dynamics of *Magnaporthe grisea* in the Indian Himalayas. *Genetics*. 1999;152:971–84.
- Saleh D, Milazzo J, Adreit H, Fournier E, Tharreau D. South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, *Magnaporthe oryzae*. *New Phytol*. 2014;201:1440–56.
- Notteghem J, Silue D. Distribution of the mating type alleles in *Magnaporthe grisea* populations pathogenic on rice. *Phytopathology*. 1992;82:421–4.
- Zeigler, RS, Leong, SA, and Teeng, PS. 1994. Rice Blast Disease. CAB International, Wallingford, UK.
- Correll J, Harp T, Guerber J, Zeigler R, Liu B, Cartwright R, et al. Characterization of *Pyricularia grisea* in the United States using independent genetic and molecular markers. *Phytopathology*. 2000;90:1396–404.
- Genovesi A, Magill CW. Heterokaryosis and parasexuality in *Pyricularia oryzae* Cavara. *Can J Microbiol*. 1976;22:531–6.
- Hebert T. The perfect stage of *Pyricularia grisea*. *Phytopathology*. 1971;61:83–7.

12. Noguchi M, Yasuda N, Fujita Y. Evidence of genetic exchange by parasexual recombination and genetic analysis of pathogenicity and mating type of parasexual recombinants in rice blast fungus, *Magnaporthe oryzae*. *Phytopathology*. 2006;96:746–50.
13. Zeigler R, Scott R, Leung H, Bordeos A, Kumar J, Nelson R. Evidence of parasexual exchange of DNA in the rice blast fungus challenges its exclusive clonality. *Phytopathology*. 1997;87:284–94.
14. Crawford MS, Chumley FG, Weaver CG, Valent B. Characterization of the heterokaryotic and vegetative diploid phases of *Magnaporthe grisea*. *Genetics*. 1986;114:1111–29.
15. Bao J, Chen M, Zhong Z, Tang W, Lin L, Zhang X, et al. PacBio sequencing reveals transposable elements as a key contributor to genomic plasticity and virulence variation in *Magnaporthe oryzae*. *Mol Plant*. 2017;10:1465–8.
16. Yoshida K, Saunders DG, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, et al. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genomics*. 2016;17:370.
17. Gladieux P, Condon B, Ravel S, Soanes D, Maciel JLN, Nhani A, et al. Gene flow between divergent cereal and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. 2017. *bioRxiv*: 161513. <https://www.biorxiv.org/content/early/2017/11/15/161513>.
18. Chiapello H, Mallet L, Guérin C, Aguilera G, Amselem J, Kroj T, et al. Deciphering genome content and evolutionary relationships of isolates from the fungus *Magnaporthe oryzae* attacking different host plants. *Genome Biol Evol*. 2015;7:2896–912.
19. Zhong Z, Norvienyeku J, Chen M, Bao J, Lin L, Chen L, et al. Directional selection from host plants is a major force driving host specificity in *Magnaporthe* species. *Sci Rep*. 2016;6:25591.
20. Couch BC, Fudal I, Lebrun MH, Tharreau D, Valent B, van Kim P, et al. Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics*. 2005;170:613–30.
21. Inoue Y, Vy TT, Yoshida K, Asano H, Mitsuoka C, Asume S, et al. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. *Science*. 2017;357:80–83.
22. Dayakar B, Narayanan N, Gnanamanickam S. Cross-compatibility and distribution of mating type alleles of the rice blast fungus *Magnaporthe grisea* in India. *Plant Dis*. 2000;84:700–4.
23. Maciel JL, Ceresini PC, Castroagudin VL, Zala M, Kema GH, McDonald BA. Population structure and pathotype diversity of the wheat blast pathogen *Magnaporthe oryzae* 25 years after its emergence in Brazil. *Phytopathology*. 2014;104:95–107.
24. Onaga G, Wydra KD, Koopmann B, Sere Y, von Tiedemann A. Population structure, pathogenicity and mating type distribution of *Magnaporthe oryzae* isolates from East Africa. *Phytopathology*. 2015;105:1137–45.
25. Park S-Y, Milgroom M, Han S, Kang S, Lee Y-H. Genetic differentiation of *Magnaporthe oryzae* populations from scouting plots and commercial rice fields in Korea. *Phytopathology*. 2008;98:436–42.
26. Tredway L, Stevenson K, Burpee L. Mating type distribution and fertility status in *Magnaporthe grisea* populations from turfgrasses in Georgia. *Plant Dis*. 2003;87:435–41.
27. Saleh D, Xu P, Shen Y, Li C, Adreit H, Milazzo J, et al. Sex at the origin: an Asian population of the rice blast fungus *Magnaporthe oryzae* reproduces sexually. *Mol Ecol*. 2012;21:1330–44.
28. Gladieux P, Ravel S, Rieux A, Cros-Arteil S, Adreit H, Milazzo J, et al. Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. 2017. *bioRxiv*: 179895. <https://www.biorxiv.org/content/early/2018/01/29/179895>.
29. Dong Y, Li Y, Zhao M, Jing M, Liu X, Liu M, et al. Global genome and transcriptome analyses of *Magnaporthe oryzae* epidemic isolate 98-06 uncover novel effectors and pathogenicity-related genes, revealing gene gain and loss dynamics in genome evolution. *PLoS Pathog*. 2015;11:e1004801.
30. Gowda M, Shirke MD, Mahesh H, Chandarana P, Rajamani A, Chattoo BB. Genome analysis of rice-blast fungus *Magnaporthe oryzae* field isolates from southern India. *Genomics Data*. 2015;5:284–91.
31. Leung H, Borromeo ES, Bernardo MA, Notteghem JL. Genetic analysis of virulence in the rice blast fungus *Magnaporthe grisea*. *Phytopathology*. 1988;78:1227–33.
32. Chen C, Lian B, Hu J, Zhai H, Wang X, Venu RC, et al. Genome comparison of two *Magnaporthe oryzae* field isolates reveals genome variations and potential virulence effectors. *BMC Genomics*. 2013;14:887.
33. Park S, Kim S, Koh H, Lee Y. A blast lesion mimic mutant of rice. D. Tharreau, M.H. Lebrun, N.J. Talbot, J.L. Notteghem (Eds.), *Advances in rice blast*, Kluwer Academic Press, Dordrecht (2000), pp. 79–85.
34. Xue M, Yang J, Li Z, Hu S, Yao N, Dean RA, et al. Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. *PLoS Genet*. 2012;8:e1002869.
35. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 2005;6:31.
36. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol*. 2008;9:R7.
37. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*. 2005;434:980–6.
38. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
39. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–8.
40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
41. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
42. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004;5:R12.
43. Shen WC, Bobrowicz P, Ebbole DJ. Isolation of pheromone precursor genes of *Magnaporthe grisea*. *Fungal Genet Biol*. 1999;27:253–63.
44. Zhao X, Huang J, Yu H, Wang L, Xie W. Genomic survey, characterization and expression profile analysis of the peptide transporter family in rice (*Oryza sativa* L.). *BMC Plant Biol*. 2010;10:92.
45. Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. 2012;4:359–61.
46. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour*. 2009;9:1322–32.
47. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007;23:2633–5.
48. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007;7:214.
49. Librado P, Rozas J. DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25:1451–2.

50. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*. 2010;8:77–80.
51. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 2000;17:32–43.
52. Guo J, Jain R, Yang P, Fan R, Kwok CK, Zheng J. Reliable and fast estimation of recombination rates by convergence diagnosis and parallel Markov Chain Monte Carlo. *IEEE/ACM Trans Comput Biol Bioinformatics*. 2014;11:63–72.
53. Vilella AJ, Blanco-Garcia A, Hutter S, Rozas J. VariScan: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics*. 2005;21:2791–3.
54. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
55. Khang CH, Berruyer R, Giraldo MC, Kankanala P, Park S-Y, Czymmek K, et al. Translocation of *Magnaporthe oryzae* effectors into rice cells and their subsequent cell-to-cell movement. *Plant Cell*. 2010;22:1388–403.
56. Li W, Wang B, Wu J, Lu G, Hu Y, Zhang X, et al. 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*. 2009;22:411–20.
57. Ma L, Lukasik E, Gawehns F, Takken FL. The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in *Nicotiana benthamiana* leaves. *Methods Mol Biol*. 2012;835:61–74.
58. Chen S, Songkumarn P, Liu J, Wang GL. A versatile zero background T-vector system for gene cloning and functional genomics. *Plant Physiol*. 2009;150:1111–21.
59. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8:785–6.
60. Stukenbrock EH. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytol*. 2013;199:895–907.
61. Ma K, Xiao J, Li X, Zhang Q, Lian X. Sequence and expression analysis of the C3HC4-type RING finger gene family in rice. *Gene*. 2009;444:33–45.
62. Li L, Xue C, Bruno K, Nishimura M, Xu JR. Two PAK kinase genes, CHM1 and MST20, have distinct functions in *Magnaporthe grisea*. *Mol Plant Microbe Interact*. 2004;17:547–56.
63. Farman ML, Eto Y, Nakao T, Tosa Y, Nakayashiki H, Mayama S, et al. Analysis of the structure of the AVR1-CO39 avirulence locus in virulent rice-infecting isolates of *Magnaporthe grisea*. *Mol Plant Microbe Interact*. 2002;15:6–16.
64. Peyyala R, Farman ML. *Magnaporthe oryzae* isolates causing gray leaf spot of perennial ryegrass possess a functional copy of the AVR1-CO39 avirulence gene. *Mol Plant Pathol*. 2006;7:157–65.
65. Stergiopoulos I, de Wit PJ. Fungal effector proteins. *Annu Rev Phytopathol*. 2009;47:233–63.
66. Xiong Z, Zhang S, Wang Y, Ford-Lloyd BV, Tu M, Jin X, et al. Differentiation and distribution of indica and japonica rice varieties along the altitude gradients in Yunnan Province of China as revealed by InDel molecular markers. *Genet Resour Crop Evol*. 2010;57:891–902.
67. Gallet R, Fontaine C, Bonnot F, Milazzo J, Tertois C, Adreit H, et al. Evolution of compatibility range in the rice—*Magnaporthe oryzae* system: an uneven distribution of *r* genes between rice subspecies. *Phytopathology*. 2016;106:348–54.
68. Liao J, Huang H, Meusnier I, Adreit H, Ducasse A, Bonnot F, et al. Pathogen effectors and plant immunity determine specialization of the blast fungus to rice subspecies. *eLife*. 2016;5:e19377.
69. Youyong Z, Hairu C, Yunyue W, Zuoshen L, Yan L, Jinghua F, et al. Diversifying variety for the control of rice blast in China. *Biodiversity*. 2001;2:10–14.
70. Zhu Y, Chen H, Fan J, Wang Y, Li Y, Chen J, et al. Genetic diversity and disease control in rice. *Nature*. 2000;406:718–22.
71. Gibbons JW, Gonzalez D, and Delgado D (2000). Use of lineage exclusion in a multi-objective rice breeding program. In: Tharreau D, Lebrun, MH, Talbot N, Notteghem JL (eds). *Advances in Rice Blast Research: Proceedings of the 2nd International Rice Blast Conference*. Springer, pp. 146–153.
72. Gnanamanickam, SS, Babujee, L, Priyadarisini, B, Dayakar, BV, Leenakumari, D, Sivaraj, R, Levy M, Leong, SA (2000). Lineage-exclusion resistance breeding: Pyramiding of blast resistance genes for management of rice blast in India. In: Tharreau D, Lebrun, MH, Talbot N, Notteghem JL (eds). *Advances in Rice Blast Research: Proceedings of the 2nd International Rice Blast Conference*. Springer, pp. 172–179.