

# 1 ***Magnaporthe oryzae* populations in Sub-Saharan Africa are diverse and show**

## 2 **signs of local adaptation**

3

4 Geoffrey Onaga<sup>1,4\*</sup>, Worrawit Suktrakul<sup>1,2\*</sup>, Margaret Wanjiku<sup>3\*</sup>, Ian Lorenzo Quibod<sup>1\*</sup>, Jean-  
5 Baka Domelevo Entfellner<sup>3</sup>, Joseph Bigirimana<sup>5</sup>, George Habarugira<sup>5</sup>, Rosemary Murori<sup>6</sup>,  
6 Godfrey Asea<sup>4</sup>, Abdelbagi M. Ismail<sup>7</sup>, Chatchawan Jantasuriyarat<sup>2</sup>, Ricardo Oliva<sup>1&</sup>

7

8 <sup>1</sup> Rice Breeding Platform, International Rice Research Institute, DAPO Box 7777, Metro Manila,  
9 Philippines

10 <sup>2</sup> Department of Genetics, Faculty of Science, Kasetsart University, Bangkhen Campus, Ladyao,  
11 Chatuchak, Bangkok 10900, Thailand

12 <sup>3</sup> Biosciences Eastern and Central Africa-International Livestock Research Institute (BecA-ILRI)  
13 Hub, P.O. Box 30709-00100 Nairobi, Kenya

14 <sup>4</sup> National Crop Resources Research Institute (NaCRRI), National Agricultural Research  
15 Organization, Namulonge, P.O. Box 7084, Kampala, Uganda

16 <sup>5</sup> Rice Breeding Platform, International Rice Research Institute, Bujumbura, Burundi

17 <sup>6</sup> Rice Breeding Platform, International Rice Research Institute, Nairobi, Kenya

18 <sup>7</sup> Regional Office for Africa, International Rice Research Institute, Nairobi, Kenya

19

20 \*all contribute equally

21 &Corresponding author: [r.oliva@irri.org](mailto:r.oliva@irri.org)

22

## 23 **Abstract**

24 Rice blast caused by *Magnaporthe oryzae* is one of the most economically damaging diseases of  
25 rice worldwide. The disease originated in Asia but was detected for the first time in Sub-Saharan  
26 Africa (SSA) around 100 years ago. Despite its importance, the evolutionary processes involved  
27 in shaping the population structure of *M. oryzae* in SSA remain unclear. In this study, we  
28 investigate the population history of *M. oryzae* using a combined dataset of 180 genomes. Our  
29 results show that SSA populations are more diverse than earlier perceived, and harbor all genetic  
30 groups previously reported in Asia. While *M. oryzae* populations in SSA and Asia draw from the  
31 same genetic pools, both are experiencing different evolutionary trajectories resulting from

32 unknown selection pressures or demographic processes. The distribution of rare alleles, measured  
33 as Tajima's  $D$  values, show significant differences at the substructure level. Genome-wide analysis  
34 indicates potential events of population contraction strongly affecting *M. oryzae* in SSA. In  
35 addition, the distribution and haplotype diversity of effectors might suggest a process of local  
36 adaptation to SSA conditions. These findings provide additional clues about the evolutionary  
37 history of *M. oryzae* outside the center of origin and help to build customized disease management  
38 strategies.

39

## 40 **Introduction**

41 Rice (*Oryza sativa*) feeds billions of people around the world and represents the only source of  
42 income for millions of smallholder farmers. While Asian countries produce most of the global  
43 supply, a growing demand in Africa is driving the expansion of rice cultivation in the sub-Saharan  
44 Africa (SSA) region (Nasrin et al., 2015). Despite this trend, the SSA rice yields remain relatively  
45 low, averaging 2.2 tonnes per hectare (t/ha) against the global average of 3.4 t/ha (Norman & Kebe,  
46 2006). This low average yield is partly because the production of rice is affected by a range of  
47 pests and diseases that reduce crop yield. In fact, recent estimates suggest that crop losses due to  
48 biotic constraints in the SSA rice sector can reach up to 30% (Savary et al., 2000). One of the most  
49 challenging diseases in the region is rice blast, caused by the ascomycetous fungus *Magnaporthe*  
50 *oryzae* (Gurr et al., 2011). The disease was first reported as "rice fever" in China in 1637 (Agrios,  
51 2005), but the first identification of symptoms matching its description in the SSA region was  
52 reported in Uganda in 1922 (Small, 1922). Subsequent studies reported similar problems in Ghana  
53 in 1923, Kenya in 1924, Congo in 1932, Egypt in 1935, Madagascar, Morocco, Senegal in 1952,  
54 and South Africa in 1956 (Asuyama, 1965). Since that time, the disease has spread throughout the  
55 sub-Saharan rice agro-ecosystems and has become a nuisance to most smallholder farmers in the  
56 region (Séré et al., 2013; Hubert et al., 2015; Onaga et al., 2019).

57 The fungus known as *M. oryzae* is actually composed of distinct genetic groups that infect multiple  
58 cereal crops (Inoue et al., 2017), but the rice-infecting isolates appear to be restricted to a single  
59 lineage (Couch et al., 2005; Choi et al., 2013). Comparative genomic studies on *M. oryzae* have  
60 provided insight into its evolutionary history in rice (Chiapello et al., 2015; Gladieux et al., 2018;  
61 Zhong et al., 2018), pointing out to a southeast Asian origin. Latorre et al. (2020) used a combined  
62 dataset to describe at least three clonal expansions from the original recombinant population

63 occurring in the last ~100-200 years. At least two of the four global genetic groups, most likely  
64 originating from Asia, were consistently present in SSA samples. While the study in Latorre et al.  
65 (2020) involved a limited number of SSA *M. oryzae* strains, other genetic studies relying on larger  
66 collections (Chuwa et al., 2013; Mutiga et al., 2017; Odjo et al., 2018; Onaga & Asea, 2016) hinted  
67 at some amount of variation in the SSA region, which is worth exploring. Significant pathogenic  
68 variations may potentially have implications for adaptation, and consequently, management of *M.*  
69 *oryzae* in Africa.

70 Similar to other plant pathogens, *M. oryzae*'s adaptation is frequently driven by coevolution  
71 processes involving gain and loss of genes (Yoshida et al., 2016). This is particularly important  
72 for effector genes, which modulate host immunity and therefore have strong co-evolutionary  
73 signatures. *M. oryzae* effectors show a high rate of presence/absence polymorphism linked to the  
74 activity of transposable elements (Chiapello et al., 2015; Yoshida et al., 2016). This feature is  
75 likely responsible for variation in host phenotype, since some of the *M. oryzae* effectors are  
76 recognized by immunoreceptors in the rice genome, collectively called *Pi* genes (Yoshida et al.,  
77 2009; Białas et al., 2018). To date, effector studies in *M. oryzae* have also focused on a small  
78 number of isolates mainly collected from Asia. It remains unknown whether the effectors  
79 polymorphism in SSA populations shows patterns similar to the ones in Asian populations, or has  
80 been changing in the process of adaptation to the SSA region. Thus, further interrogation of *M.*  
81 *oryzae* populations in SSA is required to understand the connection between effector distribution  
82 and adaptation.

83 In this study, we combined a genomic dataset of *M. oryzae* from Asia with the genome sequence  
84 of SSA isolates collected in eleven rice-growing countries. We discovered a highly diverse  
85 pathogen population that harbors all known genetic groups predominant in Asia. Interestingly, at  
86 least two *M. oryzae* populations in SSA show different evolutionary trajectories compared to those  
87 in Asia. The patterns of presence/absence of effector genes differ from the Asian members, which  
88 might indicates a process of adaptation to the local hosts. These findings provide additional clues  
89 about the evolutionary history of *M. oryzae* outside its center of origin and help to build  
90 customized disease management strategies.

91  
92

## 93 **Material and Methods**

#### 94 *Collection of M. oryzae isolates*

95 All the 42 isolates used in this study were collected from 11 African countries including Burundi  
96 (4), Kenya (2), Rwanda (3), Tanzania (9), Uganda (5), Benin (4), Burkina Faso (4), Ghana (2),  
97 Mali (2), Nigeria (4), Togo (3). Additional isolates were collected in Asia (4) and Latin America  
98 (3). Isolate metadata is listed in Table S1. Infected leaves collected from rice fields were dried and  
99 kept in filter paper at 4°C before isolation. Leaves with single lesions were placed on glass rods in  
100 Petri dishes with wet filter papers and incubated at room temperature until sporulation. The  
101 sporulating lesions were examined under a stereomicroscope and a group of conidia was  
102 aseptically transferred with a transfer needle to prune agar (PA) medium (3 pieces of prunes, 1 g  
103 yeast extract, 21 g gulaman bar, 5 g alpha-lactose monohydrate, and 1 L distilled H<sub>2</sub>O) and spores  
104 were harvested in distilled water (Meng et al., 2020). Individual germinating conidia were  
105 aseptically transferred and cultured in PA. For long-term storage, each culture was overlaid with  
106 several sterilized filter paper sections and incubated at 25°C. After 10-12 days of incubation, the  
107 colonized filter paper sections were lifted from the agar surface, placed in sterile Petri dishes,  
108 allowed to dry for 3 days at room temperature, and stored at -20°C as described by Valent et al.  
109 (1986). Isolates from Ghana, Burkina Faso, Mali, and Togo were provided in filter paper format  
110 (Table S1). All the isolates are currently curated at the International Rice Research Institute (IRRI)  
111 offices at Biosciences East and Central Africa, hosted by the International Livestock Research  
112 Institute (ILRI) based in Nairobi, Kenya.

113

#### 114 *Generation of genomic datasets*

115 The sequence datasets used in this study were obtained from *M. oryzae* isolates collected in the  
116 field or downloaded from public INSDC databases. Fungal growth and DNA extraction were  
117 performed as described previously (Mutiga et al., 2017). DNA quality checking was carried out  
118 using a NanoDrop 1000 instrument (Thermo Fisher Scientific) and agarose gel electrophoresis.  
119 Libraries were constructed by the Beijing Genomic Institute (Shenzhen, China) using Illumina  
120 paired-end reads with an insert size of 150 bp. Sequencing was performed on Illumina HiSeq4000  
121 with sequencing requirements set to an average coverage of 50x to 70x and a read length of 150  
122 bp. Sequencing quality was affirmed by the fastqc algorithm and the data were trimmed by  
123 removing low-quality sequences and adapter sequences with Trimmomatic 0.36 (Bolger et al.,  
124 2014). The whole-genome sequence of *M. oryzae* 70-15 strain, reference assembly MG8 with

125 accession number GCA\_000002495 (Dean et al., 2005), was used as the reference template for  
126 mapping using BWA-mem 0.7.17 (Li & Durbin, 2009), under default parameters. Mapped reads  
127 were sorted with Samtools 1.3.1 (Li et al., 2009). Duplicate reads were removed using  
128 the *MarkDuplicates* command and all the reads in a file were assigned to a single new read-group  
129 using the *AddOrReplaceReadGroups* command with Picard 2.7  
130 (<http://broadinstitute.github.io/picard>). Single Nucleotide Polymorphisms (SNPs) for each strain  
131 were called using the *HaplotypeCaller* command implemented in the Genome Analyses Toolkit 4  
132 (GATK4.1.6.3) (McKenna et al., 2010). Subsequently, GATK's *GenotypeGVCFs* command was  
133 applied to genotype polymorphic sequence variants for all the strains simultaneously. Hard-  
134 filtering was then performed for the raw SNP calls using the *SelectVariants* and *Variant*  
135 *Filtration* functions of GATK (De Summa et al., 2017). *M. oryzae* isolates with a mapping rate of  
136 less than 80% to the above-mentioned reference strain were discarded for population genetic  
137 analyses, but all the reads were used for effector mapping. In addition, the genomic datasets for  
138 131 *M. oryzae* isolates from a global population (Gladieux, et al., 2018; Zhong et al., 2018) were  
139 downloaded from the Sequence Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/sra>). A  
140 summary of the dataset's sequencing yield and coverage can be found in Table S2.

141

#### 142 *Phylogenetic and population analysis*

143 The phylogenetic tree was built with RAxML 8.2.9 (Stamatakis, 2014). Statistical confidence for  
144 each node was set to 1000 bootstrap runs, utilizing the general time-reversible model of nucleotide  
145 substitution with the Gamma model of rate heterogeneity. The phylogenetic tree was visualized  
146 with the *ggtree* R package (Yu et al., 2017). We also performed a phylogenetic network analysis  
147 employing the neighbor net method implemented in SplitsTree 4.16.1 (Huson & Bryant, 2006).  
148 For the population structure inference, a principal component analysis (PCA) was performed on  
149 the *genlight* object using the *glPCA* function. The population structure was calculated from a  
150 number of clusters (K) ranging from 2 to 8, using the discriminant analysis of principal  
151 components (DAPC) implemented in the *adegenet* R package (Jombart et al., 2010). The  
152 membership probability of each isolate and the most fitting number of clusters by Bayesian  
153 Information Criterion were also performed in *adegenet*. We used another approach to infer the  
154 optimum number of clusters by calculating the Silhouette score in the R package *factoextra*  
155 (Kassambara et al., 2017). To estimate genetic variation, the proportion of genetic variance due to

156 population differentiation and a significant departure from neutrality, we calculated the genome-  
157 wide nucleotide diversity ( $Pi$ ), Wright's fixation index ( $Fst$ ), and Tajima's  $D$  with the variants call  
158 format (VCF) file with VCFtools V.0.1.15 (Danecek et al., 2011) by a sliding window size of 50  
159 kb. All these analysis used the same SNP positions.

160

#### 161 *Effector mapping, distribution, and diversity*

162 To map candidate effectors in SSA *M. oryzae* genomes, we followed the methods and resources  
163 from Latorre et al. (2020) with some modifications. In summary, a total of 178 protein-coding  
164 genes (both virulent and avirulent) from *M. oryzae* isolates that infect rice, wheat, oat, millet, and  
165 wild grasses were used to create reference effector sequences. All the 180 *M. oryzae* genome reads,  
166 131 from Latorre et al., (2020) and 49 from this study, were mapped to effector reference using  
167 bwa-mem 0.7.17 (Li & Durbin, 2009). *Samtools coverage* from samtools 1.10 was used to  
168 calculate the mean coverage of each gene in each of the 180 blast isolates, with the minimum read  
169 depth set at 3x. The total number of mapped reads of each gene was divided by the length of that  
170 gene in the reference (Li et al., 2009). The threshold set to determine the presence of an effector  
171 gene was 80% coverage. A binary presence/absence matrix was created (Table S3). For clustering  
172 purposes, we used a total of 75 informative effector genes that show presence/absence  
173 polymorphisms. The hierarchical clustering analysis was done using the *hclust* function under  
174 complete linkage, and the distance matrix was computed in the R package *ade4* (Dray & Dufour,  
175 2007) with *dist.binary* function adopting the Jaccard index. The PCA and effector loading analysis  
176 were performed as described in Latorre et al., (2020).

177 Furthermore, the initially mapped effector bam files were converted into fastq files by *samtools*  
178 1.10 and *bcftools* 1.10 (Li et al., 2009) for variant calling. The consensus fasta files were created  
179 for each effector gene using *seqtk* 1.3 (Li, 2012) (Additional Data S1). To compute for genetic  
180 diversities for each effector, we chose sequences which have zero presence of "N" or unknown  
181 nucleotide base, and heterozygous base position. The effectors were then aligned using *Mafft*  
182 7.453.0 (Katoh & Standley, 2013) employing the G-INS-i strategy. The final alignments were  
183 manually curated before any test was performed. Effector diversity indices were obtained using  
184 the R package *pegas* (Paradis, 2010). The functions used were as follows: *hap.div* for haplotype  
185 diversity ( $Hd$ ), *haplotype* for the haplotype from the set of sequences, and *nuc.div* for nucleotide  
186 diversity ( $Pi$ ). The pairwise rate of synonymous and non-synonymous codon changes (Ka/Ks ratio)

187 was calculated using KaKs\_Calculator 2.0 (Wang et al., 2010) implementing the Yn00 model  
188 (Yang et al., 2000).

189

## 190 **Results and Discussion**

191

### 192 ***SSA populations of *M. oryzae* are highly diverse and harbor all known Asian genetic groups.***

193 To characterize the genetic composition of Sub-Saharan Africa (SSA) populations of *M. oryzae*,  
194 we combined previous datasets (Gladieux et al., 2018; Zhong et al., 2018) with the genome  
195 sequences of novel isolates collected in SSA rice-growing areas. The assembled genomes represent  
196 disease outbreaks that occurred across 13 different countries between 2012 and 2018 (Table  
197 S1). We identified a total of 66,744 SNPs among the global population. Based on the analysis of  
198 164 global genomes (using isolates with more than 80% read mapping rate to the *MG8* reference  
199 assembly), we found that SSA populations are more diverse than earlier perceived (Latorre et al.,  
200 2020), and harbor all genetic groups previously reported in Asia (Figure 1). To reconstruct the  
201 phylogenetic signal of SSA strains, we used a maximum-likelihood analysis and found diverse  
202 ancestry (Figure 1A), where SSA genomes showed a genetic distribution consistent with multiple  
203 origins. Principal component analysis (PCA) clearly identified four distinct genetic clusters  
204 (Figure 1B), which were also confirmed by the Bayesian Information Criterion (BIC) and  
205 Silhouette score analysis (Figure S1B-C). Phylogenetic networks using Splitree (Figure S2) were  
206 consistent with reports from Latorre et al. (2020) where SSA genomes fall under the three clonal  
207 lineages, consistent with genetic groups 2, 3, and 4; and an additional highly diverse cluster  
208 representing the group 1 (Latorre et al., 2020). Population structure analysis ( $K = 2 > 8$ ) using 164  
209 global *M. oryzae* genomes also supports the hypothesis that SSA populations are highly diverse  
210 (Figure 1C) and resemble most of the diversity found in Asia (Chuwa et al., 2013; Onaga & Asea,  
211 2016; Gladieux et al., 2018; Zhong et al., 2018; Latorre et al., 2020), representing genetic groups  
212 1, 2, 3, and 4.

213 Recent studies suggest that the rice-infecting lineage of *M. oryzae* originated in Asia, most likely  
214 from populations infecting foxtail millet (Chiapello et al., 2015; Gladieux, Condon, et al., 2018 b).  
215 It is also suggested that this lineage emerged from a recombinant population with distinct genetic  
216 backbones (Thierry et al., 2020; Latorre et al., 2020). The rich history of rice cultivation in Africa  
217 might have allowed the early establishment of *M. oryzae* groups in different waves (Choi et al.,

218 2019). We speculated that the first wave started before the proposed clonal expansion (Latorre et  
219 al., 2020), probably from a recombinant population similar to genetic group 1. We found  
220 representative genomes in East Africa (Uganda) and West Africa (Mali, Togo, and Nigeria), which  
221 show genetic features of the Asian metapopulation but are quite distant from any known genome  
222 in Asia. Population structure analysis of group 1 also suggests some level of sub-structuring  
223 (Figure S1A;  $K=6$ ). For instance, isolate E-UGD-32 from Tilda irrigation scheme in Uganda has  
224 unique features that might represent the emergence of novel variants in SSA. The significant  
225 differences of the SSA-1 genomes might point out to ancestral colonization rather than a recent  
226 one. A second wave appears to be more recent, involving the clonal groups 2, 3, and 4. The fact  
227 that blast symptoms were reported in Africa by 1922 suggests that such colonization occurred soon  
228 after the Asia expansion. This might explain why representatives of clonal genetic groups 2, 3, and  
229 4 in SSA have a similar genetic background as the Asian counterparts (Figure 1A).

230

### 231 ***SSA populations of *M. oryzae* are evolving in different patterns compared to Asia.***

232 To further investigate the overall demographic patterns of diversity in *M. oryzae*, we calculated  
233 population differentiation using genome-wide estimations of nucleotide diversity ( $P_i$ ) and  
234 population substructure ( $F_{st}$ ). Overall, the  $P_i$  values in Asia and SSA were significantly different  
235 ( $p = 0.04$ ). While mean nucleotide diversity ( $P_i$ ) in group 1 ( $P_i = 1.3e-04$ ) was higher compared  
236 to the clonal groups 2 ( $P_i = 2.5e-05$ ), 3 ( $P_i = 2.1e-05$ ), and 4 ( $P_i = 2.45e-05$ ) (Figure S3A; Table  
237 S4), Asia accumulate more diversity than SSA in every group (Figure 2A; Table S5). The fixation  
238 index ( $F_{st}$ ) across regions ( $F_{st} = 0.05$ ) or within each genetic group ( $F_{st} = 0.03$  to  $0.08$ ) (Figure  
239 2B) were low compared to  $F_{st}$  values between genetic groups ( $F_{st} = 0.25 - 0.72$ ) (Figure S4). We  
240 also detected that the major source of diversity is coming from genetic group 1 in Asia (Figure  
241 S3B; Figure S4). The observed values of nucleotide diversity and genetic structure of *M.*  
242 *oryzae* suggest that SSA and Asian populations belong to the same genetic pool and that each  
243 genetic group might exchange alleles between regions but not with other genetic groups. The  
244 observed patterns are consistent with panmixis in a rice-specialized asexual pathogen, but also  
245 aligns with independent drifts of each clonal lineage during the recent emergence and spread of  
246 the pathogen around the globe (Latorre et al., 2020).

247 To explore the selection process driving the evolution of *M. oryzae* populations in Asia and SSA,  
248 we calculated genome-wide Tajima's  $D$  and found significant differences at the substructure level.

249 Overall Tajima's  $D$  values of the global *M. oryzae* showed a deviation from a neutrally evolving  
250 population (Figure S3C; Table S4), suggesting the influence of non-random events. Similar to  
251 Latorre et al. (2020), we observed negative Tajima's  $D$  values (Tajima's  $D = -0.6$  to  $-1.2$ ) in most  
252 of the Asian genetic groups (Figure 2C; Table S5), representing the accumulation of rare alleles in  
253 the backbone genome of *M. oryzae*. These patterns are usually indicative of population size  
254 expansion that follows a bottleneck or selective sweep. In contrast to Asia, we found SSA genetic  
255 group 1 (Tajima's  $D = 0.56$ ) and genetic group 2 (Tajima's  $D = 0.58$ ) having positive  
256 Tajima's  $D$  values (Figure 2C; Table S5), which might point out to non-random removal of rare  
257 alleles from the population. Since not all the SSA groups experience significant differences in  
258 Tajima's  $D$ , one hypothesis is that a sudden population contraction occurred in the SSA region,  
259 specifically affecting genetic groups 1 and 2. Interestingly, the differences in Tajima's  $D$  appears  
260 to be scattered across the genome rather than concentrated in particular chromosomal regions  
261 (Figure 2D; Figure S5). For SSA genetic group 1, these differences appear to be more pronounced  
262 in chromosomes 4, 5, and 7 (Figure S5). A sudden population contraction could potentially mimic  
263 the observed Tajima's  $D$  in genetic group 2, and highly likely represents the retention of prevalent  
264 populations before the contraction. However, the chromosomal Tajima's  $D$  patterns observed in  
265 genetic group 1 could indicate the retention, in SSA, of cryptic or ancestral groups now  
266 disappeared from Asia.

267 Our data suggest that SSA and Asian populations of *M. oryzae* might be experiencing slightly  
268 different evolutionary trajectories resulting from unknown selection or demographic processes.  
269 Factors such as human interventions or weather patterns are known to produce significant selection  
270 pressure in agricultural pathogens. Recently, Thierry et al. (2020) linked climatic variation with  
271 the geographic distribution of *M. oryzae* groups. In addition, the management practices and the  
272 diversity of growing environments in Asia might be relatively different from Africa, where upland  
273 and rainfed ecologies occupy the largest share, and very little fertilizer input is used. Other factors,  
274 such as host composition, might be also likely shaping the diversity of *M. oryzae* in the region but  
275 its contribution needs further investigation.

276

### 277 ***Effector distribution and diversification suggests local adaptation to SSA conditions.***

278 Similar to other filamentous plant pathogens, the genome architecture of *M. oryzae* appears to be  
279 shaped by the dynamics of repetitive elements (Kelkar & Ochman, 2012), which drive the rapid

280 evolution of effector genes in response to selection (Dong et al., 2015). To understand the level of  
281 adaptation of *M. oryzae* populations in SSA, we compared the number and distribution of effector  
282 repertoires across the region and found significant differences at a substructure level. We mapped  
283 178 predicted effector references using the genome sequence of 164 *M. oryzae* isolates. The  
284 overall number of effectors ranged from 110 to 127 per isolate and varied across genetic groups  
285 (Figure S6A-B). Similar to Latorre et al., (2020) we found a different number of effector  
286 repertoires across genetic groups. This variation can be explained by clonality and divergence  
287 processes (Figure 1; Figure 2). While the number of effectors within genetic groups 2, 3, and 4  
288 were similar in Asian and SSA isolates, significant differences were observed within genetic group  
289 1 (Figure 3A). It is not clear whether the differences between genetic group 1 in Asia and SSA is  
290 the result of cryptic events of genome expansion in the ancestral backbones (Dong et al., 2015) or  
291 emerged as a process of local adaptation affecting specific effectors in the region.

292 We then used a subset of 75 effectors to assess presence/absence polymorphism in the *M.*  
293 *oryzae* genomes (Table S3). We found that effector repertoires tend to have similar but not exact  
294 patterns in each genetic group (Figure 3B), suggesting that the clonal lineages retain virulent  
295 capabilities but rapidly gain or loss of genes from the pool. The effector loading analysis identified  
296 16 informative effectors that explain this distribution (Figure S6C-E) and some are also present in  
297 a previous report from Latorre et al., (2020). Interestingly, we found that effector patterns from  
298 SSA *M. oryzae* genomes form specific sub-clusters within genetic groups 1, 3, and 4 (Figure 3B),  
299 which might indicate the presence of locally adapted groups in SSA. In addition, we repeated the  
300 analysis using 180 *M. oryzae* genomes and found the same distribution (data not shown).

301 To assess the diversity of *M. oryzae* effectors across regions, we extracted consensus sequences  
302 from all genes and identified allelic variants. The number of haplotypes in a subset of 96 effectors  
303 ranged from 1 to 16. The overall haplotype diversity ( $Hd$ ), measured as the probability of finding  
304 different alleles, was higher in Asia compare to SSA for all groups. The same was also true when  
305 computing for nucleotide diversity (Figure 4A-B). We then assessed the diversity of each effector  
306 in each genetic group and found that genetic group 1 is statistically more diverse than any other  
307 genetic groups (Figure S7A-B). This observation aligned with the genome-wide diversity in Figure  
308 S3A. While more effector haplotypes were observed in Asian populations, unique effector  
309 haplotypes were present in SSA genomes (Figure S8). The presence/absence distribution and  
310 sequence differences in SSA genomes suggesting a certain level of adaptation.

311 Interestingly, we only found a handful of effectors showing signatures of positive selection ( $Ka/Ks$   
312  $> 1$ ), with not major differences between Asia and SSA (Figure 4; Figure S7C). For instance,  
313 effector PH14.00186401 showed  $Ka/Ks > 1$  but also unique distribution of haplotypes in SSA  
314 (Figure S8). A limited number of effectors with  $Ka/Ks$  greater than 1 were also noticed by Kim et  
315 al. (2019) when comparing *M. oryzae* from different hosts. Thus, purging of allelic variants in  
316 *M. oryzae* genomes (*sensu lato*) might be a common process for this pathogen. The  $Ka/Ks$  values  
317 can be explained by purifying selection but also associated with the continuous effect of repetitive  
318 elements on the turnover of effector genes.

319 Altogether, our data support different evolutionary trajectories of *M. oryzae* groups in Asia and  
320 SSA. The number, distribution, and diversification of effector genes appear to indicate that host  
321 selection might have played a critical role in this process. Rice cultivation in Africa started around  
322 3200 BP when local farmers domesticated the rice species *O. glaberrima* (Murray, 2004). While  
323 the use of *O. glaberrima* gradually declined with the introduction of the Asian species *O.*  
324 *sativa* (Cubry et al., 2018), it was not until 1870 when its adoption intensified (Linares, 2002).  
325 Nowadays, *O. glaberrima* has been introgressed in multiple breeding programs and it is highly  
326 represented in a significant number of released popular varieties (e.g. IRAT, ROK, or NERICA  
327 types). These varieties captured the local adaptability features from *O. glaberrima*, but might have  
328 specific stress-related genes as well. In fact, blast screenings of diverse collections of *O.*  
329 *glaberrima* yielded a number of resistance phenotypes (Bidiaux, 1978; Silue, 1991; Yelome et al.,  
330 2018) suggesting the presence of cryptic *R* genes. If *M. oryzae* colonized SSA in multiple waves,  
331 it is highly likely that such adaption involved, at least partially, the interaction with *O.*  
332 *glaberrima* genotypes. The difference in host species might, in principle, explain the unique  
333 features of *M. oryzae* SSA genomes, but more detailed studies are needed to understand the broad  
334 variety of plant receptors in this African rice species.

335

### 336 **Conclusion**

337 The migration of plant pathogens to new agricultural ecosystems represent a major concern for  
338 long-term strategies of food security. Understanding the events that shaped the pathogen  
339 population structure in the new setup is likely to help to develop effective control measurements.  
340 In this report, we reconstructed the evolutionary trajectory of the rice blast pathogen *M.*  
341 *oryzae*, present in Sub-Saharan Africa (SSA) at least for the last 100 years. The SSA populations

342 harbor all the genetic complexity from the Asian population but display additional features that  
343 suggest local adaptation. The distribution and diversity of effector repertoires also indicate  
344 elevated virulent potential and cryptic diversity that is worth exploring in the future. The report  
345 provides important insights into the complexity of this pathogen that threatens rice cultivation in  
346 Africa and advice on the directions of future resistance management deployment strategies.

347

## 348 **Figure Legends**

349

350 **Figure 1.** The genetic composition of rice-infecting populations of *Magnaporthe oryzae* in Sub-  
351 Saharan Africa (SSA) involves all know genetic groups from Asia. (A) A maximum-likelihood  
352 phylogenetic tree constructed with single nucleotide polymorphism alignment from 164 global  
353 isolates. Black nodes depict bootstrap scores greater than 90%. Genetic groups 1, 2, 3, and 4 were  
354 based from the results of Figure 1B and Figure S1A-C. The origin of the genomic dataset is  
355 indicated in colors. (B) A two dimensions principal correspondence analysis (PCA) showing the  
356 distribution of the genetic variability in the *M. oryzae* datasets. (C) Spatial distribution of *M.*  
357 *oryzae* genetic groups collected in rice-growing regions across the globe.

358

359 **Figure 2.** Patterns of diversity and population differentiation across rice-infecting *Magnaporthe*  
360 *oryzae* genetic groups in Sub-Saharan Africa (SSA) and Asia. (A) Nucleotide diversity ( $P_i$ ) was  
361 calculated for each group. SSA1-4 and Asia1-4 represent each of the genetic groups described in  
362 Figure 1. (B) Population substructure ( $F_{st}$ ) values calculated for each genetic group in SSA and  
363 Asia. (C) Genome-wide Tajima's  $D$  computations of genetic groups from SSA and Asia showing  
364 substantial differences within groups 1 and 2. To compare the distribution of each genome-wide  
365 diversity test analysis in each genetic group within different regions, Mann-Whitney test was  
366 performed as shown above in the boxplot. (D) Example of the dramatic change in Tajima's  $D$   
367 values on *M. oryzae* chromosome 7 within genetic group 2 from SSA and Asia. Sequence variation  
368 of *M. oryzae* genes within a 5 kb window representing color-coded Single Nucleotide  
369 Polymorphisms (SNPs).

370

371 **Figure 3.** Variation in the number and distribution of candidate effector repertoires  
372 in *Magnaporthe oryzae* genetic groups collected from Sub-Saharan Africa (SSA). (A) A box plot

373 representing the average number of effector genes in each genetic group from SSA and Asia. To  
374 compare the distribution of each effector content within the genetic group between different  
375 regions, Mann-Whitney test was performed as shown above in the boxplot. SSA1-4 and Asia1-4  
376 represent each of the genetic groups described in Figure 1. (B) A hierarchical heatmap representing  
377 the presence/absence patterns of candidate effectors in 164 genomes. SSA (triangles), Asia  
378 (circles), and Other (squares) regions are depicted. Color labels in the tree represent each of the  
379 genetic group. Gray and white colors in the heat map represent the presence/absence of effectors  
380 as <80% of coverage. Gene names (rows) and isolate names (columns) are described in Table S1.  
381 Sub clusters in genetic groups 1, 3, and 4 are indicated as gray nodes in the tree. Complete-linkage  
382 clustering was performed as visualized in the dendrogram.

383  
384 **Figure 4.** Sequence variation and signatures of selection among candidate effectors repertoires  
385 in *Magnaporthe oryzae* genetic groups collected from Sub-Saharan Africa (SSA) and Asia. (A)  
386 Haplotype diversity ( $Hd$ ) of effector genes in each genetic group from SSA and Asia. SSA1-4 and  
387 Asia1-4 represent each of the genetic groups described in Figure 1. The box plot represents only  
388 effectors with more than one haplotype. (B) Box plot representing nucleotide diversity ( $Pi$ ) of the  
389 same dataset. (C) Distribution of synonymous and non-synonymous codon changes (average  
390  $Ka/Ks$  ratio) across candidate effectors from *M. oryzae* genetic groups collected in SSA and Asia.  
391 To compare the distribution of each effector diversity within the genetic group between different  
392 regions, Mann-Whitney test was performed as shown above in the boxplot.

### 393 394 **Supplementary Material**

395 **Table S1.** Name, origin, and year of collection of all rice-infecting *Magnaporthe oryzae* isolates  
396 used in this study.

397 **Table S2.** The number of mapped reads obtained by each of the *Magnaporthe oryzae* datasets used  
398 in the study.

399 **Table S3.** A binary matrix representing presence/absence of effectors genes in 180 *Magnaporthe*  
400 *oryzae* genomes worldwide.

401 **Table S4.** Mean and median genome-wide nucleotide diversity ( $Pi$ ) and Tajima's  $D$   
402 of *Magnaporthe oryzae* genetic groups globally.

403 **Table S5.** Mean and median genome-wide nucleotide diversity ( $P_i$ ), and Tajima's  $D$   
404 of *Magnaporthe oryzae* genetic groups in Asia and SSA.

405  
406 **Figure S1.** Four genetic groups were inferred in 164 global *Magnaporthe oryzae* isolates based on  
407 the whole genome SNPs. (A) Bar plot showing the membership probability of each genome from  
408  $K=2$  to  $K=8$  populations. The clusters were built using a discriminant analysis of principal  
409 components (DAPC). (B) The Bayesian information criterion (BIC), and (C) the Silhouette score  
410 both hint at 4 genetic groups as the optimum number of clusters (elbow in the BIC curve and  
411 maximum Silhouette score).

412 **Figure S2.** Phylogenetic network analysis of global *Magnaporthe oryzae* populations using the  
413 neighbor net method showing the four inferred genetic groups. The color denotes the four genetic  
414 groups as inferred by the PCA in Figure 1B and confirmed by the clustering analysis in Figure  
415 S1B-C.

416 **Figure S3.** Genome-wide genetic analysis of *Magnaporthe oryzae* genetic groups. (A) Nucleotide  
417 diversity ( $P_i$ ) analysis of each genetic group. The  $P_i$  analysis presented genetic group 1 as the most  
418 diverse genetic group. (B) Fixation index ( $F_{st}$ ) among genetic groups. The  $F_{st}$  analysis reveals  
419 genetic group 1 as a major source of genetic flow between genetic groups. The color in each box  
420 plot designates the pairwise comparison between genetic groups. (C) Tajima's  $D$  computation  
421 among genetic groups shows negative Tajima's  $D$  values. The overall patterns are similar to  
422 previous reports (Latorre et al., 2020).

423 **Figure S4.** Fixation index ( $F_{st}$ ) between different *Magnaporthe oryzae* genetic groups across  
424 different regions. The pattern shows genetic group 1 in Asia (Asia-1) shares diversity to almost all  
425 populations. In contrast, genetic group 1 from SSA (SSA-1) shares less diversity with all the  
426 groups. Clonal lineages in Asia or SSA also show a range of differentiation. The color in each box  
427 plot designates the pairwise comparison between genetic groups in each region.

428 **Figure S5.** Genome-wide Tajima's  $D$  from *Magnaporthe oryzae* chromosomes reveals different  
429 patterns in genetic group 1 and 2 across regions. Tajima's  $D$  of each genetic group from  
430 chromosomes one to seven was built using a 50kb sliding window. The different genetic group  
431 colors correspond to the predicted groups from Figure 1B and Figure S1B-C. Tajima's  $D$  values  
432 for Asia and Sub-Saharan Africa (SSA) regions are depicted as grey and black lines.

433 **Figure S6.** Effector repertoires in *Magnaporthe oryzae* reveal distinct patterns of diversification  
434 in each genetic group. (A) An assortment of the total number of effectors per isolate from highest  
435 (CH0333 = 127) to lowest (IN0072 = 110). (B) A box plot of the total effector content in each  
436 genetic group. To compare the distribution of the total effector for every isolate in each of the four  
437 genetic groups, Mann-Whitney test was performed as shown above in the boxplot. (C) PCA biplot  
438 using 82 subsets of effectors from the presence/absence matrix (Figure 3B). The effector loading  
439 vectors are indicated in the arrow. (D) The bar plot shows the product for each effector loading  
440 vectors. The redline reveals 90% of the cumulative sum from the data, or in this case, sixteen  
441 effectors that can explain the distribution. (E) Complete hierarchical clustering dendrogram of the  
442 sixteen effectors based on the results in Figure S6E. The distance matrix was computed using the  
443 Jaccard index.

444 **Figure S7.** Diversity of *Magnaporthe oryzae* effectors associates with a strong purifying selection  
445 (A) Haplotype diversity (*Hd*) analysis of each effector from the four genetic groups. (B) Nucleotide  
446 diversity (*Pi*) of each effector from the four genetic groups. (C) Ka/Ks effector distribution in each  
447 genetic group. To compare the distribution of each effector diversity test analysis in each of the  
448 four genetic groups, Mann-Whitney test was performed as shown above in the boxplot.

449 **Figure S8.** Haplotype diversity (*Hd*) of effector repertoires in *Magnaporthe oryzae* genetic groups  
450 from Asia and Sub Saharan Africa (SSA). The heatmap shows haplotype frequency from 96  
451 effectors present in each genetic group across regions. Effector haplotypes range from one to  
452 sixteen. There are more effector haplotypes in Asia compared to SSA, regardless of genetic group.  
453 Low frequency (yellow) and high frequency (red) is based on the total number of isolates that have  
454 that particular haplotype in each effector. The color in the text is based on the four genetic groups  
455 inferred in Figure S2.

456  
457 **Additional Data S1.** The complete effector sequences files for each of the 180 strains in fasta file

#### 459 **Data availability**

460 The sequences produced in this study are stored under BioProject PRJNA670311. Genomic files  
461 in Genbank format are available at NCBI. Effector sequences from each of the 180 strains are  
462 provided in Additional Data 1.

463

464 **Acknowledgment**

465 The authors would like to thank Mr. Inosters Nzuki for technical assistance. Blast isolates from  
466 Burkina Faso, Ghana, Mali, and Togo were kindly provided by Dr. Samuel Mutiga (Biosciences  
467 eastern and central Africa – International Livestock Research Institute) to whom we are grateful.  
468 We are grateful to the DOST–Advanced Science and Technology Institute (DOST-ASTI) for free  
469 access to high-performance computing services. We also like to thank Scientists at IRRI are  
470 partially funded by the Research Program on Rice Agri-food System (RICE). Funding support also  
471 came from The Bill and Melinda Gates Foundation through the STRASA project.

472

473 **Author contribution**

474 GO, JBDE, CJ, AI, and RO designed the research; GO, ILQ, WS, MW, performed research  
475 experiments; and analyzed the data; GO, JB, GH, RM arrange disease collection; CJ, GA, GO,  
476 RO, JB, CJ, supervised the students; RO, GO, ILQ wrote the paper

477

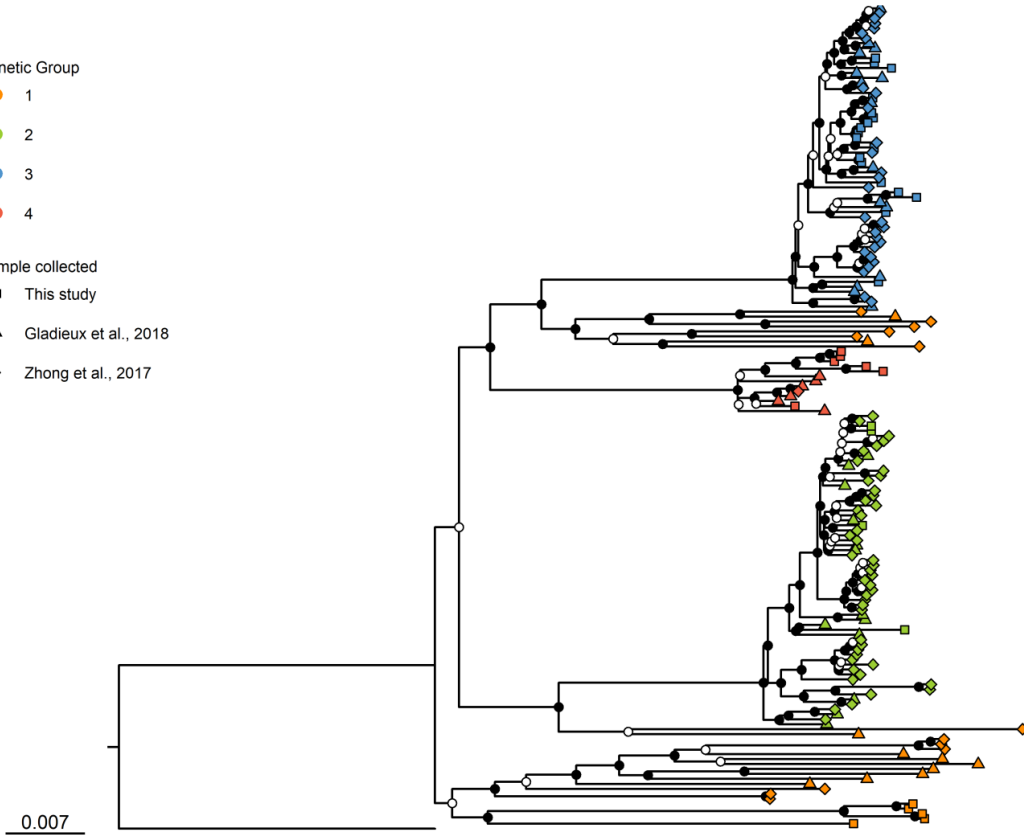
478

479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509

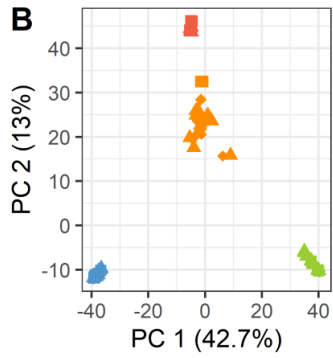
**A**

Genetic Group  
● 1  
● 2  
● 3  
● 4

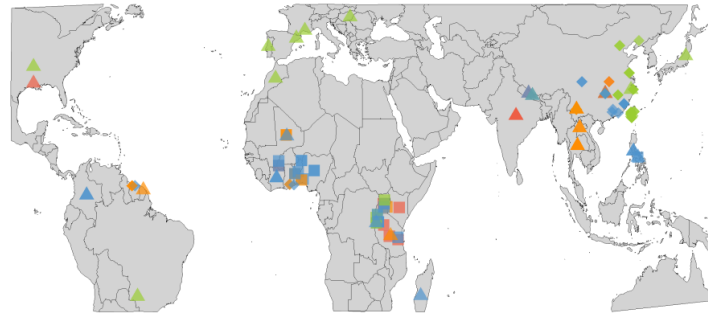
Sample collected  
■ This study  
▲ Gladieux et al., 2018  
◆ Zhong et al., 2017



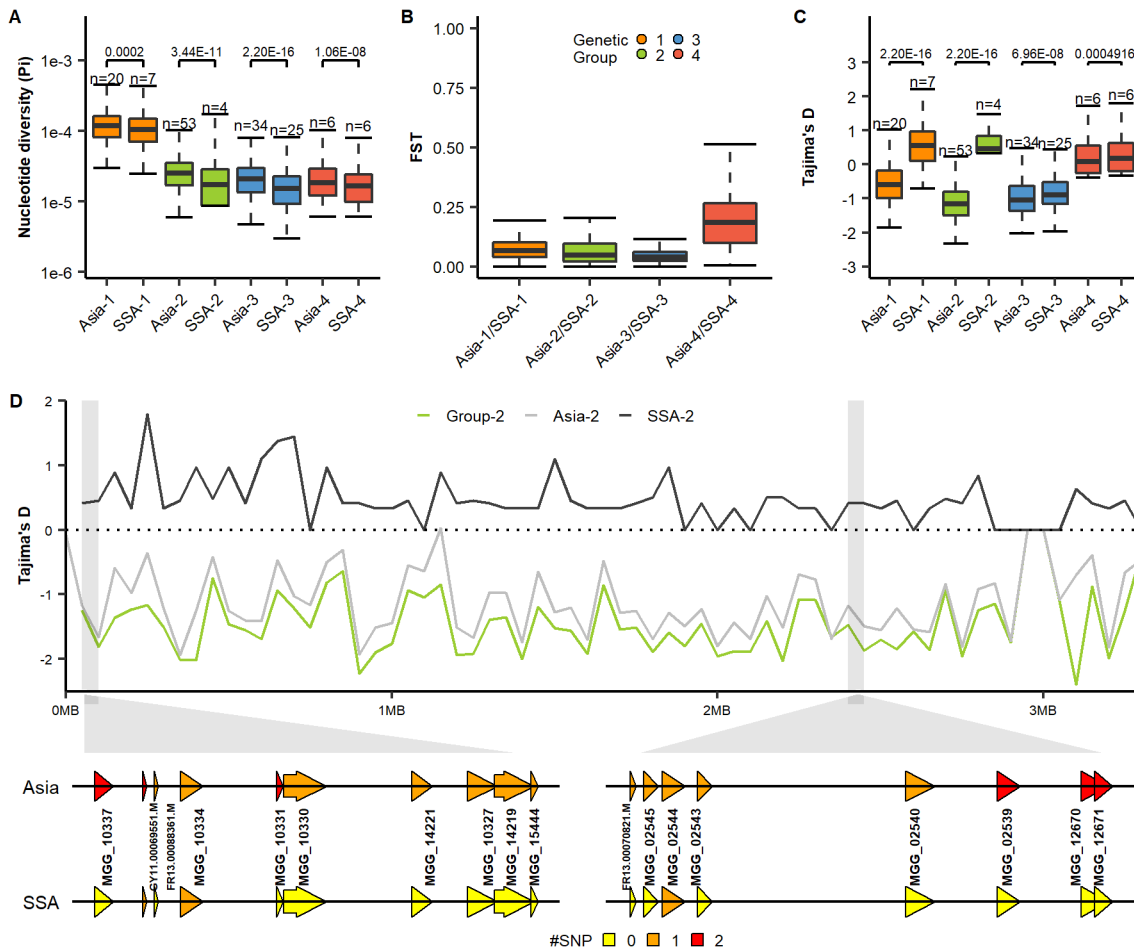
**B**



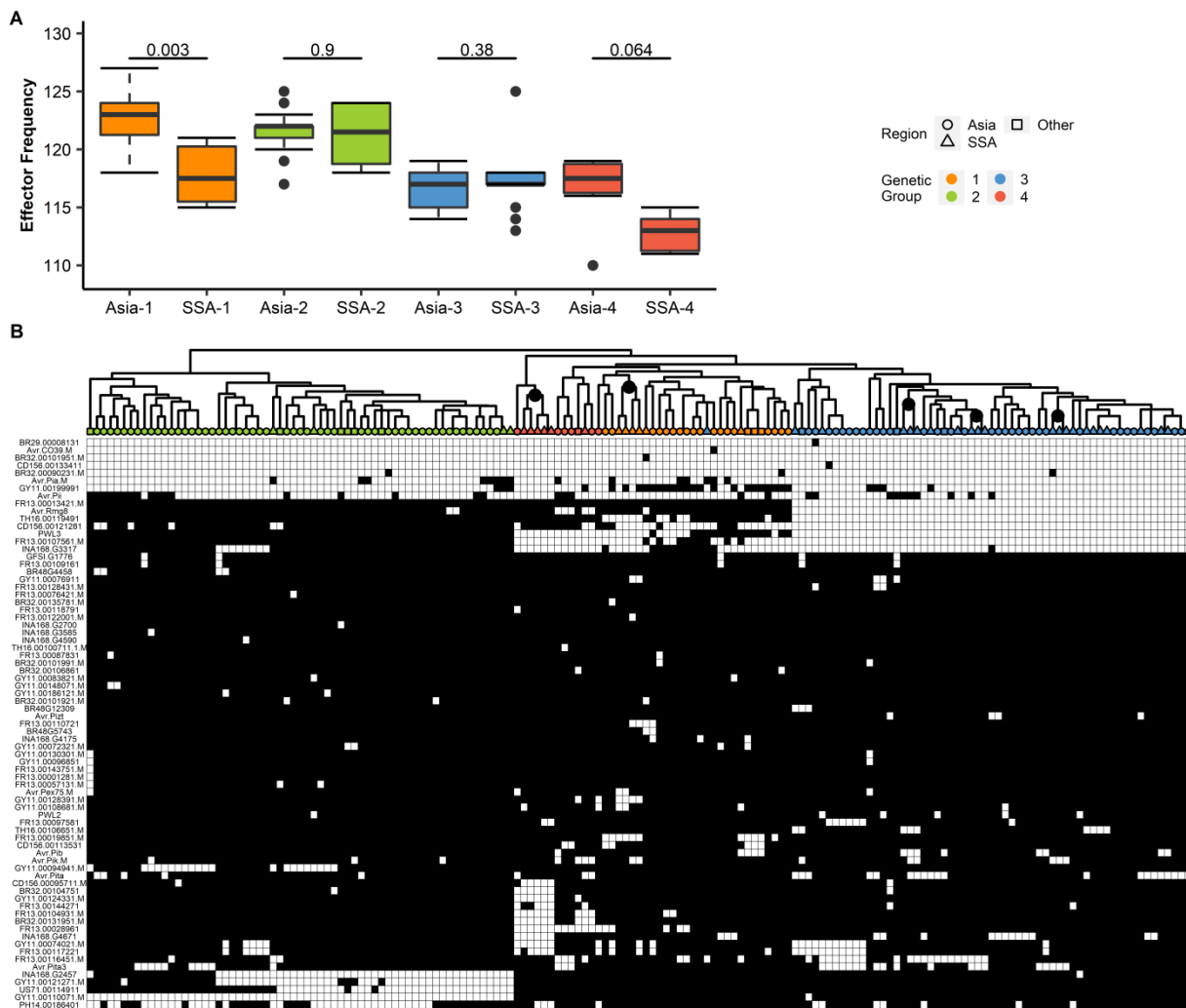
**C**



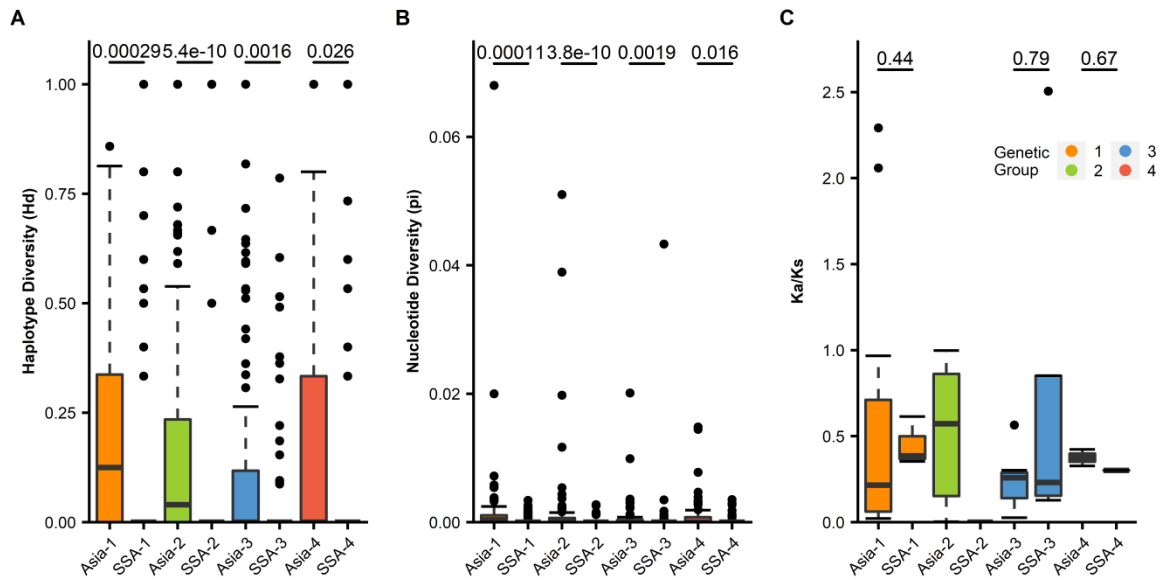
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540



541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571



572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592



## 593 Reference

- 594 Agrios, G. N. (2005). Plant diseases caused by fungi. *Plant Pathology*, 4.
- 595 Asuyama, H. (1965). Morphology, taxonomy, host range, and life cycle of *Piricularia*
- 596 *oryzae*. *The Rice Blast Disease*, 9–12.
- 597 Białas, A., Zess, E. K., De la Concepcion, J. C., Franceschetti, M., Pennington, H. G.,
- 598 Yoshida, K., Upson, J. L., Chanclud, E., Wu, C.-H., & Langner, T. (2018). Lessons in effector
- 599 and NLR biology of plant-microbe systems. *Molecular Plant-Microbe Interactions*, 31(1), 34–
- 600 45.
- 601 Bidaux, J. (1978). Screening for horizontal resistance to rice blast (*Pyricularia oryzae*) in
- 602 Africa. *Rice in Africa*, 159–174.
- 603 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for
- 604 Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.

- 605 Chiapello, H., Mallet, L., Guerin, C., Aguilera, G., Amselem, J., Kroj, T., Ortega-  
606 Abboud, E., Lebrun, M.-H., Henrissat, B., & Gendrault, A. (2015). Deciphering genome content  
607 and evolutionary relationships of isolates from the fungus *Magnaporthe oryzae* attacking  
608 different host plants. *Genome Biology and Evolution*, 7(10), 2896–2912.
- 609 Choi, J., Park, S.-Y., Kim, B.-R., Roh, J.-H., Oh, I.-S., Han, S.-S., & Lee, Y.-H. (2013).  
610 Comparative analysis of pathogenicity and phylogenetic relationship in *Magnaporthe grisea*  
611 species complex. *PloS One*, 8(2), e57196.
- 612 Chuwa, C. J., Mabagala, R. B., & Reuben, M. (2013). Pathogenic Variation and  
613 molecular characterization of *Pyricularia oryzae*, Causal agent of rice blast disease in Tanzania.  
614 *Int. J. Sci and Res*, 4(11), 1131–1139.
- 615 Couch, B. C., Fudal, I., Lebrun, M.-H., Tharreau, D., Valent, B., Van Kim, P.,  
616 Nottéghem, J.-L., & Kohn, L. M. (2005). Origins of host-specific populations of the blast  
617 pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic  
618 clones on rice and weeds of rice. *Genetics*, 170(2), 613–630.
- 619 Cubry, P., Tranchant-Dubreuil, C., Thuillet, A.-C., Monat, C., Ndjiondjop, M.-N.,  
620 Labadie, K., Cruaud, C., Engelen, S., Scarcelli, N., & Rhoné, B. (2018). The rise and fall of  
621 African rice cultivation revealed by analysis of 246 new genomes. *Current Biology*, 28(14),  
622 2274–2282.
- 623 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A.,  
624 Handsaker, R. E., Lunter, G., Marth, G. T., & Sherry, S. T. (2011). The variant call format and  
625 VCFtools. *Bioinformatics*, 27(15), 2156–2158.
- 626 De Summa, S., Malerba, G., Pinto, R., Mori, A., Mijatovic, V., & Tommasi, S. (2017).  
627 GATK hard filtering: Tunable parameters to improve variant calling for next generation  
628 sequencing targeted gene panel data. *BMC Bioinformatics*, 18(5), 119.
- 629 Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J.,  
630 Thon, M., Kulkarni, R., Xu, J.-R., & Pan, H. (2005). The genome sequence of the rice blast  
631 fungus *Magnaporthe grisea*. *Nature*, 434(7036), 980–986.
- 632 Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous  
633 pathogens: Waltz with plants. *Current Opinion in Genetics & Development*, 35, 57–65.
- 634 Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram  
635 for ecologists. *Journal of Statistical Software*, 22(4), 1–20.
- 636 Gladieux, P., Condon, B., Ravel, S., Soanes, D., Maciel, J. L. N., Nhani, A., Chen, L.,  
637 Terauchi, R., Lebrun, M.-H., & Tharreau, D. (2018). Gene flow between divergent cereal-and  
638 grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. *MBio*, 9(1).
- 639 Gladieux, P., Ravel, S., Rieux, A., Cros-Arteil, S., Adreit, H., Milazzo, J., Thierry, M.,  
640 Fournier, E., Terauchi, R., & Tharreau, D. (2018). Coexistence of multiple endemic and  
641 pandemic lineages of the rice blast pathogen. *MBio*, 9(2).
- 642 Gurr, S., Samalova, M., & Fisher, M. (2011). The rise and rise of emerging infectious  
643 fungi challenges food security and ecosystem health. *Fungal Biology Reviews*, 25(4), 181–188.

- 644 Hubert, J., Mabagala, R. B., & Mamiro, D. P. (2015). *Efficacy of selected plant extracts*  
645 *against Pyricularia grisea, causal agent of rice blast disease.*
- 646 Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary  
647 studies. *Molecular Biology and Evolution*, 23(2), 254–267.
- 648 Inoue, Y., Vy, T. T., Yoshida, K., Asano, H., Mitsuoka, C., Asuke, S., Anh, V. L.,  
649 Cumagun, C. J., Chuma, I., & Terauchi, R. (2017). Evolution of the wheat blast fungus through  
650 functional losses in a host specificity determinant. *Science*, 357(6346), 80–83.
- 651 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal  
652 components: A new method for the analysis of genetically structured populations. *BMC*  
653 *Genetics*, 11(1), 94.
- 654 Kassambara, A., & Mundt, F. (2017). Factoextra: Extract and visualize the results of  
655 multivariate data analyses. *R Package Version*, 1(5), 337–354.
- 656 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software  
657 version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4),  
658 772–780.
- 659 Kelkar, Y. D., & Ochman, H. (2012). Causes and consequences of genome expansion in  
660 fungi. *Genome Biology and Evolution*, 4(1), 13–23.
- 661 Kim, K.-T., Ko, J., Song, H., Choi, G., Kim, H., Jeon, J., Cheong, K., Kang, S., & Lee,  
662 Y.-H. (2019). Evolution of the genes encoding effector candidates within multiple pathotypes of  
663 *Magnaporthe oryzae*. *Frontiers in Microbiology*, 10, 2575.
- 664 Latorre, S. M., Reyes-Avila, C. S., Malmgren, A., Win, J., Kamoun, S., & Burbano, H.  
665 A. (2020). Differential loss of effector genes in three recently expanded pandemic clonal  
666 lineages of the rice blast fungus. *BMC Biology*, 18(1), 1–15.
- 667 Li, H. (2012). Seqtk Toolkit for processing sequences in FASTA/Q formats. *GitHub*, 767,  
668 69.
- 669 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–  
670 Wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
- 671 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,  
672 Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools.  
673 *Bioinformatics*, 25(16), 2078–2079.
- 674 Linares, O. F. (2002). African rice (*Oryza glaberrima*): History and future potential.  
675 *Proceedings of the National Academy of Sciences*, 99(25), 16360–16365.
- 676 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A.,  
677 Garimella, K., Altshuler, D., Gabriel, S., & Daly, M. (2010). The Genome Analysis Toolkit: A  
678 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*,  
679 20(9), 1297–1303.
- 680 Meng, X., Xiao, G., Telebanco-Yanoria, M. J., Siazon, P. M., Padilla, J., Opulencia, R.,  
681 Bigirimana, J., Habarugira, G., Wu, J., & Li, M. (2020). The broad-spectrum rice blast resistance  
682 (R) gene *Pita2* encodes a novel R protein unique from *Pita*. *Rice*, 13(1), 1–15.

- 683 Murray, S. S. (2004). Searching for the origins of African rice domestication. *Antiquity*,  
684 78(300), 1–3.
- 685 Mutiga, S. K., Rotich, F., Ganeshan, V. D., Mwongera, D. T., Mgonja, E. M., Were, V.  
686 M., Harvey, J. W., Zhou, B., Wasilwa, L., & Feng, C. (2017). Assessment of the virulence  
687 spectrum and its association with genetic diversity in *Magnaporthe oryzae* populations from sub-  
688 Saharan Africa. *Phytopathology*, 107(7), 852–863.
- 689 Nasrin, S., Lodin, J. B., Jirström, M., Holmquist, B., Djurfeldt, A. A., & Djurfeldt, G.  
690 (2015). Drivers of rice production: Evidence from five Sub-Saharan African countries.  
691 *Agriculture & Food Security*, 4(1), 1–19.
- 692 Norman, J., & Kebe, B. (2006). African smallholder farmers: Rice production and  
693 sustainable livelihoods. *International Rice Commission Newsletter*, 55, 33–44.
- 694 Odjo, T., KASSANKOGNO, A. I., Adreit, H., Milazzo, J., Ravel, S., Gumedzoé, Y. M.  
695 D., Ouedraogo, I., Silue, D., & Tharreau, D. (2018). Diversity and structure of African  
696 populations of *Magnaporthe* (*Pyricularia*) *oryzae* from rice. *Phytopathology* (Submitted).
- 697 Onaga, G., & Asea, G. (2016). Occurrence of rice blast (*Magnaporthe oryzae*) and  
698 identification of potential resistance sources in Uganda. *Crop Protection*, 80, 65–72.
- 699 Onaga, G., Bigirimana, J., Murori, R., Cruz, C. M. V., Oliva, R., & Séré, Y. (2019).  
700 Section 4. Importance of Plant Diseases on Rice Production in Africa. In *Rice diseases: Biology*  
701 *and selected management practices*. International Rice Research Institute. [http://rice-](http://rice-diseases.irri.org)  
702 [diseases.irri.org](http://rice-diseases.irri.org).
- 703 Paradis, E. (2010). pegas: An R package for population genetics with an integrated–  
704 modular approach. *Bioinformatics*, 26(3), 419–420.
- 705 Savary, S., Willocquet, L., Elazegui, F. A., Castilla, N. P., & Teng, P. S. (2000). Rice  
706 pest constraints in tropical Asia: Quantification of yield losses due to rice pests in a range of  
707 production situations. *Plant Disease*, 84(3), 357–369.
- 708 Séré, Y., Fargette, D., Abo, M. E., Wydra, K., Bimerew, M., Onasanya, A., & Akator, S.  
709 K. (2013). 17 Managing the Major Diseases of Rice in Africa. *Realizing Africa's Rice Promise*,  
710 213.
- 711 Silue, D. (1991). Resistance of 99 *O. glaberrima* varieties to blast. *Int. Rice Res. Notes*,  
712 16, 13–14.
- 713 Small, W. (1922). Annual report of the government mycologist. *Uganda Dept. Agr. Ann.*  
714 *Rept*, 27–29.
- 715 Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-  
716 analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.
- 717 Thierry, M., Gladieux, P., Fournier, E., Tharreau, D., & Ioos, R. (2020). A genomic  
718 approach to develop a new qPCR test enabling detection of the *Pyricularia oryzae* lineage  
719 causing wheat blast. *Plant Disease*, 104(1), 60–70.
- 720 Valent, B., Crawford, M. S., Weaver, C. G., & Chumley, F. G. (1986). Genetic studies of  
721 fertility and pathogenicity in *Magnaporthe grisea* (*Pyricularia oryzae*). *Iowa State J. Res.*, 60(4),  
722 569–594.

723 Wang, D., Zhang, Y., Zhang, Z., Zhu, J., & Yu, J. (2010). KaKs\_Calculator 2.0: A toolkit  
724 incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics &*  
725 *Bioinformatics*, 8(1), 77–80.

726 Yang, Z., Nielsen, R., Goldman, N., & Pedersen, A.-M. K. (2000). Codon-substitution  
727 models for heterogeneous selection pressure at amino acid sites. *Genetics*, 155(1), 431–449.

728 Yelome, O. I., Audenaert, K., Landschoot, S., Dansi, A., Vanhove, W., Silue, D.,  
729 Damme, P. V., & Haesaert, G. (2018). Combining high yields and blast resistance in rice (*Oryza*  
730 *spp.*): A screening under upland and lowland conditions in Benin. *Sustainability*, 10(7), 2500.

731 Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., Tosa,  
732 Y., Chuma, I., Takano, Y., & Win, J. (2009). Association genetics reveals three novel avirulence  
733 genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *The Plant Cell*, 21(5), 1573–  
734 1591.

735 Yoshida, K., Saunders, D. G., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., Inoue,  
736 Y., Chuma, I., Tosa, Y., & Cano, L. M. (2016). Host specialization of the blast fungus  
737 *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable  
738 elements. *BMC Genomics*, 17(1), 370.

739 Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: An R package  
740 for visualization and annotation of phylogenetic trees with their covariates and other associated  
741 data. *Methods in Ecology and Evolution*, 8(1), 28–36.

742 Zhong, Z., Chen, M., Lin, L., Han, Y., Bao, J., Tang, W., Lin, L., Lin, Y., Somai, R., &  
743 Lu, L. (2018). Population genomic analysis of the rice blast fungus reveals specific events  
744 associated with expansion of three main clades. *The ISME Journal*, 12(8), 1867–1878.

745  
746  
747