

Article

Comparative Genomics and Gene Pool Analysis Reveal the Decrease of Genome Diversity and Gene Number in Rice Blast Fungi by Stable Adaption with Rice

Qi Wu^{1,2,3,†} , Yi Wang^{1,†}, Li-Na Liu^{1,4}, Kai Shi⁵ and Cheng-Yun Li^{1,*}

¹ State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming 650201, China; wuqi_ynkm@ynau.edu.cn (Q.W.); wyi_0114@ynau.edu.cn (Y.W.); handyliu@126.com (L.-N.L.)

² College of Science, Yunnan Agricultural University, Kunming 650201, China

³ Yunnan Organic Tea Industry Intelligent Engineering Research Center, Key Laboratory of Intelligent Organic Tea Garden Construction in Universities of Yunnan Province, Key Laboratory for Crop Production and Smart Agriculture of Yunnan Province, Yunnan Agricultural University, Kunming 650201, China

⁴ Key Laboratory of Green Prevention and Control of Agricultural Transboundary Pests of Yunnan Province, Agricultural Environment and Resources Institute, Yunnan Academy of Agricultural Sciences, Kunming 650201, China

⁵ School of Foreign Language, Yunnan Agricultural University, Kunming 650201, China; sksdasda@163.com

* Correspondence: lichengyun@ynau.edu.cn

† These authors contributed equally to this work.

Abstract: *Magnaporthe oryzae* caused huge losses in rice and wheat production worldwide. Comparing to long-term co-evolution history with rice, wheat-infecting isolates were new-emerging. To reveal the genetic differences between rice and wheat blast on global genomic scale, 109 whole-genome sequences of *M. oryzae* from rice, wheat, and other hosts were reanalyzed in this study. We found that the rice lineage had gone through stronger selective sweep and fewer conserved genes than those of *Triticum* and *Lolium* lineages, which indicated that rice blast fungi adapted to rice by gene loss and rapid evolution of specific loci. Furthermore, 228 genes associated with host adaptation of *M. oryzae* were found by presence/absence variation (PAV) analyses. The functional annotation of these genes found that the fine turning of genes gain/loss involved with transport and transcription factor, thiol metabolism, and nucleotide metabolism respectively are major mechanisms for rice adaption. This result implies that genetic base of specific host plant may lead to gene gain/loss variation of pathogens, so as to enhance their adaptability to host. Further characterization of these specific loci and their roles in adaption and evaluation of the fungi may eventually lead to understanding of interaction mechanism and develop new strategies of the disease management.

Keywords: *Magnaporthe oryzae*; comparative genomics; gene pool; selective sweep; host adaption



Citation: Wu, Q.; Wang, Y.; Liu, L.-N.; Shi, K.; Li, C.-Y. Comparative Genomics and Gene Pool Analysis Reveal the Decrease of Genome Diversity and Gene Number in Rice Blast Fungi by Stable Adaption with Rice. *J. Fungi* **2022**, *8*, 5. <https://doi.org/10.3390/jof8010005>

Academic Editor: Ana Cristina Esteves

Received: 3 November 2021

Accepted: 17 December 2021

Published: 22 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

M. oryzae is an ascomycete pathogen that causes blast disease in more than 50 species of plants, especially many economic cereal crops such as rice, wheat, barley, millet, and oats [1–6]. Furthermore, some grasses including goosegrass, annual ryegrass, perennial ryegrass, tall fescue, and St. Augustine grass can also become hosts for it [7]. In the past 400 years, people used to focus on the rice blast disease management because of its great threat to rice, which is an important crop to human beings. However, unlike the rice blast, the epidemic of wheat blast disease caused by *M. oryzae* was first recorded in Brazil in 1985 followed by its spread to the United States in 2011 and then to Bangladesh [8–12]. Blast fungi threaten productions of two main crops, which indicates its significant economics, thus it is urgent to explore the genetic difference among *M. oryzae* populations. Recently, research showed that the population structure of *M. oryzae* is subdivided into multiple lineages with limited host range [13]. Southeast Asia is the center of origin, diversity,

and dispersion of the rice blast fungus. Except for the complex population structure of epidemic isolates in Southeast Asia, most of the epidemic isolates in Europe and Africa come from the same lineage [14]. The adaptive evolution of wheat-infecting isolates to host is complicated, some of which have clustered with *Lolium*-infecting isolates. At the same time, although rice-infecting isolates around the world are subdivided into different lineages, their diversity was lower than that of wheat-infecting isolates [13,15], which indicates that they are well-adapted to rice and have relatively stable genome.

During the co-evolution between plant and pathogen, the evolution of pathogens was directly related to the differentiation of hosts [16]. Due to domestication of crops with morphological and genetic variation, pathogens are subject to selection pressures from their hosts. Strong positive (natural) selection of a beneficial allele could lead to the selective sweep of adjacent linked sites, and eventually decreases of genetic diversity [17]. It is generally believed that positive selection plays an important role in the outcome of new phenotypes, and balancing selection is regarded as the main mechanism to maintain population diversity [18–20]. Compared with the core genome, many effector genes in pathogens present a trend of accelerated evolution to gain advantages in the arms race with their hosts. The RxLR effectors in *Phytophthora sojae* and Hopz family of type III secreted effectors in *Pseudomonas syringae* showed high polymorphism and positive selection during co-evolution with host [21,22]. Previous studies have showed that the selective sweep can have a significant impact on the level of population subdivision, especially when the sweep has not yet spread to all populations within one species [23–26]. For example, the genes in the strongest sweep region involving biotic and abiotic responses were found in the formation of two clusters of *Rhynchosporium commune* from Central Europe and Ethiopia, respectively [27].

Population genetic analysis of 446 wild rice and 1083 indica and japonica genomes showed that selective sweeps were accompanied with the whole domestication process and population subdivision in rice [28]. Rice and *Setaria* infecting isolates were diverged followed by rice domestication [13]. Therefore, there is a long co-adaptation history between *M. oryzae* and rice, and whether the events of selective sweeps happened in population of rice-infecting isolates needs to be discovered. In this study, population genetic statistics calculated on 10 kb window showed that compared with the alternating distribution of two different characteristic regions in the genome of wheat-infecting isolates, rice-infecting isolates were under selective sweep at the whole genome level and showed lower diversity. In addition, gene pool analysis indicated that rice-infecting isolates contained fewer genes than wheat- and *Lolium*-infecting isolates, and discovered some candidate loci of *M. oryzae* adapted to host and potential virulence factors at the same time. Our results showed that this comparison could be helpful to understand genomic instability, putative target genes for host adaptation and to understand how stable adaptation with host was realized to achieve a balanced population.

2. Materials and Methods

2.1. Genome Sequences

One hundred and nine published genome data of *M. oryzae* from eleven host genera were downloaded from the National Centre of Biotechnology Information assembly database (<http://www.ncbi.nlm.nih.gov/assembly>, accessed on 10 December 2021, Table S1) [10–13,16,29–40]. Among them, there were 63 rice-infecting isolates (the FR13, GUY11 and FJ81278 had both NGS and SMRT assembly data) and 43 non-rice-infecting isolates. In the population genetic statistics and gene pool analysis between different lineages of *M. oryzae*, 48 isolates were used. Among them, 24 non-rice-infecting isolates, including 13, 6, 2, 2, and 1 isolates from *Triticum aestivum*, *Lolium perenne*, *Lolium multiorum*, *Bromus tectorum*, and *Triticum durum* respectively, were combined into non-rice lineage (NRL) and compared with rice lineage (RL) which was composed of 24 rice-infecting isolates (Table S1, Figure S1). As a supplement, seven isolates of *Eleusine* lineage (EL) were used to verify the conclusions obtained from population genetic analysis. In order

to reveal the PAV of genes in rice and non-rice isolates, the 24 isolates of RL were selected according to the rank of genome size from NCBI database with a total ungapped length of each genome sequence over 39.0 Mb. These rice-infecting isolates came from 8 countries in South America, North America, Europe, Asia, and Africa, ranging from 1988 to 2015, while wheat-, *Lolium*-, and *Bromus*-infecting isolates come from Brazil, Bolivia, Paraguay, the United States and Bangladesh, ranging from 1988 to 2016. In the Figures S2, S4, S6 and S8, it could be seen that when comparing RL with different non-rice lineages of different population sizes, the selected isolates were distributed in the major branches of the phylogenetic tree constructed by SNP dataset of 148 global distributing rice-infecting isolates (all the genomes submitted to the NCBI assembly database before May 2020) rather than coming from a single branch. Other isolates in the Table S1 were used together with the 48 isolates mentioned above to analyze the diversified distribution patterns of 11 cloned AVR genes in *M. oryzae* from different hosts, and served as assistants and contrasts in gene pool analysis. Two *Stenotaphrum*-infecting isolates were included as outgroups. The reference genome 70-15 version 8 and other two Magnaportheaceae species (*Gaeumannomyces graminis* var. *tritici* and *Magnaporthe poae*) were downloaded from the Broad Institute (<https://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/magnaporthe-comparative-genomics-proj>, accessed on 10 December 2021).

2.2. Single-Nucleotide Polymorphism Discovery and Population Genetics

The genome-to-genome SNP calling approach [34] was used in this study as the pair-end reads for 27 published genome sequences of the 48 isolates were not obtained online. First, each assembly sequence was mapped to the reference genome by a program NUCmer of MUMmer v3.23 with parameters -maxmatch -c 100 [41]. Final SNP data were obtained by program show-snps with parameter -Clr. Our results showed that a total of 6.5 M SNPs obtained from the 48 genomes compared with 70-15, which covered 547.9 K sites of reference genome, were identified for population structures (Table S2). Due to the close genetic relationship between BdJes16-1, BdMeh16-1, BdBar16-1, and B71, we also carried out the genome-to-genome SNP calling again with the sequences spliced to chromosome level of B71 (GCA_004785725.2) as reference genome. It could be seen that except BdJes16-1, BdMeh16-1 and BdBar16-1, which only contain 679, 981, and 1212 SNPs, other isolates of NRL contain much more SNPs (ranging from 52.6–127.0 K, average 104.1 K, covering 350.9 K sites of B71) than those of rice-infecting isolates again with 70-15 (ranging from 10.3–23.6 K, average 13.0 K, covering 77.2 K sites of 70-15). These results indicate that there are high differences among these isolates of NRL, which can support our comparative genomic analysis later. Based on the SNP dataset, VariScan v2.0.3 was used to calculate nucleotide diversity (π) and Tajima's D [42] with PHYLIP input data files and parameter RunMode = 12, and population differentiation (F_{st}) was calculated by Genepop v4.7.2 [43] with MenuOptions = 6.2. All data were estimated with the 10 kb windows according to the reference genome. The grouping-specific SNPs (S-SNPs) which were all X in RL and Y in NRL (X, Y is one of A, T, G, and C) were screened out based on the occurrence frequency of SNPs at the same site in RL and NRL. The dS values were calculated by Paml v4.9c with yn00 model [44].

2.3. Gene Prediction, Annotation, and Secreted Proteins Prediction

Protein-coding genes were predicted by Augustus V3.0.3 [45]. The *Magnaporthe grisea* model was selected for Augustus prediction. Predicted genes with deduced protein length shorter than 50 amino acids, or overlapping with repeat sequences for more than 50% of their transcript length were removed. Some genes apparently belonging to the rice genome had also been eliminated. Protein sequences were functionally annotated against the EggNOG database [46]. The putative secreted proteins were identified by SignalP 3.0 [47]. The PAV of AVR genes were detected by searching the nucleotide sequence against each isolate with BLASTN ($E < 1 \times 10^{-10}$) [34].

2.4. Gene Pool Analysis

Homologous genes with sequence identities of 100%, 80–100%, and 50–80% were defined as identical, similar, and divergent, respectively, while those below 50% were considered non-homologous [48]. We performed BLASTP ($E < 1 \times 10^{-5}$) comparison on the protein-encoded genes predicted by Augustus in each genome, and removed the homologous genes with similarity more than 50%. The gene pool of 48 *Magnaporthe* isolates with 11,914 genes was obtained. The PAV of genes in the gene pool were detected by searching the protein sequences against each isolate with TBLASTN ($E < 1 \times 10^{-10}$), and the genome would contain the gene when the sequence identity and matching overlap were above 50% [16]. To identify genes of *Magnaporthe* under selection from different hosts, their present frequencies in different lineages were derived. The significance of the differences of the presence frequencies for each gene between the two lineages was determined by the Fisher's exact test. The resulting raw p values of 1599 identified PAVs were then corrected via false discovery rate (FDR). As the number of isolates we used for population comparison was small, genes with significantly different frequencies (FDR < 0.0001) were identified as those related to host adaptability of *M. oryzae* to lock the candidate gene as accurately as possible. The maximum-likelihood trees were constructed based on the SNP data with default parameter values by FastTree v2.1.9 [49] or the binary PAV data with 1000 bootstraps by IQ-TREE v2.1.4 [50], and displayed with FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>, accessed on 15 December 2021).

2.5. Identification of Expressed Genes

We used RNA-Seq data from samples 2, 5, 7, and 12 from Bangladesh wheat fields to identify the expression of grouping-specific genes of non-rice lineage, including comparisons between asymptomatic and symptomatic samples [12]. The specific genes expressed only after the symptoms were considered to be related to the host adaptability of wheat isolates. In addition, the RNA-Seq data of 98-06's mycelium and conidia infection at 0, 8, 24, 48, and 72 h post inoculation were used to compare the expression of grouping-specific genes of rice lineage *in planta* and *in vitro* culture [31]. The genes were selected according to two situations, the first is that genes were not expressed in one situation of *planta* or culture, the second is that genes showed differential expression.

2.6. Statistical Analysis

The statistical tests used are illustrated in the article and the figure legends. To be more precise, since the data did not conform to the normal distribution, we performed Spearman correlation analysis between the number of genes contained in the isolate and the four genomic features (total length, total ungapped length, number of sequences and N50). At the same time, Mann–Whitney U test was used to analyze whether there were significant differences in the four genomic features mentioned above between rice- and non-rice-infecting isolates. Fisher's exact test with FDR corrected for multiple comparisons was used in the search for specific genes of different lineages.

2.7. Data Availability

The data to confirm the conclusions are completely listed in the article and Supplementary Materials. The NCBI database accession number of the genome used for analysis is showed in Table S1. Detailed information such as predicted protein sequences, occurrence frequencies in different lineages, and gene function annotations of the 228 genes associated with host adaptation of *M. oryzae* are recorded in Table S4.

3. Results

3.1. Population Genomic Analysis of Rice- and Non-Rice-Infecting Isolates

Previous studies showed that the wheat-infecting isolates have closer genetic relationship with *Lolium*- and *Brome*- than that of rice-infecting isolates. Wheat-infecting isolates were divided into two lineages when phylogenetic trees were built by analyses

of single copy gene and pairwise distance. The *Triticum* lineage (TL) consists almost all of wheat-infecting isolates and the *Lolium* lineage (LL) comprises *Lolium*-, *Festuca*-, and *Brome*-infecting isolates besides several wheat-infecting isolates [13]. Compared with other host infecting isolates, rice-infecting isolates seem to possess strict host specificity and showed lower diversity [13,51]. To better understand the selection effect of rice on *M. oryzae*, the genetic divergence between 48 rice and non-rice isolates were analyzed. When the phylogenetic tree was constructed with genome-wide SNP data, it could be seen that isolate from same host were in the same branch except a few wheat-infecting isolates which was consistent with previous study (Figure S1).

To reveal the genetic diversity and differentiation between RL and NRL, F_{st} , π , and Tajima's D were used. The average F_{st} value of all 4045 windows was 0.73, indicating that there was high genetic differentiation between the two groups (Figure 1a). Moreover, the π and Tajima's D also showed different patterns between the two groups, indicating that RL had lower diversity than that of NRL (Figure 1b,c). When considering $\pi > 0.001$, more windows were found in NRL (58.9%) than that in RL (5%, Figure S2a–c). When calculating the diversity of intragenic and intergenic regions, the diversity of RL was also the lowest among all groups or lineages, which was similar to the previous results (Table 1) [13]. In addition, the windows with Tajima's $D > 0$ were estimated from RL (81.3%) and NRL (71.8%). The distribution patterns of windows with Tajima's $D > 0$ or < 0 were obviously different between the two groups (Figure 1c and Figure S2d–f). For NRL, more enriched windows with Tajima's $D > 0$ or < 0 seemed to cluster in certain regions on 2–7 chromosomes. Compared with the regions enriched by windows with Tajima's $D < 0$, most of the regions enriched by windows with Tajima's $D > 0$ showed higher diversity and lower F_{st} at the same time, which indicated that balancing selection played a dominant role in these regions, and there were many alleles with medium frequency. The F_{st} of most windows in these regions was still higher than 0.25, suggesting that there was a great genetic differentiation between RL and NRL [52]. When using NRL after removing BdJes16-1, BdMeh16-1, and BdBar16-1 and RL of the same population size to calculate the F_{st} , π , and Tajima's D , the results were similar to those before removing (Figures S3 and S4). Other similar results were also obtained when we compared RL with TL and LL, respectively (Figure S5–S8). In addition, after comparing the F_{st} between different lineages, it was found that the degree of genetic differentiation between rice lineage and non-rice lineages was significantly higher than that between non-rice lineages (Figure S9a,b). At the same time, the diversity of RL was significantly lower than that of non-rice lineages with the same number of isolates, and EL had the highest diversity among all the non-rice lineages (Figure S9c). These results indicated that some regions in the genomes of NRL were influenced by the balance selection and showed significantly higher diversity than RL. However, the lower diversity in RL implied that in the long history, breeding of rice varieties has brought strong selection, which might lead the genetics of rice infecting isolate populations toward simplification and specificity.

Table 1. Diversity in intronic regions and intergenic regions among different lineages.

Lineage	Intronic Regions	Intergenic Regions
RL ($n = 24$)	0.0002665	0.0005304
NRL ($n = 24$)	0.0021783	0.0035670
TL ($n = 14$)	0.0021997	0.0033026
LL ($n = 10$)	0.0016510	0.0029569
EL ($n = 7$)	0.0032822	0.0050494

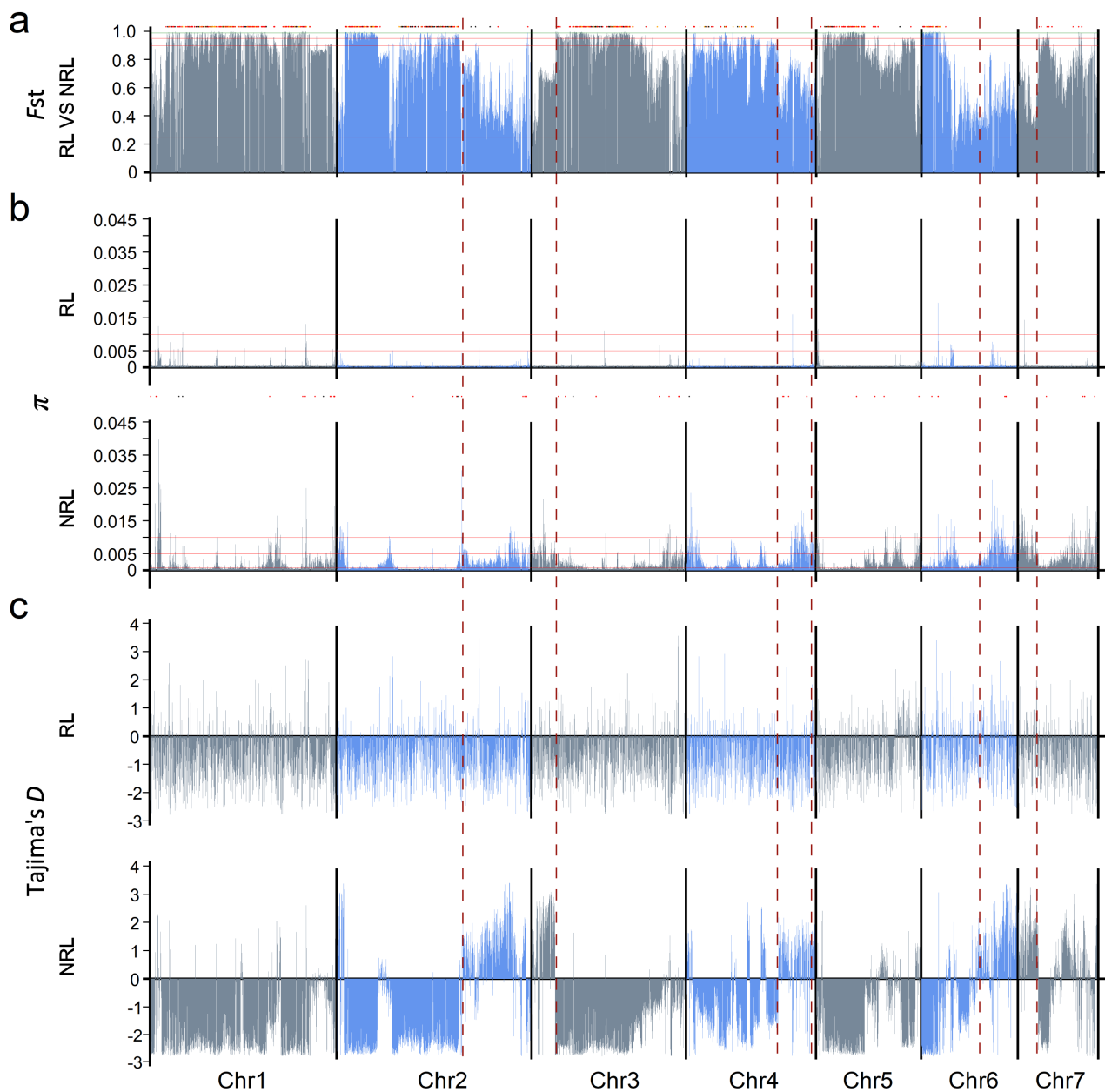


Figure 1. Genome-wide view in 10 kb windows of differentiation (F_{st}), diversity (π) and Tajima's D . (a) The F_{st} compared between RL and NRL. The RL included 24 rice-infecting isolates with the largest ungapped length of genome in NCBI database, and the NRL included 14 wheat-, 8 *Lolium* and 2 *Bromus*-infecting isolates. The locations of 856 genes mutated by S-SNPs were showed at the top of F_{st} . The orange dot indicated that there was at least one S-SNP is in CDS, the red dot indicated that non-synonymous substitutions had occurred in CDS, and the black dot indicated that all S-SNPs were located in intron or untranslated region. The distributions of π (b) and Tajima's D (c) of RL and NRL. The area sandwiched by the red dotted line in the vertical direction represented the area enriched by window of Tajima's $D < 0$ in NRL. In these regions, the π of NRL increased while the F_{st} between RL and NRL decreased. The dots between π of RL and NRL in (b) were the favorable genes present in 70-15. Among them, red and black indicated favorable genes of RL and NRL respectively.

3.2. SNPs with Host Selection

According to previous genetic differentiation analysis of *M. oryzae* population, 1411 windows were found based on $F_{st} > 0.9$ and 97.7% of windows contained 1–1500 SNPs

(Figure 2a). A total of 119,637 S-SNPs covering 8525 genes were obtained and 510 windows (12.6%) did not contain S-SNPs (Figure 2b,c). Among 8525 genes possessing at least one S-SNP, 856 genes only contained S-SNPs, and 7669 genes contained both S-SNPs and moderate-frequency alleles. There were 702 genes with S-SNPs distributed in coding sequences (CDS), and 509 of which with $dS > 0$. Among these 856 genes with S-SNPs, it could be seen that almost all the genes (84.2%) were distributed in the windows with $F_{st} > 0.9$, a few genes (15.8%) were found in several regions of all chromosomes with $F_{st} < 0.9$ (Figure 1a), especially the last third of chromosome 2. Therefore, a large number of S-SNPs involved in genetic differentiation between RL and NRL were widely distributed in the genomes, indicating that these two groups were evolving toward separate trajectories, and it was difficult to find out the genes related to the adaptability of *M. oryzae* by population genetic statistics or association analysis based on the SNP dataset.

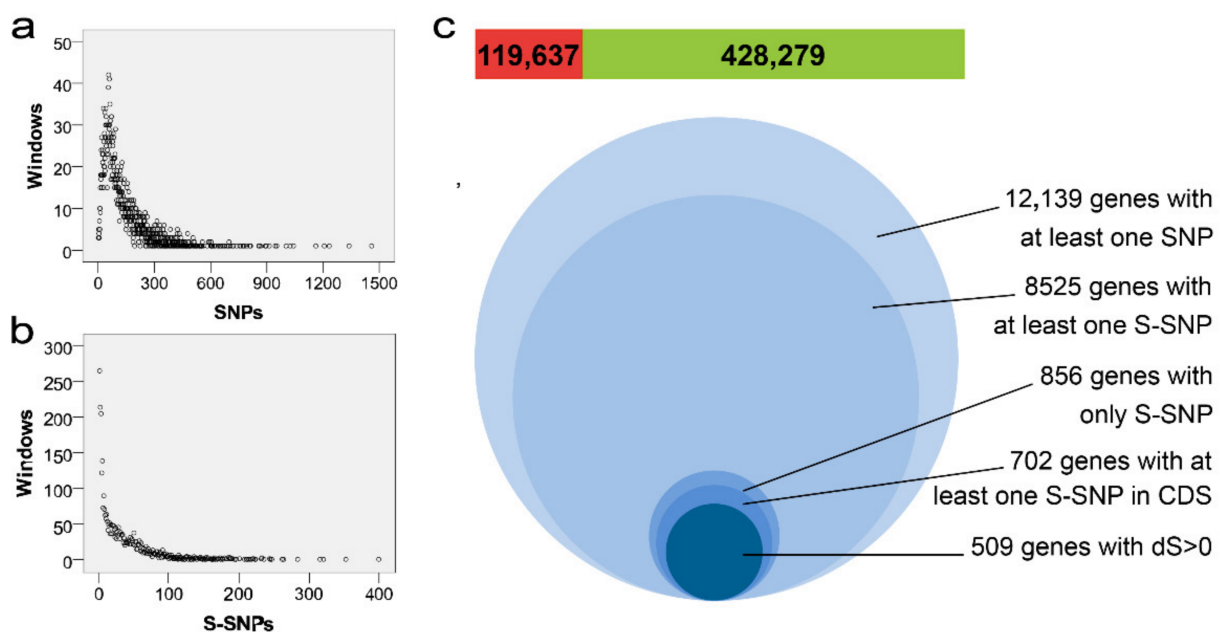


Figure 2. Windows and Genes affected by SNPs. Distribution of windows on chromosome against the corresponding SNPs (a) and S-SNPs (b). (c) The number of sites in the two classes and the genes affected by different mutations. The numbers in the green and red bars indicated the number of identified sites for S-SNPs and SNPs, respectively. Circles with different sizes represented the number of genes with different mutations due to SNP.

3.3. Highly Polymorphic Distribution of 11 Cloned AVR Genes in Two Groups

Rice and *M. oryzae* were regarded as an ideal model for studying host-pathogen interaction. The arms race between *R* and *AVR* genes were involved in the recognition between host and pathogen. There were more than 40 *AVR* genes identified in *M. oryzae*. Eleven *AVR* genes have been cloned and characterized by their cognate Rgenes, including *PWL1* [53], *PWL2* [54], *AVR-Pita* [55], *ACE1* [56], *AVR-Pia*, *AVR-Pii*, *AVR-Pik* [57], *AvrPiz-t* [58], *AVR1-CO39* [59], *AVRPib* [60], and *AVRPi9* [61]. Except for 48 genomes used for SNP analysis, 57 *Magnaporthe* genomes from 11 different hosts were added to select the *AVR* genes that were involved in the genetic differentiation together (Figure 3).

Host	Isolate	AVR1-CO39	AVR-Phi	PWL1	AVR-Pia	AVR-Pik	AVR-Pita	PWT4	PWL2	AVR-Pib	PWT3	AerPiz-t	AVR-P19	ACE1		
<i>Pennisetum americanum</i>	PM1	/	/	/	/	/	D	/	D	/	/	/	/	D		
	VO107	/	/	/	D	/	D	/	D	/	/	/	/	D		
	Digitalia sanguinalis	DsL1Z	/	/	B	/	D	/	D	/	/	/	/	D		
<i>Eletusine indica</i>	EI9411	B	/	D	/	/	/	/	/	D	/	D	D	D		
	EI9604	B	D	/	/	/	/	/	/	D	D	D	D	D		
	B51	B	D	/	/	/	/	/	/	D	D	D	D	D		
<i>Eletusine coracana</i>	G22	B	/	D	/	/	/	/	/	D	D	D	D	D		
	MG12	B	D	D	/	/	/	/	/	D	D	D	D	D		
	MG34	B	/	D	/	/	/	/	/	D	D	D	D	D		
<i>Stenotaphrum secundatum</i>	PH42	D	D	/	/	/	/	/	/	D	D	D	D	D		
	SSFL02	D	/	/	/	/	/	/	/	D	D	D	D	D		
	SSFL14.3	D	D	/	/	/	/	/	/	D	D	D	D	D		
<i>Setaria viridis</i>	SV9610	B	/	/	/	/	D	/	/	D	B	D	A	/		
	SV9623	B	/	/	/	/	D	/	/	D	C	D	A	B		
	MG08	B	/	/	/	/	D	/	/	D	C	D	B	B		
<i>Cenchrus ciliaris</i>	MG07	D	D	/	/	/	/	/	/	D	/	D	D	D		
	Brachiaria distachya	Bd4801	D	/	/	D	/	/	C	/	D	D	A	B		
<i>Brachiaria mutica</i>	Bm88324	D	D	/	/	/	/	/	/	/	/	D	D	A	D	
	<i>Eragrostis curculula</i>	G17	D	/	/	/	/	/	/	/	D	D	D	D	D	
<i>Triticum aestivum</i>	Bdjes16-1	D	/	/	/	/	/	/	/	D	D	D	D	D		
	BdMch16	D	/	/	/	/	/	/	/	D	D	D	D	D		
	BdBar16-1	D	/	/	/	/	/	/	/	D	D	D	D	D		
	Py22.1	D	/	/	/	/	/	/	/	D	D	D	D	B		
	Py5020	D	/	/	/	/	/	/	/	D	D	D	D	B		
	B2	D	/	/	/	/	/	/	/	D	D	D	D	B		
	B71	D	/	/	/	/	/	/	/	D	D	D	D	D		
	B67	D	/	/	/	/	/	/	/	/	D	D	D	D	B	
	B680	D	/	/	/	/	/	/	/	/	D	D	D	D	A	B
	Bt130	D	/	/	/	/	/	/	/	/	D	D	D	D	A	D
<i>Triticum durum</i>	T25	D	/	/	/	/	/	/	/	D	D	D	D	D	B	
	WHITQ	D	/	/	/	/	/	/	/	D	D	D	D	D	D	
	WBKY11	D	/	/	/	/	/	/	/	D	D	D	D	A	D	
	F3	D	/	/	/	/	/	/	/	D	D	D	D	D	B	
	CHRF	D	/	/	/	/	/	/	/	D	D	/	D	A	B	
<i>Lolium perenne</i>	CHW	D	/	/	/	/	/	/	/	D	D	D	A	B		
	FH	D	/	/	/	/	/	/	/	D	D	D	A	D		
	GG11	D	/	/	/	/	/	/	/	D	D	D	A	B		
<i>Lolium multiflorum</i>	HO	D	/	/	/	/	/	/	/	D	D	D	A	B		
	LpKY97	D	/	/	/	/	/	/	/	D	D	D	A	B		
	PL2-1	D	/	/	/	/	/	/	/	D	D	D	A	B		
<i>Bromus tectorum</i>	PL3-1	D	/	/	/	/	/	/	/	D	D	D	A	B		
	P-0028	D	/	/	/	/	/	/	/	D	D	D	A	B		
	P-0029	D	/	/	/	/	/	/	/	D	D	D	D	D		
	KJ201	/	/	/	/	B	D	A	D	C	A	A	A	A	B	
	98-06	/	/	/	/	B	D	A	D	C	A	A	A	A	B	
	IB33	/	/	/	/	D	/	D	C	A	A	A	A	B		
	FR13	/	/	/	B	B	D	C	D	C	C	A	A	A	B	
	IE1K	/	/	/	/	/	/	/	/	B	C	A	A	A	B	
	IC17	/	/	/	/	/	D	/	/	D	C	A	A	A	B	
	ML33	/	/	/	/	B	D	D	D	D	D	A	A	A	B	
<i>Oryza sativa</i>	IA1	/	/	/	/	B	D	/	/	D	C	A	C	A	B	
	HO81d	/	/	/	/	D	D	D	B	/	D	A	A	A	D	
	Ks131	/	/	/	/	B	D	D	B	D	C	A	A	A	B	
	MG02	/	/	/	D	/	D	D	D	C	C	C	B	D		
	YN126311	/	/	/	D	D	/	B	D	C	A	A	A	B		
	YN08182d	/	/	/	/	D	D	C	B	C	C	A	A	B		
	YN072310	/	/	/	D	D	D	B	D	C	C	A	A	B		
	Ks211	/	/	/	/	D	D	/	C	/	C	C	A	B		
	Ssz2201	/	/	/	/	D	/	C	D	/	D	A	A	A	B	
	HO81c	/	/	/	/	D	D	C	/	D	D	A	A	A	B	
	FW111B1	/	/	/	/	D	D	B	D	C	A	A	A	A	B	
	YN072311	/	/	/	/	D	D	B	D	C	A	B	A	B		
	FW272A1	/	/	/	/	D	D	C	D	C	A	A	A	A	B	
	FW1N42	/	/	/	/	B	D	B	D	C	B	A	A	A	B	
	Nich274	/	/	/	/	B	D	C	B	/	B	C	A	A	B	
Nich232	/	/	/	/	B	D	C	B	/	C	C	A	A	B		
1106.2	D	/	/	/	B	D	/	B	C	A	A	A	A	B		
2303.1	D	D	/	/	/	D	/	B	C	A	A	A	A	B		
1801.4	/	/	/	/	B	D	C	D	C	C	A	A	A	D		
K84-01	/	/	/	/	D	D	/	D	C	C	A	C	D			
K88-24	/	/	/	/	D	/	C	D	C	C	A	C	D			
K88-07	D	/	/	/	D	D	C	D	D	C	A	C	D			
<i>Oryza sativa</i>	K91-13	/	/	/	/	D	D	C	D	C	C	A	C	D		
	K91-30	/	D	/	/	D	D	/	D	D	C	/	A	D		
	K96-07	/	/	/	/	D	D	/	B	C	C	A	A	D		
	K91-10	/	/	/	/	D	D	/	C	D	C	D	A	D		
	K93-16	/	/	/	/	D	D	C	D	C	C	A	A	D		
	K96-11	/	/	/	/	D	D	C	D	C	C	A	C	D		
	K98-10	D	/	/	/	D	D	/	D	C	C	A	C	D		
	401.4	/	/	/	/	B	D	C	/	C	C	A	C	D		
	913.4	/	/	/	/	D	D	C	D	C	C	A	A	D		
	4603.4	/	/	/	/	D	D	C	D	C	C	A	A	B		
	1836.3	/	/	/	/	D	D	C	/	C	C	A	C	D		
	4403.2	/	/	/	/	D	D	D	C	B	C	C	A	C	D	
	IS25	/	/	/	/	D	D	C	D	C	B	A	C	B		
	NX37	/	/	/	/	D	D	D	C	D	C	C	A	D		
	Z115	/	/	/	/	D	D	D	C	D	C	C	A	B		
AH06	/	/	/	/	D	D	D	C	B	C	C	A	B			
IL10	/	/	/	/	D	D	D	C	D	C	B	C	A	D		
PR72	/	/	/	/	D	D	D	C	D	C	C	A	A	D		
IT10	/	/	/	/	D	D	D	C	B	C	C	A	C	D		
SC05	/	/	/	/	D	D	D	C	D	C	C	C	C	D		
GD22	/	/	/	/	D	D	D	C	D	C	A	C	A	D		
GX01	/	/	/	/	D	D	D	C	B	C	C	A	A	D		
HN06	/	/	/	/	D	D	C	D	C	D	C	A	C	D		
HN19311	/	/	/	/	D	D	/	/	C	A	A	A	A	B		
JR1278	/	/	/	/	B	D	/	/	D	C	A	A	A	D		
HB12	/	/	/	/	D	D	D	C	D	C	C	C	C	D		
CQ11	/	/	/	/	D	D	D	C	D	C	D	C	A	D		
9157	/	/	/	/	/	D	/	/	C	A	A	A	A	B		
MG01	/	/	/	/	D	/	D	C	D	C	C	A	A	D		
MG10	/	D	/	/	/	D	/	C	B	D	C	A	A	B		
Mic-moi-38	/	/	/	/	/	/	/	/	C	D	C	A	A	D		
Y86910	/	/	/	/	/	/	/	/	C	D	D	A	A	D		

A No variation
B SNP
C InDel
D SNP+InDel
/ Absent

Figure 3. Highly polymorphic distribution of 11 cloned AVR genes in *M. oryzae* isolates from different hosts. The isolates with red color came from rice, and the gray from light to dark indicated wheat-, *Lolium*- and *Bromus*-infecting isolates. The other non-rice-infecting isolates were not filled with color.

When an isolate contains the detected *AVR* gene, we would distinguish and label the variant types of the gene with the cloned *AVR* genes. No variation, SNPs, InDel, and the combination of SNPs and InDel were marked as A, B, C, and D respectively. When describing the genotypes of *PWT3* or *PWT4* in different strains, A indicated that the sequences in this strain were consistent with those in 70-15 or Ina168 respectively. According to the sequence alignment results, the distribution patterns of the 11 *AVR* genes in RL and NRL could be divided into four types. First, as this gene was widely distributed in NRL and rarely appeared in RL, the distribution pattern of *AVR1-CO39* in isolates from different hosts was singled out. This was consistent with the previous results that almost all the pathogenic isolates of *M. oryzae* lack *AVR1-CO39*, so they showed strong pathogenicity to *CO39* rice cultivars [62]. Second, *AVR* genes only appeared in individual strains of *M. oryzae* from different hosts, including *AVR-Pii* and *PWL1*. Among them, *PWL1* was only distributed in four *Eleusine*-infecting isolates, while all rice isolates did not contain this gene. The third distribution pattern of *AVR* genes was contrary to frequency of *AVR1-CO39*, these genes appeared more frequently in RL than in NRL, including *AVR-Pia*, *AVR-Pik*, *AVR-Pita*, *PWT4*, *PWL2*, and *AVR-Pib*. In particular, *AVR-Pia*, *AVR-Pik*, and *AVR-Pita* were absent in all wheat-, *Lolium*-, and *Bromus*-infecting isolates. The last distribution pattern included the remaining four *AVR* genes of *PWT3*, *AVR-Piz-t*, *AVR-Pi9*, and *ACE1*. It could be seen that these four *AVR* genes were widely distributed in *M. oryzae* from different hosts except *Digitaria sanguinalis* and *Pennisetum americanum*. Compared with the sequences in RL, *PWT3* containing an insertion sequence of about 394 bp in NRL was basically consistent with the previous report [63]. Similar variations of *AVR-Piz-t* were observed among rice and non-rice isolates. *AVR-Piz-t* was conserved in 35 of 62 rice strains, while SNP and InDel mutations were found in all non-rice strains. Although the relationship between the distributions and variations of different *AVR* genes in different strains seemed to be complicated, the various genotypes of *Avr1-CO39*, *AvrPik* and *AvrPita* indicated their important roles in host differentiation of *M. oryzae*.

3.4. Gene Pool Analysis between RL and NRL

In order to show the selection pressure from various hosts on *M. oryzae* population structure at larger scale such as PAV, the gene pool was constructed with 48 *M. oryzae* genomes (Table S1). Among these genomes, wheat-infecting isolate BdBAr16-1 contained the most predicted genes (11,263), while rice-infecting isolate H081c contained the least predictive genes (10,422). The average numbers of predicted genes in RL and NRL were 10,578.7 and 10,779.3. The average lengths of predicted proteins in RL and NRL were 504 and 500.3 amino acids, respectively. By BLASTP, the predicted genes from 48 genomes were merged into the whole gene pool and 11,914 genes were obtained after removal of duplicate homologous genes (Figure 4a). Then, TBLASTN was performed to obtain the common genes shared in gene pool and every genome (Figure 4b). Our results showed that the number of shared genes in every member of NRL is more than the number of genes in different isolates of RL, except that in Py5020. Moreover, the average number of genes also showed significant differences between NRL (11,477.8) and RL (11,350.2).

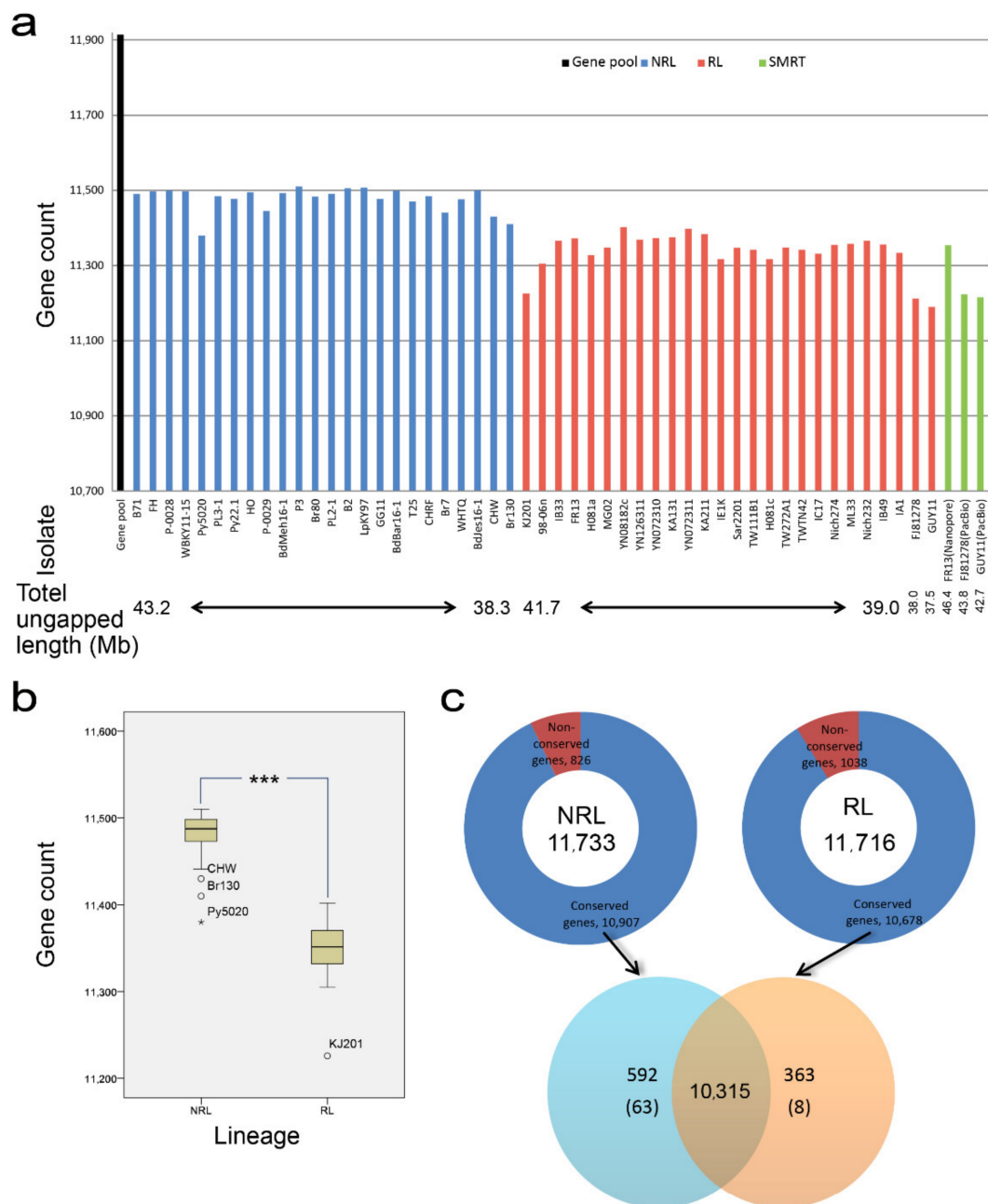


Figure 4. Gene pool analysis. (a) The distribution of the genes in gene pool shared in each isolate of RL and NRL. The black, blue, red, and green columns represent the number of genes in the gene pool, isolates of RL, isolates of NRL, and 3 rice-infecting genomes by SMRT, respectively. (b) Boxplot of the genes in RL and NRL. The symbol ★ indicated Py5020 in NRL with the extreme value of number of genes. However, Py5020 contained more genes than 87.5% rice-infecting isolates. The symbol ° represented the outliers CHW and Br130 in NRL, which also contained higher number of genes than all isolates of RL except YN08182c. ***, $p < 0.001$ by Student's t -test. (c) The numbers of homologous and non-homologous genes within the group, as well as common genes shared between groups and grouping-specific genes. The number in the intersection of the two solid circles represented the common genes shared in all 48 isolates. Combination 592(63) indicated that 592 genes were common in NRL but absent in at least one isolate of RL, and 63 genes only appeared in all isolates of NRL. 363(8) indicated that 363 genes were common in RL but absent in at least one isolate of NRL, and 8 genes only appeared in all isolates of RL.

To clarify the effects of different sequencing and assembly methods on gene pool results, the top 24 genomes with the total ungapped length of *M. oryzae* were selected in NCBI database (Figure 4a, Table S1). It could be seen that there was no relationship between the number of genes shared in the gene pool and genome size. Additional isolates FR13, FJ81278, and GUY11 were added because of their genomic data generated by both NGS (second-generation sequencing) and SMRT (single-molecule real-time sequencing), which can be used for further comparison and evaluation. Although the genome sizes of FR13, FJ81278, and GUY11 based on NGS were less than those generated by SMRT (6.6 Mb, 5.8 Mb and 5.4 Mb, respectively), the variance in number of genes generated by these two sequencing techniques were 18, 12, and 25 in FR13, FJ81278, and GUY11, respectively, which could be neglected compared with the average number of genes (11,350.2) in RL. Furthermore, Spearman correlation analysis was applied between the number of shared genes and the four genomic features used for evaluation of the genome assembly quality (Table 2 and Table S3). When considering the whole gene pool, our results showed that the number of shared genes contained in the genome was significantly correlated with the total length and total ungapped length of genome, which is independent with the number of sequences and N50. However, total length and total ungapped length were also significantly correlated with different hosts, and similar results were obtained by Mann Whitney U test. The number of sequences and N50 between the groups of different hosts were not significantly different. Moreover, there was no relationship between the number of shared genes contained in the genome and the four parameters when RL and NRL were regarded as separate groups. It was concluded that there were significant differences in the genome size and the number of genes between RL and NRL. This may also be related to the adaptation of rice-infecting isolate to hosts.

Table 2. Correlation analyses between four genomic features and the number of shared genes in each isolate and hosts.

Isolates	Features	Number of Genes		Hosts (Rice or Non-Rice)				
		Spearman ^a		Spearman ^b		Mann-Whitney U Test ^c		
		ρ	Sig	ρ	Sig	U	Z	Sig
48 isolates of RL and NRL	Total length	0.626 **	0.000	0.767 **	0.000	33.000	−5.258	0 **
	Total ungapped length	0.693 **	0.000	0.752 **	0.000	38.000	−5.155	0 **
	Scaffold (contig) number	−0.079	0.595	−0.244	0.095	207.000	−1.670	0.095
	N50	−0.081	0.585	−0.066	0.655	266.000	−0.454	0.650

^a The Spearman correlation analysis between four genomic features and the number of genes shared in isolates.

^b The Spearman correlation analysis between four genomic features and hosts of *M. oryzae* (rice or non-rice). ^c The Mann–Whitney U tests of four genomic features between isolates of RL and NRL. **, significant correlation at the 0.01 level in Spearman correlation analysis and $p < 0.01$ by Mann–Whitney U *t*-test, respectively.

In addition to the gene pool mentioned above, we also divided the 48 *Magnaporthe* isolates into two groups of RL and NRL. Our results showed that the number of genes in NRL (11,733) was similar to that in RL (11,716, Figure 4c). There were 10,678 conserved genes and 1038 non-conserved genes in RL, and 10,907 conserved genes and 826 non-conserved genes in NRL, respectively. While the number of conserved genes in RL is bigger than that in NRL, the number of non-conserved genes in RL is less than that in NRL. Among the gene pools containing 11,914 genes of all 48 isolates, 10,315 genes were found in the genomes of all isolates of RL and NRL, while 1599 PAV genes were absent in at least one isolate. We also compared the differences in gene pool analysis between RL and NRL with the same population size after BdJes16-1, BdMeh16-1, and BdBar16-1 removal, and similar results were obtained (Figure S10). The PAV dataset was used to screen candidate genes associated with host adaptation of *M. oryzae*.

3.5. Genes Associated with Host Adaptation of *M. oryzae* Identified by PAV Calling

Among 1599 specific genes, there were two genetic variation patterns, one was the genes only presented in one of RL and NRL but absented in another, the other was the genes with significantly different occurrence frequencies between RL and NRL. Fortunately, we found 8 and 63 genes (with the highest sequence identity less than 50% in the other group, Table S4) completely belonging to all isolates in RL and NRL, respectively. Maximum-likelihood methods were used to construct the polygenetic trees of 48 previous isolates and 6 extra isolates based on 1599 identified PAV genes. Among these 6 isolates, 2 *Stenotaphrum*-infecting isolates were regarded as an outgroup, 2 isolates of EL and 2 isolates of *Setaria* lineages (SL) were close to rice and wheat isolates, respectively. As showed in results, except for that P-0028 was closer to isolates of TL at this time, the clusters of 54 Magnaporthe isolates based on PAV data were consistent with those based on SNP data (Figure 5 and Figure S1) and previous studies [13]. In addition, the number of 8 grouping-specific genes of RL found in other lineages of *M. oryzae* decreased with the increase of genetic distance to the rice isolates. Similarly, the number of 63 grouping-specific genes of NRL found in EL, SL, and *Stenotaphrum* lineages decreased with the increase of genetic distance to TL and LL. However, a few of 71 grouping-specific genes were found in *G. graminis var. tritici* and *M. poae*, indicating that most grouping-specific genes occurred after the differentiation of *M. oryzae*, *G. graminis var. tritici*, and *M. poae*. BLASTP and TBLASTN searches against the NCBI non-redundant database revealed that 8 grouping-specific genes of RL were orphan genes for *M. oryzae* and 11 of 63 grouping-specific genes of NRL had homologs in other species, such as *Colletotrichum sp.* According to previous reports [12,31], the expressions of 25 grouping-specific genes of NRL were induced during the compatibility interaction between wheat and wheat blast (Table S5), and 7 grouping-specific genes of RL were regulated during the development and infection in *M. oryzae* (Table S6). Therefore, our results suggested that these 71 grouping-specific genes not only completely participated in host adaptability among various lineages of *M. oryzae*, but also played an important role in pathogenicity.

As for the screening of partial selection genes, the genes with higher occurrence frequencies in RL than those in NRL were regarded as favorable genes for RL and those in opposite situations were favorable genes for NRL. Our results showed that 60 and 168 favorable genes for RL and NRL were identified, respectively (Figure 6a,b, Table S4). Except for 71 complete selection genes, there were 52 and 105 favorable genes for RL and NRL respectively, indicating that these genes were undergoing the partial selection by different hosts.

To determine the putative function of 60 and 168 favorable genes for RL and NRL respectively, EggNOG database was used for GO annotation. However, there were only 14 and 31 favorable genes with GO annotation, indicating that the functions of most genes were unknown. These 45 annotation genes were classified according to previous reports (Table S7). We found that these genes were involved in various functions including recognition, cell wall modification, compound metabolism, transcript regulation, post-translational modification and so on. However, it is hard for us to enrich these genes with similar functions, the genes belonging to cell wall modification, transport, transcription factor, as well as compound metabolism might play an important role in host differentiation. Although almost all of these PAV genes do not have annotations, it could be testified by further experiments. Our analyses provided the vital information for genetic divergence and host adaption of pathogens, and these candidate genes could be the important candidates for preventing emergence of new virulence isolate in the future.

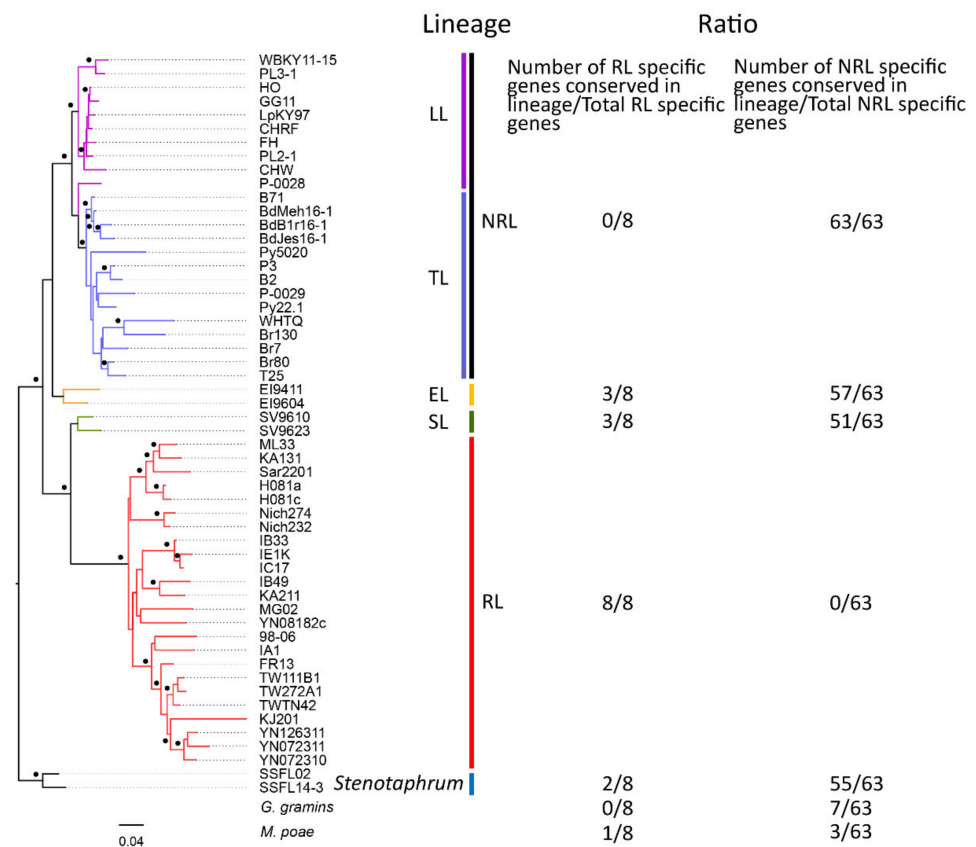


Figure 5. Maximum-likelihood phylogenetic tree, and occurrence ratio of RL and NRL specific genes in different lineages and two related species of Magnaporthaceae. The maximum-likelihood tree was constructed based on 1599 identified PAVs. Nodes with more than 90% bootstrap support were marked by dots. The isolates of different lineages were labeled with different colors according to the result of Figure S1. SL represented the *Setaria* lineage.

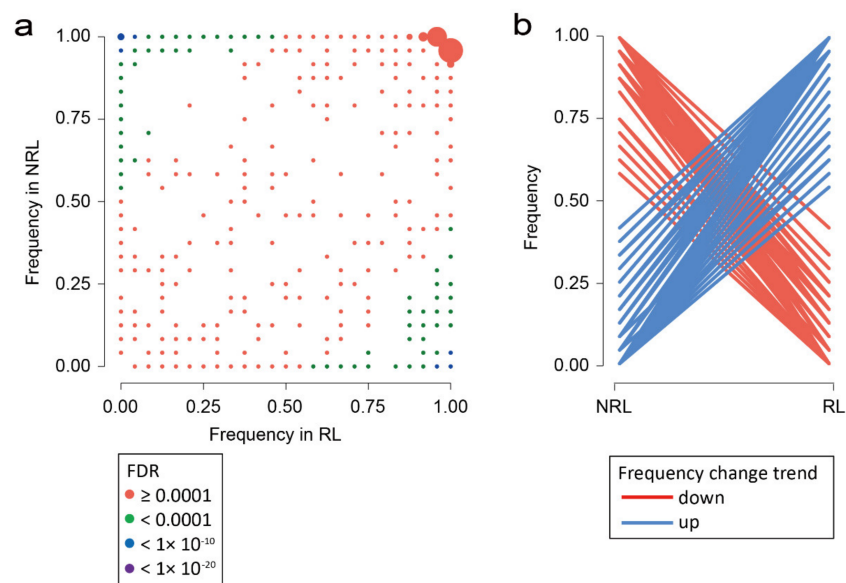


Figure 6. Gene selection preference by rice and non-rice hosts. (a) Scatter plots show occurrence frequencies of 1599 genes with PAV in RL and NRL. The different colors of the points represented the different value ranges of FDR, and the size of circle represented the number of genes with the same occurrence frequency. (b) Two kinds of occurrence frequency patterns of the putative selected genes by rice and non-rice host.

To determine the putative function of 60 and 168 favorable genes for RL and NRL, EggNOG database was used for GO annotation. However, there were only 14 and 31 favorable genes with GO annotation respectively, indicating that the functions of most genes were unknown. These 45 annotation genes were classified according to previous reports (Table S7). We found that these genes were involved in various functions including recognition, cell wall modification, compound metabolism, transcript regulation, post-translational modification, and so on. However, it is hard for us to enrich these genes with similar functions, the genes belonging to cell wall modification, transport, transcription factor, as well as compound metabolism might play an important role in host differentiation. Although almost all of these PAV genes do not have annotations, it could be testified by further experiments. Our analyses provided the vital information for genetic divergence and host adaptation of pathogens, and these candidate genes could be the important candidates for preventing emergence of new virulence isolate in the future.

4. Discussion

4.1. Comparative Genomes Analyses Present the Genome Regions Involving Host Selection between RL and NRL

For crop pathogens, the domestication of hosts and the established time of co-evolution between hosts and pathogens might determine the pathogen genetic diversity. Previous studies have showed that there is no evidence that multi-speed genome evolutions are found in different fungi of the Magnaporthaceae family [64], but the nucleotide diversity (π) of *M. oryzae* from different hosts is various. Rice with high domestication exerts high selection pressure on *M. oryzae*, which makes rice-infecting isolates with lower genetic diversity than that from other hosts. The isolates from *Setaria* have more than twice the π value of the RL in both the gene coding (synonymous) and non-coding regions [51]. When using single copy genes to calculate diversity, the wheat lineage was almost two times larger than rice lineage, while the *Lolium* and *Setaria* lineages were close to rice lineage [13]. In our studies, the genetic differences of 48 *Magnaporthe* isolates from rice, wheat, *Lolium*, and *Bromus* were analyzed by comparative genomics, which were collected from 12 different countries in America, Asia, Europe, and Africa from 1988 to 2016. The diversities of isolates from non-rice hosts including wheat, *Lolium*, and *Eleusine* were significantly higher than that of rice-infecting isolates, which was similar to previous results. However, our results presented more detailed information on the regions involved in host differentiation combining with F_{st} and Tajima's D calculations.

To our surprise, a large number of S-SNPs (119,637) were distributed in most windows (87.4%) which mainly led to 34.9% of windows with $F_{st} > 0.9$ and developed high degree of genetic differentiation between RL and NRL. However, such a large genetic differentiation with high value of F_{st} was rare in the comparison among other non-rice lineages (Figure S9a,b). Gladieux et al., pointed out that when the population divergence is older, more regions of divergence (with high F_{st}) will appear in the genome than similar regions (with low F_{st}) [65]. From the distribution of F_{st} , it could be seen that the divergence of ancestors of RL and NRL was older. Tajima's D is often used to distinguish the DNA sequences with random and non-random evolution [66]. When Tajima's $D > 0$, it can be used to infer bottleneck effect or balancing selection, and on the contrary, it refers to group size enlargement or directional selection. Through comparative analysis, the windows with Tajima's $D < 0$ in RL and NRL accounted for more than 70% and the large genetic difference between the two groups indicated that the selective sweep promoted the subdivision within a species [23–26]. While there were also significantly enriched windows with Tajima's $D > 0$ and higher diversity in NRL, it suggested that balancing selection not only significantly increased the diversity of NRL in these regions, but also reduced the F_{st} between RL and NRL. Therefore, two kinds of different signatures took place between RL and NRL in whole genomes. One is that both groups were strongly affected by the selective sweep in most chromosomal regions where a large number of S-SNPs and alleles were found. Both RL and NRL were developing toward the individual direction of homogenization which

led to strong genetic differentiation. The other one is that a few regions with high diversity appeared in NRL because of balancing selection, while RL retained homogenization in these regions. Similar results appeared in the population genetic statistical analysis of *Neurospora crassa* [67], and this study also showed that there are genes, which related to local photoperiod adaption of *Neurospora crassa* populations from different latitudes, in the genomic island similar to the second region with high diversity and Tajima's $D > 0$ mentioned above. By marking the location of the favorable genes related to host adaptability of RL and NRL in the reference genome, it can be seen that most of these genes were distributed in the region with high diversity and Tajima's $D > 0$ (Figure 1b).

Earlier studies have shown that balancing selection increases variability within a population which is an important mechanism to maintain population diversity [19,26]. However, directional selection like crop domestication reduces the diversity [28,68,69]. According to the pathotypes among different host infecting lineages, the rice-infecting isolates possess strict host specificity, while the isolates from wheat show broader hosts [70], implying balancing selection developing a lot of elasticity and variation in the wheat blast genome might be related with host range. Balance selection also provides more opportunities for wheat blast towards uncertain or risk direction. Moreover, much longer interaction history between rice and rice blast caused lower values of diversity and Tajima's D , indicating that rice-infecting isolates might undergo genome-wide selective sweeps to promote the subdivision of *M. oryzae*.

4.2. Gene Pool Analysis Deepen the PAV Genes Promoting Host Specification of Rice Infecting Isolates

As gene pool analysis can distinguish the core genes and specific genes in the group effectively and comprehensively, integrated PAVs can better explain genetic differentiation. For example, genome evolution across 1011 *Saccharomyces cerevisiae* isolates found that genome evolution in wild isolates is mainly driven by the accumulation of SNP with very low frequencies, while the copy-number changes have a greater phenotypic effect than single nucleotide polymorphisms [71,72]. For host jump, the pathogen needs various determinants such as effectors and virulence proteins to overcome new host resistance. After successful colonization was established, a strong selection pressure was exerted on the pathogen to improve its performance and adapt to the new host environment. Present genes might promote the adaption to new host while absent genes might be necessary for previous host or recognized by new host which should be abandoned. Therefore, PAV analysis might be used to explain large variations which are responsible for the occurrence of new phenotypes.

According to the previous study [73], large host jumps might be associated with marked gene loss. For successful host jump, losses of unnecessary genes help fungi better colonize in the new host. Our analysis revealed the average number of genes shared in each rice-infecting isolate was significantly less than that in non-rice-infecting isolate, and the number of conserved genes in RL was also smaller than that in NRL. Moreover, we also found that the decreased gene number among the 24 isolates in RL was independent of the quality of sequencing data. A similar result can be found in the variation of tomato genomes selected by constant domestication. The number of genes in the wild-type tomato genome is significantly higher than that in the domesticated tomato, while the number of genes in the improved tomato is the least [68]. By comparing the frequencies of 1599 PAV genes between RL and NRL, we obtained 60 and 168 favorable genes for RL and NRL during co-evolution with hosts respectively, a quarter of which were probably secreted proteins. Among these 228 genes selected by hosts, 8 genes were present in each isolate of RL and absent in all wheat-, *Lolium*-, and *Bromus*-infecting isolates, while 63 genes were only found in each isolate from wheat, *Lolium*, and *Bromus*. It is important to note that there were not more than 3 (37.5%) grouping-specific genes of RL conserved in other lineages of *M. oryzae*, such as the *Eleusine*, *Stenotaphrum*, and *Setaria* lineages which possessed no less than 51 (81.0%) grouping-specific genes of NRL. In order to better adapt to the host

and gain advantages in the arms race with the host, the genome of RL had undergone considerable simplification and specialization under stronger pressure, such as the loss of useless or unnecessary genes, while the selection of beneficial or specific genes was prudent and effective. Therefore, gene pool analysis suggested that, compared with the NRL, the RL experienced a higher level of natural selection through the gain or loss of gene.

4.3. Functions of the Genes Related to Host Adaptation

Gene pool analysis revealed 60 and 168 favorable genes for host adaption of RL and NRL respectively, but there were only 14 and 31 favorable genes with GO annotation, which indicated that the functions of a large number of genes need to be further identified. According to the annotated results and functional classification, some favorable genes for RL are likely concentrated on transport, while many favorable genes for NRL are enriched in cell wall modification, transcription factor, and compound metabolism. The studies related to the transport-related genes in *M. oryzae* which involve fungal development, xenobiotic degradation, and virulence have been reported. However, the functions of outer membrane auto transporter and importin are unknown in *M. oryzae*. According to previous results, importin family proteins, importin α and β play different roles in cargoes with nuclear localization signal transport from cytoplasm to nucleus. Importin α acts as an adaptor to bind importin β and cargo to enter into the nucleus. For hosts, different importin α isoforms govern the host adaptation of influenza virus [74]. Some importin-related genes involved in oxidative stress response, pathogenicity and transport of cargoes were identified [75]. For example, the β -importin KAP8 (Pse1/Kap121) is required for nuclear import of the cellulose transcriptional regulator XYR1, asexual sporulation, and stress resistance in *Trichoderma reesei* [76].

Compared with the situation of RL, more favorable genes for NRL were obtained and annotated. Gene families related to Glycosyl hydrolase (GH) have participated in cell wall modification [77]. Chitin oligosaccharides, as important constitutions of fungal cell wall, are the representative general elicitors to induce defense responses in rice. Recognition and modification for chitin become a co-evolutionary strategy between rice and *M. oryzae*. There are many chitin-related receptors such as CEBiP and OsCERK1 with LysM motifs to induce defense reaction in rice [78], while the effector Slp1 (secreted LysM protein 1) competes with CEBiP for the binding of chitin oligosaccharides [79]. Chitinase used for chitin degradation, which are regarded as members of GH family, can be recognized by rice [80,81], while a series of proteases involving chitinase cleavage also improves the fungal virulence [82]. Therefore, the modification of chitin and chitin-related enzymes is usually regarded as a strategy to escape recognition of rice. Moreover, the favorable gene encoding pectinesterase in our results is also involved in host specialization of *Epichloë* [83], indicating that the genes related to cell wall play a universal role in host differentiation among fungi. We also found a favorable gene for NRL encoding conserved glycine-rich protein (GRP), and previous report identified GRP1 as a splicing factor to regulate post-transcription and fungal virulence and growth in *M. oryzae* [84]. Furthermore, two avirulence effectors *PWL1* and *PWL2* containing glycine-rich and hydrophilic domains prevent the rice-infecting isolates from infecting weeping lovegrass with host specificity, indicating that conserved glycine-rich proteins might be recognized by the host [53,54].

Among the favorable genes for NRL, two of which encoding heterokaryon incompatibility protein (HET) were screened. Heterokaryon incompatibility in filamentous fungi is regarded as a rejection response for genetically different entities. Heterokaryon incompatibility is controlled by het genes, whose incompatible alleles co-expressed in a common cytoplasm initiate cell death [85]. Coincidentally, two genes containing caspase structural domain were obtained in the favorable genes for NRL, and the induction of caspase can be used as the biomarker of programmed cell death [86]. Compared with favorable genes of RL, these genes mentioned above might be associated with the adaption to host species.

4.4. The Gain or Loss of AVR and Effector Affects the Host Differentiation of *M. oryzae*

Many genes of *M. oryzae* are under positive selection and most of them are unique sequences compared with other *Sordariomycetes*. Moreover, in order to gain an advantage in the interaction between host and pathogen, *M. oryzae* contains more secretory proteins and effectors than other *Sordariomycetes* [87]. AVR genes have been reported to involve in recognition for different hosts by *M. oryzae*. For instance, polymorphisms of *Pik* and *AVR-Pik* alleles suggest that the mutant of AVR genes drives the fast adaption of *M. oryzae* to new resistant varieties [88]. In addition, *AVR-Pia* has been identified as a possible determinant of the specialization of *M. oryzae* to local japonica rice in rice production area of Yuanyang county [89].

PWL2 identified from weeping lovegrass participates in host jump of *M. oryzae* among different crop hosts [54]. In this study, our results showed that the presence or absence variations of *AVR1-CO39*, *AVR-Pita*, and *AVR-Pik* involved host differentiation, especially between rice blast and wheat blast. However, the relationship between AVR genes and host adaption seems not strict due to different genomes used for analysis [90]. As expected, the effectors showed the richest diversity in all the genes of *M. oryzae* [51] and many PAV genes with signal peptide were identified from our analysis, which indicated that there might be potential effectors responsible for host recognition.

For hosts, the cognate *R* genes are considered as the effective defense weapon to pathogens. Almost all of identified *R* genes encoding NBS–LRR proteins seem conserved in animals and plants [91]. Moreover, some *R* genes from maize, sorghum, and brachypodium confer the resistance to rice blast, indicating that the abundant rice blast *R* genes occur not only within species but also among species [92]. On the one hand, the host jump means the pathogens possess enough resource to overcome the resistance from new hosts. On the other hand, the hosts might not be ready or have no ability to cope with novel pathogens. Domestication of crops toward preference traits often loses some genetic resources which increase the risk to face unknown challenges for hosts. When the wheat blast disease has not been reported, it is difficult to determine whether the *R* genes against *M. oryzae* might be lost in wheat due to neglect of potential value and domestication of wheat. Therefore, the loss of genetic diversity in hosts should be considered as an alert for host jump.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/jof8010005/s1>. Figure S1: Phylogenetic tree of 48 RL, NRL isolates, 2 EL isolates and 2 SL isolates based on genome-wide SNP data, Figure S2: Statistical distribution of π and Tajima's *D* of 4045 Calculation window in RL and NRL, Figure S3: F_{st} , π and Tajima's *D* for 21 isolates of RL and 21 isolates of NRL (without BdJes16-1, BdMeh16-1 and BdBar16-1), Figure S4: Statistical distribution of π and Tajima's *D* of 4045 calculation window in RL(21) and NRL(21), and the 21 RL isolates used in this section, Figure S5: F_{st} , π and Tajima's *D* for RL(14) and TL, respectively, Figure S6: Statistical distribution of π and Tajima's *D* of 4045 calculation window in RL(14) and TL, and the 14 RL isolates used in this section, Figure S7: F_{st} , π and Tajima's *D* for RL(10) and LL, respectively, Figure S8: Statistical distribution of π and Tajima's *D* of 4045 calculation window in RL(10) and LL, and the 10 RL isolates used in this section, Figure S9: Statistics of F_{st} and π in comparison between different lineages, Figure S10: Gene pool analysis of 21 RL isolates and NRL without BdJes16-1, BdMeh16-1 and BdBar16-1, Table S1: *M. oryzae* isolates used in this study, Table S2: Statistics of SNPs number scanned by comparing 48 isolates with different reference genomes, Table S3: Correlation analyses between four genomic features and the number of shared genes in RL and NRL, Table S4: Genes detected in the 48 isolates with PAV calling and significantly different frequencies between RL and NRL, Table S5: Expression of 25 NRL Specific genes in field samples collected during wheat blast epidemic in Bangladesh 2016, Table S6: Expression of 7 RL Specific genes during the compatible interaction between 98-06 and rice plants (cv. CO-39), Table S7: Classification of genes related to host adaptation for *M. oryzae* with annotation.

Author Contributions: Conceptualization, Q.W. and Y.W.; methodology, Q.W.; software, Q.W. and L.-N.L.; validation, Q.W. and Y.W.; formal analysis, Q.W.; investigation, Q.W. and Y.W.; resources, C.-Y.L.; data curation, Q.W. and L.-N.L.; writing—original draft preparation, Q.W. and Y.W.; writing—review and editing, C.-Y.L. and K.S.; visualization, Q.W.; supervision, C.-Y.L.; project administration, C.-Y.L.; funding acquisition, C.-Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key R&D Program of China with grant number 2017YFD0200400, the Yunling Scholar Training Project, Foundation of State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan with grant number GZJS201903, the Basic Research Special Project-Youth Project in Yunnan with grant number 202101AU070101, the Yunnan Province Major Science and Technology Project Sub-project with grant number 202002AE09001004, the Science Research Foundation of Yunnan Education Bureau with grant number 2021J0138 and the Open Project of State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in this article and its Supplementary Materials.

Acknowledgments: Special thanks to Hou-Xiang Kang (Institute of Plant Protection, Chinese Academy of Agricultural Sciences) for his help and support in data analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Kato, H.; Yamamoto, M.; Yamaguchi-Ozaki, T.; Kadouchi, H.; Iwamoto, Y.; Nakayashiki, H.; Tosa, Y.; Mori, S.M. Pathogenicity, mating ability and DNA restriction fragment length polymorphisms of *Pyricularia* populations isolated from Gramineae, Bambusidae and Zingiberaceae plants. *J. Gen. Plant Pathol.* **2000**, *66*, 30–47. [[CrossRef](#)]
- Couch, B.C.; Kohn, L.M. A multilocus gene genealogy concordant with host preference indicates segregation of a new species, *Magnaporthe oryzae*, from *M. grisea*. *Mycologia* **2002**, *94*, 683–693. [[CrossRef](#)] [[PubMed](#)]
- Couch, B.C.; Fudal, I.; Lebrun, M.H.; Tharreau, D.; Valent, B.; van Kim, P.; Notteghem, J.L.; Kohn, L.M. 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*, 613–630. [[CrossRef](#)]
- Murakami, J.; Tosa, Y.; Kataoka, T.; Tomita, R.; Kawasaki, J.; Chuma, I.; Sesumi, Y.; Kusaba, M.; Nakayashiki, H.; Mayama, S. Analysis of host species specificity of *Magnaporthe grisea* toward wheat using a genetic cross between isolates from wheat and foxtail millet. *Phytopathology* **2000**, *90*, 1060–1067. [[CrossRef](#)] [[PubMed](#)]
- Oh, H.S.; Tosa, Y.; Takabayashi, N.; Nakagawa, S.; Tomita, R.; Don, L.D.; Kusaba, M.; Nakayashiki, H.; Mayama, S. Characterization of an *Avena* isolate of *Magnaporthe grisea* and identification of a locus conditioning its specificity on oat. *Can. J. Bot.* **2002**, *80*, 1088–1095. [[CrossRef](#)]
- Tosa, Y.; Hirata, K.; Tamba, H.; Nakagawa, S.; Chuma, I.; Isobe, C.; Osue, J.; Urashima, A.S.; Don, L.D.; Kusaba, M.; et al. Genetic constitution and pathogenicity of *Lolium* isolates of *Magnaporthe oryzae* in comparison with host species-specific pathotypes of the blast fungus. *Phytopathology* **2004**, *94*, 454–462. [[CrossRef](#)] [[PubMed](#)]
- Ou, S.H. *Rice Diseases*; Kew Commonwealth Mycological Institute: Kew, UK, 1985.
- Maciel, J.L.; Ceresini, P.C.; Castroagudin, V.L.; Zala, M.; Kema, G.H.; McDonald, B.A. Population structure and pathotype diversity of the wheat blast pathogen *Magnaporthe oryzae* 25 years after its emergence in Brazil. *Phytopathology* **2014**, *104*, 95–107. [[CrossRef](#)]
- Cruz, C.D.; Valent, B. Wheat blast disease: Danger on the move. *Trop. Plant Pathol.* **2017**, *42*, 210–222. [[CrossRef](#)]
- Farman, M.; Peterson, G.; Chen, L.; Starnes, J.; Valent, B.; Bachi, P.; Murdock, L.; Hershman, D.; Pedley, K.; Fernandes, J.M.; et al. The *Lolium* pathotype of *Magnaporthe oryzae* recovered from a single blasted wheat plant in the United States. *Plant Dis.* **2017**, *101*, 684–692. [[CrossRef](#)]
- Malaker, P.K.; Barma, N.; Tewari, T.P.; Collis, W.J.; Valent, B. First report of wheat blast caused by *Magnaporthe oryzae* pathotype *tritium* in Bangladesh. *Plant Dis.* **2016**, *100*, 2330. [[CrossRef](#)]
- Islam, M.T.; Croll, D.; Gladieux, P.; Soanes, D.M.; Persoons, A.; Bhattacharjee, P.; Hossain, M.S.; Gupta, D.R.; Rahman, M.M.; Mahboob, M.G.; et al. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biol.* **2016**, *14*, 84. [[CrossRef](#)] [[PubMed](#)]
- Gladieux, P.; Condon, A.B.; Ravel, B.S.; Soanes, A.D.; Maciel, C.J.L.N. Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. *mBio* **2018**, *9*, e01219-17. [[CrossRef](#)] [[PubMed](#)]

14. 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*, 1440–1456. [[CrossRef](#)] [[PubMed](#)]
15. Gladieux, P.; Ravel, S.; Rieux, A.; Cros-Arteil, S.; Adreit, H.; Milazzo, J.; Thierry, M.; Fournier, E.; Terauchi, R.; Tharreau, D. Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. *mBio* **2018**, *9*, e01806–e01817. [[CrossRef](#)] [[PubMed](#)]
16. Zhong, Z.; Norvinyeku, J.; Chen, M.; Bao, J.; Lin, L.; Chen, L.; Lin, Y.; Wu, X.; Cai, Z.; Zhang, Q.; et al. Directional selection from host plants is a major force driving host specificity in *Magnaporthe* species. *Sci. Rep.* **2016**, *6*, 25591. [[CrossRef](#)] [[PubMed](#)]
17. Smith, J.M.; Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **1974**, *23*, 23–35. [[CrossRef](#)] [[PubMed](#)]
18. Anisimova, M.; Liberles, D. Detecting and understanding natural selection. In *Codon Evolution: Mechanisms and Models*; Cannarozzi, G.M., Schneider, A., Eds.; Oxford University Press: Oxford, UK, 2012.
19. Kojima, K.-I. Is there a constant fitness value for a given genotype? NO! *Evolution* **1971**, *25*, 281–285. [[CrossRef](#)]
20. Rech, G.E.; Sanz-Martin, J.M.; Anisimova, M.; Sukno, S.A.; Thon, M.R. Natural selection on coding and noncoding DNA sequences is associated with virulence genes in a plant pathogenic fungus. *Genome Biol. Evol.* **2014**, *6*, 2368–2379. [[CrossRef](#)]
21. Ye, W.; Wang, Y.; Tyler, B.M.; Wang, Y. Comparative genomic analysis among four representative isolates of *Phytophthora sojae* reveals genes under evolutionary selection. *Front. Microbiol.* **2016**, *7*, 1547. [[CrossRef](#)] [[PubMed](#)]
22. Ma, W.; Dong, F.; Stavrinides, J.; Guttman, D.S. Type III effector diversification via both pathoadaptation and horizontal transfer in response to a coevolutionary arms race. *PLoS Genet.* **2007**, *2*, e209. [[CrossRef](#)]
23. Charlesworth, B.; Nordborg, M.; Charlesworth, D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **1997**, *70*, 155–174. [[CrossRef](#)]
24. Majewski, J.; Cohan, F.M. Adapt globally, act locally: The effect of selective sweeps on bacterial sequence diversity. *Genetics* **1999**, *152*, 1459–1474. [[CrossRef](#)] [[PubMed](#)]
25. Slatkin, M.; Wiehe, T. Genetic hitch-hiking in a subdivided population. *Genet. Res.* **1998**, *71*, 155–160. [[CrossRef](#)] [[PubMed](#)]
26. Nielsen, R. Molecular signatures of natural selection. *Annu. Rev. Genet.* **2005**, *39*, 197–218. [[CrossRef](#)] [[PubMed](#)]
27. Mohd-Assaad, N.; McDonald, B.A.; Croll, D. Genome-wide detection of genes under positive selection in worldwide populations of the barley scald pathogen. *Genome Biol. Evol.* **2018**, *10*, 1315–1332. [[CrossRef](#)] [[PubMed](#)]
28. Huang, X.; Kurata, N.; Wei, X.; Wang, Z.X.; Wang, A.; Zhao, Q.; Zhao, Y.; Liu, K.; Lu, H.; Li, W.; et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **2012**, *490*, 497–501. [[CrossRef](#)]
29. Pieck, M.L.; Ruck, A.; Farman, M.L.; Peterson, G.L.; Stack, J.P.; Valent, B.; Pedley, K.F. Genomics-based marker discovery and diagnostic assay development for wheat blast. *Plant Dis.* **2017**, *101*, 103–109. [[CrossRef](#)]
30. Park, S.; Kim, S.; Koh, H.; Lee, Y. A blast lesion mimic mutant of rice. In *Advances in Rice Blast*; Tharreau, D., Lebrun, M.H., Talbot, N.J., Nottoghem, J.L., Eds.; Kluwer Academic Press: Dordrecht, The Netherlands, 2000.
31. Dong, Y.; Li, Y.; Zhao, M.; Jing, M.; Liu, X.; Liu, M.; Guo, X.; Zhang, X.; Chen, Y.; Liu, Y.; 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. [[CrossRef](#)] [[PubMed](#)]
32. Chiapello, H.; Mallet, L.; Guerin, C.; Aguileta, G.; Amselem, J.; Kroj, T.; Ortega-Abboud, E.; Lebrun, M.H.; Henrissat, B.; Gendrault, A.; et al. Deciphering genome content and evolutionary relationships of isolates from the fungus *Magnaporthe oryzae* attacking different host plants. *Genome Biol. Evol.* **2015**, *7*, 2896–2912. [[CrossRef](#)] [[PubMed](#)]
33. Wang, Z.Q.; Meng, F.Z.; Zhang, M.M.; Yin, L.F.; Yin, W.X.; Lin, Y.; Hsiang, T.; Peng, Y.L.; Wang, Z.H.; Luo, C.X. A putative Zn2Cys6 transcription factor is associated with isoprothiolane resistance in *Magnaporthe oryzae*. *Front. Microbiol.* **2018**, *9*, 2608. [[CrossRef](#)] [[PubMed](#)]
34. Zhong, Z.; Chen, M.; Lin, L.; Han, Y.; Bao, J.; Tang, W.; Lin, L.; Lin, Y.; Somai, R.; Lu, L.; et al. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. *ISME J.* **2018**, *12*, 1867–1878. [[CrossRef](#)] [[PubMed](#)]
35. Shirke, M.D.; Mahesh, H.B.; Gowda, M. Genome-wide comparison of *Magnaporthe* species reveals a host-specific pattern of secretory proteins and transposable elements. *PLoS ONE* **2016**, *11*, e0162458. [[CrossRef](#)]
36. Chen, C.; Lian, B.; Hu, J.; Zhai, H.; Wang, X.; Venu, R.C.; Liu, E.; Wang, Z.; Chen, M.; Wang, B.; et al. Genome comparison of two *Magnaporthe oryzae* field isolates reveals genome variations and potential virulence effectors. *BMC Genom.* **2013**, *14*, 887. [[CrossRef](#)] [[PubMed](#)]
37. Ray, S.; Singh, P.K.; Gupta, D.K.; Mahato, A.K.; Sarkar, C.; Rathour, R.; Singh, N.K.; Sharma, T.R. Analysis of *Magnaporthe oryzae* genome reveals a fungal effector, which is able to induce resistance response in transgenic rice line containing resistance gene, *Pi54*. *Front. Plant Sci.* **2016**, *7*, 1140. [[CrossRef](#)]
38. Zhu, K.; Bao, J.; Zhang, L.; Yang, X.; Li, Y.; Zhu, M.; Lin, Q.; Zhao, Z.; Zhou, B.; Lu, G. Comparative analysis of the genome of the field isolate V86010 of the rice blast fungus *Magnaporthe oryzae* from Philippines. *J. Integr. Agric.* **2017**, *16*, 2222–2230. [[CrossRef](#)]
39. Gowda, M.; Shirke, M.D.; Mahesh, H.B.; Chandarana, P.; Rajamani, A.; Chattoo, B.B. Genome analysis of rice-blast fungus *Magnaporthe oryzae* field isolates from southern India. *Genome Data* **2015**, *5*, 284–291. [[CrossRef](#)]
40. Bao, J.; Chen, M.; Zhong, Z.; Tang, W.; Lin, L.; Zhang, X.; Jiang, H.; Zhang, D.; Miao, C.; Tang, H.; et al. PacBio sequencing reveals transposable elements as a key contributor to genomic plasticity and virulence variation in *Magnaporthe oryzae*. *Mol. Plant* **2017**, *10*, 1465–1468. [[CrossRef](#)] [[PubMed](#)]

41. Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S.L. Versatile and open software for comparing large genomes. *Genome Biol.* **2004**, *5*, R12. [[CrossRef](#)] [[PubMed](#)]
42. Vilella, A.J.; Blanco-Garcia, A.; Hutter, S.; Rozas, J. VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics* **2005**, *21*, 2791–2793. [[CrossRef](#)] [[PubMed](#)]
43. Rousset, F. genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **2008**, *8*, 103–106. [[CrossRef](#)] [[PubMed](#)]
44. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)] [[PubMed](#)]
45. Keller, O.; Kollmar, M.; Stanke, M.; Waack, S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **2011**, *27*, 757–763. [[CrossRef](#)] [[PubMed](#)]
46. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, D.R.; Sunagawa, S.; Kuhn, M.; et al. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **2016**, *44*, D286–D293. [[CrossRef](#)]
47. Bendtsen, J.D.; Nielsen, H.; von Heijne, G.; Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **2004**, *340*, 783–795. [[CrossRef](#)] [[PubMed](#)]
48. Xue, M.; Yang, J.; Li, Z.; Hu, S.; Yao, N.; Dean, R.A.; Zhao, W.; Shen, M.; Zhang, H.; Li, C.; et al. Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. *PLoS Genet.* **2012**, *8*, e1002869. [[CrossRef](#)] [[PubMed](#)]
49. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **2009**, *26*, 1641–1650. [[CrossRef](#)] [[PubMed](#)]
50. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)] [[PubMed](#)]
51. Yoshida, K.; Saunders, D.G.; Mitsuoka, C.; Natsume, S.; Kosugi, S.; Saitoh, H.; Inoue, Y.; Chuma, I.; Tosa, Y.; Cano, L.M.; et al. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genom.* **2016**, *17*, 370. [[CrossRef](#)]
52. Wright, S. *Evolution and the Genetics of Populations: Experimental Results and Evolutionary Deductions*; University of Chicago Press: Chicago, IL, USA, 1977; Volume 3.
53. Kang, S.; Sweigard, J.A.; Valent, B. The *PWL* host specificity gene family in the blast fungus *Magnaporthe grisea*. *Mol. Plant Microbe Interact* **1995**, *8*, 939–948. [[CrossRef](#)] [[PubMed](#)]
54. Sweigard, J.A.; Carroll, A.M.; Kang, S.; Farrall, L.; Chumley, F.G.; Valent, B. Identification, cloning, and characterization of *PWL2*, a gene for host species specificity in the rice blast fungus. *Plant Cell* **1995**, *7*, 1221–1233. [[CrossRef](#)] [[PubMed](#)]
55. Orbach, M.J.; Farrall, L.; Sweigard, J.A.; Chumley, F.G.; Valent, B. A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. *Plant Cell* **2000**, *12*, 2019–2032. [[CrossRef](#)] [[PubMed](#)]
56. Bohnert, H.U.; Fudal, I.; Dioh, W.; Tharreau, D.; Notteghem, J.L.; Lebrun, M.H. A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. *Plant Cell* **2004**, *16*, 2499–2513. [[CrossRef](#)] [[PubMed](#)]
57. Yoshida, K.; Saitoh, H.; Fujisawa, S.; Kanzaki, H.; Matsumura, H.; Yoshida, K.; Tosa, Y.; Chuma, I.; Takano, Y.; Win, J.; et al. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* **2009**, *21*, 1573–1591. [[CrossRef](#)] [[PubMed](#)]
58. Li, W.; Wang, B.; Wu, J.; Lu, G.; Hu, Y.; Zhang, X.; Zhang, Z.; Zhao, Q.; Feng, Q.; Zhang, H.; 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*, 411–420. [[CrossRef](#)]
59. Zheng, Y.; Zheng, W.; Lin, F.; Zhang, Y.; Yi, Y.; Wang, B.; Lu, G.; Wang, Z.; Wu, W. *AVR1-CO39* is a predominant locus governing the broad avirulence of *Magnaporthe oryzae* 2539 on cultivated rice (*Oryza sativa* L.). *Mol. Plant Microbe Interact.* **2011**, *24*, 13–17. [[CrossRef](#)]
60. Zhang, S.; Wang, L.; Wu, W.; He, L.; Yang, X.; Pan, Q. Function and evolution of *Magnaporthe oryzae* avirulence gene *AvrPib* responding to the rice blast resistance gene *Pib*. *Sci. Rep.* **2015**, *5*, 11642. [[CrossRef](#)]
61. Wu, J.; Kou, Y.; Bao, J.; Li, Y.; Tang, M.; Zhu, X.; Ponaya, A.; Xiao, G.; Li, J.; Li, C.; et al. Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector *AvrPi9* that triggers *Pi9*-mediated blast resistance in rice. *New Phytol.* **2015**, *206*, 1463–1475. [[CrossRef](#)] [[PubMed](#)]
62. Farman, M.L.; Eto, Y.; Nakao, T.; Tosa, Y.; Nakayashiki, H.; Mayama, S.; Leong, S.A. 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. [[CrossRef](#)] [[PubMed](#)]
63. Inoue, Y.; Vy, T.T.P.; Yoshida, K.; Asano, H.; Mitsuoka, C.; Asume, S.; Anh, V.L.; Cumagun, C.J.R.; Chuma, I.; Terauchi, R.; et al. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. *Science* **2017**, *357*, 80–83. [[CrossRef](#)] [[PubMed](#)]
64. Okagaki, L.H.; Sailsbery, J.K.; Eyre, A.W.; Dean, R.A. Comparative genome analysis and genome evolution of members of the magnaporthaceae family of fungi. *BMC Genom.* **2016**, *17*, 135. [[CrossRef](#)] [[PubMed](#)]
65. Taylor, J.W.; Branco, S.; Gao, C.; Hann-Soden, C.; Montoya, L.; Sylvain, I.; Gladieux, P. Sources of fungal genetic variation and associating it with phenotypic diversity. *Microbiol. Spectr.* **2017**, *5*. [[CrossRef](#)] [[PubMed](#)]

66. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**, *123*, 585–595. [[CrossRef](#)] [[PubMed](#)]
67. Ellison, C.E.; Hall, C.; Kowbel, D.; Welch, J.; Brem, R.B.; Glass, N.L.; Taylor, J.W. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2831–2836. [[CrossRef](#)] [[PubMed](#)]
68. Gao, L.; Gonda, I.; Sun, H.; Ma, Q.; Bao, K.; Tieman, D.M.; Burzynski-Chang, E.A.; Fish, T.L.; Stromberg, K.A.; Sacks, G.L.; et al. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* **2019**, *51*, 1044–1051. [[CrossRef](#)]
69. Zhao, J.; Sauvage, C.; Zhao, J.; Bitton, F.; Bauchet, G.; Liu, D.; Huang, S.; Tieman, D.M.; Klee, H.J.; Causse, M. Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavor. *Nat. Commun.* **2019**, *10*, 1534. [[CrossRef](#)]
70. Ceresini, P.C.; Castroagudin, V.L.; Rodrigues, F.A.; Rios, J.A.; Eduardo Aucique-Perez, C.; Moreira, S.I.; Alves, E.; Croll, D.; Maciel, J.L.N. Wheat blast: Past, present, and future. *Annu. Rev. Phytopathol.* **2018**, *56*, 427–456. [[CrossRef](#)]
71. Peter, J.; De Chiara, M.; Friedrich, A.; Yue, J.X.; Pflieger, D.; Bergstrom, A.; Sigwalt, A.; Barre, B.; Freel, K.; Llored, A.; et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **2018**, *556*, 339–344. [[CrossRef](#)]
72. Dunn, B.; Richter, C.; Kvitek, D.J.; Pugh, T.; Sherlock, G. Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res.* **2012**, *22*, 908–924. [[CrossRef](#)]
73. Sharma, R.; Mishra, B.; Runge, F.; Thines, M. Gene loss rather than gene gain is associated with a host jump from monocots to dicots in the Smut Fungus *Melanopsichium pennsylvanicum*. *Genome Biol. Evol.* **2014**, *6*, 2034–2049. [[CrossRef](#)]
74. Gabriel, G.; Klingel, K.; Otte, A.; Thiele, S.; Hudjetz, B.; Arman-Kalceck, G.; Sauter, M.; Shmidt, T.; Rother, F.; Baumgarte, S.; et al. Differential use of importin- α isoforms governs cell tropism and host adaptation of influenza virus. *Nat. Commun.* **2011**, *2*, 156. [[CrossRef](#)]
75. Yang, X.; Ding, F.; Zhang, L.; Sheng, Y.; Zheng, X.; Wang, Y. The importin α subunit PsIMPA1 mediates the oxidative stress response and is required for the pathogenicity of *Phytophthora sojae*. *Fungal Genet. Biol.* **2015**, *82*, 108–115. [[CrossRef](#)]
76. Ghassemi, S.; Lichius, A.; Bidard, F.; Lemoine, S.; Rossignol, M.N.; Herold, S.; Seidl-Seiboth, V.; Seiboth, B.; Espeso, E.A.; Margeot, A.; et al. The ss-importin KAP8 (Pse1/Kap121) is required for nuclear import of the cellulase transcriptional regulator XYR1, asexual sporulation and stress resistance in *Trichoderma reesei*. *Mol. Microbiol.* **2015**, *96*, 405–418. [[CrossRef](#)] [[PubMed](#)]
77. Van Vu, B.; Itoh, K.; Nguyen, Q.B.; Tosa, Y.; Nakayashiki, H. Cellulases belonging to glycoside hydrolase families 6 and 7 contribute to the virulence of *Magnaporthe oryzae*. *Mol. Plant Microbe Interact.* **2012**, *25*, 1135–1141. [[CrossRef](#)]
78. Shimizu, T.; Nakano, T.; Takamizawa, D.; Desaki, Y.; Ishii-Minami, N.; Nishizawa, Y.; Minami, E.; Okada, K.; Yamane, H.; Kaku, H.; et al. Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. *Plant J.* **2010**, *64*, 204–214. [[CrossRef](#)] [[PubMed](#)]
79. Mentlak, T.A.; Kombrink, A.; Shinya, T.; Ryder, L.S.; Otomo, I.; Saitoh, H.; Terauchi, R.; Nishizawa, Y.; Shibuya, N.; Thomma, B.P.; et al. Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell* **2012**, *24*, 322–335. [[CrossRef](#)] [[PubMed](#)]
80. Yang, C.; Yu, Y.; Huang, J.; Meng, F.; Pang, J.; Zhao, Q.; Islam, M.A.; Xu, N.; Tian, Y.; Liu, J. Binding of the *Magnaporthe oryzae* chitinase MoChia1 by a rice tetratricopeptide repeat protein allows free chitin to trigger immune responses. *Plant Cell* **2019**, *31*, 172–188. [[CrossRef](#)] [[PubMed](#)]
81. Han, Y.; Song, L.; Peng, C.; Liu, X.; Liu, L.; Zhang, Y.; Wang, W.; Zhou, J.; Wang, S.; Ebbolle, D.; et al. A *Magnaporthe* chitinase interacts with a rice jacalin-related lectin to promote host colonization. *Plant Physiol.* **2019**, *179*, 1416–1430. [[CrossRef](#)]
82. Jashni, M.K.; Mehrabi, R.; Collemare, J.; Mesarich, C.H.; de Wit, P.J. The battle in the apoplast: Further insights into the roles of proteases and their inhibitors in plant-pathogen interactions. *Front Plant Sci.* **2015**, *6*, 584. [[CrossRef](#)]
83. Schirrmann, M.K.; Zoller, S.; Croll, D.; Stukenbrock, E.H.; Leuchtman, A.; Fior, S. Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Mol. Ecol.* **2018**, *27*, 3070–3086. [[CrossRef](#)]
84. Gao, X.; Yin, C.; Liu, X.; Peng, J.; Chen, D.; He, D.; Shi, W.; Zhao, W.; Yang, J.; Peng, Y.L. A glycine-rich protein MoGrp1 functions as a novel splicing factor to regulate fungal virulence and growth in *Magnaporthe oryzae*. *Phytopathol. Res.* **2019**, *1*, 2. [[CrossRef](#)]
85. Paoletti, M. Vegetative incompatibility in fungi: From recognition to cell death, whatever does the trick. *Fungal Biol. Rev.* **2016**, *30*, 152–162. [[CrossRef](#)]
86. Kumar, S. Caspase function in programmed cell death. *Cell Death Differ.* **2007**, *14*, 32–43. [[CrossRef](#)]
87. Zhang, N.; Cai, G.; Price, D.C.; Crouch, J.A.; Gladieux, P.; Hillman, B.; Khang, C.H.; LeBrun, M.H.; Lee, Y.H.; Luo, J.; et al. Genome wide analysis of the transition to pathogenic lifestyles in Magnaporthales fungi. *Sci. Rep.* **2018**, *8*, 5862. [[CrossRef](#)]
88. Kanzaki, H.; Yoshida, K.; Saitoh, H.; Fujisaki, K.; Hirabuchi, A.; Alaux, L.; Fournier, E.; Tharreau, D.; Terauchi, R. Arms race co-evolution of *Magnaporthe oryzae* AVR-Pik and rice Pik genes driven by their physical interactions. *Plant J.* **2012**, *72*, 894–907. [[CrossRef](#)] [[PubMed](#)]
89. Liao, J.; Huang, H.; Meusnier, I.; Adreit, H.; Ducasse, A.; Bonnot, F.; Pan, L.; He, X.; Kroj, T.; Fournier, E.; et al. Pathogen effectors and plant immunity determine specialization of the blast fungus to rice subspecies. *eLife* **2016**, *5*, e19377. [[CrossRef](#)]
90. Wang, B.-h.; Ebbolle, D.J.; Wang, Z.-h. The arms race between *Magnaporthe oryzae* and rice: Diversity and interaction of Avr and R genes. *J. Integr. Agric.* **2017**, *16*, 2746–2760. [[CrossRef](#)]

-
91. Jones, J.D.; Vance, R.E.; Dangl, J.L. Intracellular innate immune surveillance devices in plants and animals. *Science* **2016**, *354*. [[CrossRef](#)] [[PubMed](#)]
 92. Yang, S.; Li, J.; Zhang, X.; Zhang, Q.; Huang, J.; Chen, J.Q.; Hartl, D.L.; Tian, D. Rapidly evolving *R* genes in diverse grass species confer resistance to rice blast disease. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 18572–18577. [[CrossRef](#)] [[PubMed](#)]