Open access • Posted Content • DOI:10.1101/2020.06.02.129296

Ecological Differentiation Among Globally Distributed Lineages of the Rice Blast Fungus Pyricularia Oryzae

Maud Thierry, Florian Charriat, Joëlle Milazzo, Henri Adreit ...+9 more authors

Institutions: ANSES

Published on: 28 Oct 2021 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Pyricularia

Ecological Differentiation and Incipient Speciation in the Fungal Pathogen Causing Rice Blast

Maud THIERRY[1,2,3], Florian CHARRIAT[1], Joëlle MILAZZO[1,2], Henri ADREIT[1,2], Sébastien RAVEL[1,2], Sandrine CROS-ARTEIL[1], Sonia BORRON[1], Violaine SELLA[3], Thomas KROJ[1], Renaud IOOS[3], Elisabeth FOURNIER[1], Didier THARREAU[1,2,4], Pierre GLADIEUX[1,4]

[1]PHIM Plant Health Institute, Univ Montpellier, INRAE, CIRAD, Institut Agro, IRD, Montpellier, France

[2]CIRAD, UMR PHIM, 34090 Montpellier, France.

[3]ANSES Plant Health Laboratory, Mycology Unit, Domaine de Pixérécourt, Bâtiment E, F-54220 Malzéville, France

[4]pierre.gladieux@inrae.fr; didier.tharreau@cirad.fr

ABSTRACT

Many invasive fungal species coexist as multiple populations on the same host, but the factors behind the origin and maintenance of population structure remain largely unknown. Here, we analyzed genetic and phenotypic diversity in isolates of the rice blast fungus (*Pyricularia oryzae*) covering a broad geographical range. We showed that the four lineages of *P. oryzae* were found in areas that differ in terms of prevailing environmental conditions and types of rice grown. Pathogenicity tests revealed that specialization to rice subspecies contributed to niche separation between lineages, and differences in repertoires of putative pathogenicity effectors were consistent with differences in host range. Crossing experiments revealed that female sterility and early post-mating genetic incompatibilities acted as strong barriers to gene flow between these lineages. Our results demonstrate that the spread of a pathogen across heterogeneous habitats and divergent populations of a crop species can lead to niche separation and incipient speciation.

Short title: Incipient speciation in the rice blast pathogen

INTRODUCTION

Understanding and controlling natural variation in fungal plant pathogens is of great importance for maintaining food security and ecosystem health. In addition to the overwhelming phylogenetic diversity of pathogens (Burdon 1993), many pathogens also harbor substantial diversity at the species scale (Taylor & Fisher 2003). Most fungal pathogens of plants form distinct populations (Taylor *et al.* 2006), and this population structure is a compound outcome of migration, selection and drift, acting at multiple scales, and patterned by factors such as time (Ali *et al.* 2014; Pagliaccia *et al.* 2018), admixture (Robert *et al.* 2015), geography (Ali *et al.* 2014; Marin *et al.* 2009; Yang *et al.* 2018), mode of reproduction (Ali *et al.* 2014; Bueker *et al.* 2016; Gibson *et al.* 2012; Ropars *et al.* 2016) and environmental conditions (Walker *et al.* 2015). The identity of the host plants with which fungal pathogens interact is widely thought to be the environmental factor structuring the pathogen population with the strongest impact. Many plant pathogens reproduce within or on their host, resulting in assortative mating on the basis of host use and a strong association between adaptation to the host and reproductive isolation (Giraud *et al.* 2010; Servedio *et al.* 2011). However, more generally, population structure can result from limited dispersal (i.e. limited gene flow due to distance) or limited adaptation (i.e. limited gene flow because of differences in the capacity to exploit resources), both of which may be due to a wealth of potential factors largely unexplored in plant pathogens. Developing an understanding of how population structure emerges and is maintained in fungal plant pathogens is a major goal in evolutionary microbiology, because it will improve our comprehension of both disease emergence and the origin of fungal biodiversity.

Pyricularia oryzae (Ascomycota; syn. *Magnaporthe oryzae*) is a widespread model plant pathogen displaying population subdivision. Several lineages, each associated mostly with a single main cereal/grass host, have been characterized within *P. oryzae* (Gladieux *et al.* 2018a). The lineage infecting Asian rice (*Oryza sativa*) has been characterized in detail. Previous population genomic studies of population structure in the rice-infecting lineage revealed genetic subdivision into three (Saleh *et al.* 2014; Tharreau *et al.* 2009; Zhong *et al.* 2018) to six (Gladieux *et al.* 2018b) lineages,

51 estimated to have diverged approximately 1000 years ago (Gladieux *et al.* 2018b; Latorre *et al.* 2020;
52 Zhong *et al.* 2018). *Pyricularia oryzae* is a heterothallic fungus with a sexual cycle involving mating
53 between individuals of opposite mating types, with at least one of the partners capable of producing
54 female structures (i.e. "female-fertile"). One single lineage (lineage 1, prevailing in East and Southeast
55 Asia) displayed a genome-wide signal of recombination, a balanced ratio of mating type alleles and a
56 high frequency of fertile females, consistent with sexual reproduction (Gladieux *et al.* 2018b; Saleh *et*
57 *al.* 2014; Saleh *et al.* 2012b). All the other lineages had clonal population structures, highly biased
58 frequencies of mating types and few fertile females, suggesting strictly asexual reproduction
59 (Gladieux *et al.* 2018b; Saleh *et al.* 2014). In theory, gene flow remains possible between the majority
60 of the lineages, because there are rare fertile strains and some lineages have fixed opposite mating
61 types. However, while resequencing studies have revealed signatures of recent shared ancestry
62 between clonal lineages and the recombinant lineage from South-West Asia, this is not the case
63 between the different clonal lineages, consistent with a lack of recent gene flow between lineages of
64 opposite mating types and despite their wide and overlapping distributions (Gladieux *et al.* 2018b).
65 The loss of female fertility is probably a key factor in the emergence of clonal lineages, inducing a
66 shift towards strictly asexual reproduction, and protecting populations adapting to new conditions
67 from maladaptive gene flow from ancestral, recombining populations (Giraud *et al.* 2010). This raises
68 questions about the nature of the factors underlying the emergence of different rice-infecting
69 lineages in *P. oryzae*, and contributing to their maintenance, because, according to the competitive
70 exclusion principle, different lineages should not be able to co-exist on the same host (Abbate *et al.*
71 2018; Giraud *et al.* 2017; Hardin 1960). Previous studies have helped to cement the status of *P. oryzae*
72 as a model for studies of the population biology of fungal pathogens, but most efforts to understand
73 the population structure of the pathogen have been unable to provide a large-scale overview of the
74 distribution of rice-infecting lineages and of the underlying phenotypic differences, because the
75 number of genetic markers was limited (Saleh *et al.* 2014; Tharreau *et al.* 2009), the number of

76 isolates was relatively small (Gladieux *et al.* 2018b; Zhong *et al.* 2018) and because few phenotypic
77 traits were scored (Saleh *et al.* 2014; Tharreau *et al.* 2009; Zhong *et al.* 2018).

78 We provide here a detailed genomic and phenotypic overview of natural variation in rice-
79 infecting *P. oryzae* isolates sampled across all rice-growing areas of the world. With a view to
80 inferring and understanding the global population structure of rice-infecting *P. oryzae*, we analyzed
81 genetic variation, using genotyping data with a high resolution in terms of genomic markers and
82 geographic -coverage, and measured phenotypic variation for a representative subset of the sample
83 set. We also characterized the genetic variability and content of repertoires of effectors –secreted
84 proteins that modulate host responses – using whole-genome resequencing data for a representative
85 set of isolates. We addressed the following specific questions: (i) How many different lineages of *P.*
86 *oryzae* infect rice? (ii) Do lineages of *P. oryzae* display footprints of recombination? (iii) Do lineages of
87 *P. oryzae* present evidence of recent admixture? (iv) Are the different lineages sexually competent
88 and compatible? (v) Are the different lineages differentially adapted to host and temperature? (vi)
89 Can we find differences in number, identity and molecular evolution of putative effectors between
90 lineages? (vii) Can we find differences in the geographic and climatic distribution of lineages?

RESULTS

The rice blast pathogen comprises one recombining, admixed lineage and three clonal, non-admixed lineages

We characterized the global genetic structure of the rice blast pathogen, using an Illumina genotyping beadchip to score 5,657 SNPs distributed throughout the genome of 886 *P. oryzae* isolates collected from cultivated Asian rice in 51 countries (Supplementary file 1). The final dataset included 3,686 SNPs after the removal of positions with missing data, which identify 264 distinct multilocus genotypes out of 886 isolates. Clustering analyses based on sparse nonnegative matrix factorization algorithms, as implemented in the sNMF program (Frichot et al., 2014), revealed four clusters, hereafter referred to as "lineages" (Figure 1C). The model with K=4 clusters was identified as the best model on the basis of the cross-entropy criterion, as models with K>4 induced only a small decrease in cross-entropy, suggesting that K=4 captured the deepest subdivisions in the dataset (Figure 1 – figure supplement 1). The neighbor phylogenetic network inferred with SPLITSTREE also supported the subdivision into four widely distributed lineages, with long branches separating three peripheral lineages branching off within a central lineage (Figure 1A; 1B). A comparison with previous findings revealed that the central lineage corresponded to the combination of recombining lineage 1 and lineages 5 and 6, represented by two and one individuals, respectively, in (Gladieux *et al.* 2018b) (Supplementary file 2). This central lineage is referred to hereafter as lineage 1. The three peripheral lineages in the network corresponded to the previously described lineages 2, 3 and 4 (Gladieux *et al.* 2018b) with lineages 2 and 3 similar to lineages B and C, respectively, identified on the basis of microsatellite data in a previous study (Saleh *et al.* 2014)(Supplementary file 2). Genetic differentiation between the four lineages was strong and significant (Weir and Cockerham's F_{ST}> 0.54), indicating strong barriers to gene flow between the lineages. All genotypes from lineages 2 to 4 had high proportions of membership q in a single cluster (q > 0.89), whereas shared ancestry with the other three lineages was detected in lineage 1, with 32% of genotypes having q > 0.10 in lineages 2-4

(Figure 1C; Figure 1 – source data 3). Admixture may account for the lower F_{ST} observed in comparisons between lineage 1 and the other lineages. We detected no footprints of recombination in lineages 2, 3 and 4 with the pairwise homoplasy index (PHI) test and a null hypothesis of clonality (Table 1), confirming previous findings (Gladieux *et al.* 2018b; Latorre *et al.* 2020; Saleh *et al.* 2014; Zhong *et al.* 2018).

Geographic structure and admixture within recombining lineage 1

Lineage 1 was the only lineage for which there was population genetic evidence of recombination in the PHI test (Table 1). Most of the lineage 1 isolates were collected in Asia (79%), but the lineage was present in all continents in which the pathogen was sampled (Europe: one isolate; North America: 10, Central and South America: 7, Africa: 22) (Figure 2). Clustering analysis on this single lineage with sNMF detected four clusters with different geographic distributions (Figure 2). Estimates of differentiation were lower between the clusters within lineage 1 (F_{ST} < 0.49) than between the main lineages (F_{ST} > 0.54), consistent with a longer history of restricted gene flow between the main lineages and a more recent history of admixture between clusters within lineage 1 (Supplementary file 3; Figure 1C).

Two clusters within lineage 1 (referred to hereafter as *Baoshan* and *Yule*) consisted mostly of isolates sampled from two different sites in the Yunnan province of China (Baoshan and Yule, respectively, located about 370 km apart). The third cluster consisted mostly of isolates from Laos and South China (referred to hereafter as *Laos*), and the fourth brought together 95% of the isolates from lineage 1 collected outside Asia (referred to hereafter as *International*). In Asia, the *International* cluster was found mostly in the Yunnan province of China, India and Nepal (Figure 2). PHI tests for recombination rejected the null hypothesis of clonality in all four clusters (Table 1). Admixture was widespread in lineage 1 (Figure 2), with most of the isolates (78%) displaying membership proportions q > 0.10 for two or more clusters (Figure 2). Only five genotypes were detected in multiple countries (genotype ID [number of countries]: 2 [6], 18 [2], 58 [2], 98 [3] and 254 [3],

142 corresponding to 41 isolates in total). All these genotypes belonged to the *International* cluster.

143

144 Table 1. PHI test for recombination (null hypothesis: clonality) and distribution of mating types and fertile females in
145 lineages, and in clusters within lineage 1. PHI tests were carried out using source files Figure 1 – source data 1 and Figure 2 –
146 source data 1. Information about mating types and female fertility is available in Supplementary file 1.

Lineage/*Cluster*	*p*-value PHI test	Mat1.1/Mat1.2 (%)	Female-fertile isolates (%)
1	<0.0001	53/47	43
Baoshan	<0.0001	69/31	67
International	<0.0001	55/45	4
Laos	<0.0001	68/32	37
Yule	<0.0001	40/60	79
2	0.25	97/3	0
3	0.18	3/97	7
4	0.22	97/3	0

147

148 **Reproductive barriers caused by female sterility and postmating genetic incompatibilities**

149 Our analysis of genotyping data revealed the existence of a large, ancient, recombinant population
150 from which several clonal lineages have more recently emerged. This raised the question as to why
151 these populations became clonal, with no further gene flow. We addressed this question by
152 determining the capacity of the different lineages to engage in sexual reproduction, by characterizing
153 the distribution of mating types and the ability to produce female structures (i.e. female fertility).
154 Using *in vitro* crosses with tester strains, we found that lineages 2, 3 and 4 were composed almost
155 exclusively of a single mating type (97% of lineage 2 and 4 isolates tested carried the *Mat1.1* allele;
156 97% of lineage 3 isolates carried the *Mat1.2* allele), and had only small proportions (0-7%) of isolates
157 with female fertility (Table 1). By contrast to these clonal lineages, the mating type ratio was more
158 balanced in lineage 1 (52% of *Mat1.1*; Table 1), with *Mat1.1*/*Mat1.2* ratios ranging from 40/60 in the
159 *Yule* cluster to 69/31 in the *Baoshan* cluster. Female fertility rates were higher in three of the clusters
160 in lineage 1 (*Yule*: 79%; *Baoshan*: 67%; *Laos*: 37%; Table 1), but that in the International cluster (4%

female fertility) was as low as those in the clonal clusters (Table 1). These observations indicate that there is a potential for sexual reproduction between lineages with compatible mating types and, albeit rare, fertile females, but not between lineages 2 and 4, due to the low female fertility rates and highly biased mating type ratios.

We further assessed the likelihood of sexual reproduction within and between lineages, by evaluating the formation of sexual structures (perithecia), the production of asci (i.e. meiotic octads within the perithecia) and the germination of ascospores (i.e. meiospores in meiotic octads) in crosses between randomly selected isolates of opposite mating types from all four lineages (Figure 3). This experiment revealed a marked heterogeneity in the rate of perithecium formation across lineages. Three clusters in lineage 1 had the highest rates of perithecia formation, with isolates in the *Yule* cluster, in particular, forming perithecia in 93% of crosses with isolates from the same cluster, and in more than 46% of crosses with isolates from other lineages (Figure 3A). Isolates from lineages 2, 3 and 4 could only be crossed with isolates from other lineages, given their highly biased mating type ratios, and the percentage of these crosses producing perithecia was highly variable (ranging from 0% to 83%, depending on the lineages involved; Figure 3A). The rate of perithecium formation was similar in crosses involving the *International* cluster of lineage 1 and in crosses involving clonal lineages 2-4, and none of the crosses attempted between isolates from the *International* cluster led to perithecium formation (Figure 3A). Perithecium dissection for a subset of crosses involving some of the most fertile isolates revealed that most inter-lineage crosses produced perithecia that did not contain asci or contained asci with low ascospore germination rates (Figure 3B). Whereas 100% of crosses between isolates of the *Yule* cluster produced numerous germinating ascospores, ascospore germination rates were only 33%, 56% and 7% in *Yule* x lineage 2, *Yule* x lineage 3 and *Yule* x lineage 4 crosses, respectively. Together, these results indicate that the clonal structure of lineages 2-4 is caused by highly biased mating types and female sterility, and that the three clonal lineages and the *International* cluster of lineage 1 are isolated from each other by strong pre- and early-postmating barriers, including breeding system isolation (differences in mating type and female sterility), and a

187 combination of gametic incompatibility, zygotic mortality or hybrid non-viability. However, three of
188 the clusters in lineage 1 had biological features consistent with an ability to reproduce sexually,
189 suggesting possible sexual compatibility with clonal lineages 2-4 and the *International* cluster.

190 **Specialization to temperature conditions and rice subspecies**

191 Specialization can further reduce gene flow, and may facilitate niche separation between lineages,
192 which should favor the emergence of new lineages and their maintenance on the same host species
193 in conditions of competitive exclusion. The strong reproductive barriers demonstrated in our analysis
194 of crossing experiments would favor the establishment of such specialization, because reproductive
195 isolation facilitates adaptation to new conditions, by preventing maladaptive gene flow from
196 ancestral, non-adapted populations (Giraud *et al.* 2010). Given the very broad distribution of *P.*
197 *oryzae* and the strong environmental heterogeneity with which it is confronted in terms of the
198 nature of its hosts and the climate in which it thrives, we evaluated the variation of pathogen fitness
199 over different types of rice and temperature conditions.

200 We measured growth rate and sporulation rate of representative isolates cultured at different
201 temperatures to test the hypothesis of adaptation to temperature. Experiments were conducted with
202 41 isolates (lineage 1 [yule cluster]: 11; lineage 2: 10; lineage3: 10; lineage 4: 10). For all lineages,
203 mycelial growth rate increased with incubation temperature, although this trend was more visible
204 from 10°C to 15°C (increased mean mycelial growth of +2.22 mm/day) than from 25°C to 30°C (+0.05
205 mm/day) (growth curves and full statistical treatment of data are presented in Supplementary file 4;
206 data are reported in Supplementary file 5). Fitting a linear mixed-effects model with incubation time,
207 experimental replicate, and lineage of origin as fixed effects and isolate as a random effect, revealed a
208 significant lineage effect at each incubation temperature [10°C: $F(1,364)=7988$, $p<0.001$; 15°C:
209 $F(1,419)=33161$, $p<0.001$; 20°C: $F(1,542)=40335$, $p<0.001$; 25°C: $F(1,413)=30156$, $p<0.001$; 30°C:
210 $F(1,870)=52681$, $p<0.001$]. Comparing least-squares means resulting from linear mixed-effects
211 models at each temperature revealed a significantly lower growth rate of lineage 4 at 10°C compared

to other lineages and a significantly higher growth rate of lineage 1 at 15°C, 20°C, 25°C and 30°C compared to other lineages. As regards sporulation rates, sporulation was nearly absent at 10°C after 20 days of culture, with the few spores observed being not completely formed and divided by only one septum instead of two septa in mature conidia (sporulation curves and full statistical treatment of data are presented in Supplementary file 6; data are reported in Supplementary file 7). For all lineages, sporulation rates increased with temperature from 15 to 25 °C and dropped at 30°C, although isolates were cultured 7 days only at this latter temperature instead of 10 days at other temperatures. Isolates were cultured 7 days only at 30°C because at this stage the mycelium had reached the edges of the plates. Significant lineage effects were observed at 10°C, 15°C and 30°C (Kruskal-Wallis tests; 10°C: $H(3)=9.61$, $p=0.022$; 15°C: $H(3)=16.8$, $p<0.001$; 30°C: $H(3)=8.96$, $p=0.030$). Pairwise non-parametric comparisons using Dunn's test of lineages revealed significant differences between lineages 1 and 2 at 10°C, between lineage 1 and lineages 3 and 4 at 15°C, and between lineage 1 and 4 at 30°C. Together, measurements of mycelial growth and sporulation at different temperatures revealed differences in performance between lineages, but no clear pattern of temperature specialization.

We assessed the importance of adaptation to the host, by challenging 45 rice varieties representative of the five main genetic groups of Asian rice (i.e. the temperate japonica, tropical japonica, aus, aromatic and indica subgroups of *Oryza sativa*) with 70 isolates representative of the four lineages of *P. oryzae* and the four clusters within lineage 1 (Supplementary file 8). Interaction phenotypes were assessed qualitatively by scoring resistance (from full to partial resistance: scores 1 to 3) or disease symptoms (from weak to full susceptibility: scores 4 to 6) and analysed by fitting a proportional-odds model, followed by an analysis of variance (full statistical treatment of data is presented in Supplementary file 9). This analysis revealed significant differences between groups of isolates ($\chi^2(6)=100$, $p<0.001$), and between rice genetic groups ($\chi^2(4)=161$, $p<0.001$), and a significant interaction between these two variables ($\chi^2(24)=97$, $p<0.001$). The finding of a significant interaction between groups of isolate (lineage or cluster) and rice types indicates that the effect of the group of

238 isolates on the proportion of compatible interactions differed between rice types, suggesting
239 adaptation to the host. This is also supported by the specific behaviour of certain lineages. Isolates
240 from lineage 2 showed e.g. much lower symptom scores than the other lineages on all rice types
241 except temperate japonica and the isolates of the *Yule* cluster possess particularly high virulence on
242 indica varieties (Figure 4A; Supplementary file 8). In comparisons of rice genetic groups, significantly
243 higher symptom scores were observed on temperate japonica rice than on the other types of rice,
244 whereas the varieties of the aromatic genetic group were significantly more resistant to rice blast
245 (Figure 4B). Together, these experiments therefore revealed significant differences in host range
246 between lineages, but this specialization to the host is not strict since host ranges overlap (Figure 4C-
247 G). Adaptation to the host or small differences in temperature optima could, nonetheless, further
248 increase niche separation and reduce gene flow between lineages, via pre-mating barriers such as
249 immigrant non-viability (i.e., lower rates of contact between potential mates due to the mortality of
250 immigrants; (Gladieux *et al.* 2011; Nosil *et al.* 2005)), and post-mating barriers, such as ecological
251 hybrid non-viability (i.e., lower survival of ill-adapted hybrid offspring).

252 **Differences in number and genetic variability of putative effector genes among *P. oryzae* lineages**

253 Small secreted proteins named effectors are the most prominent class of fungal virulence factors.
254 They act by manipulating host cellular pathways and are key determinants of the host range and
255 fitness on compatible hosts (Liao *et al.* 2016; Yoshida *et al.* 2016). We therefore determined the
256 differences of effector repertoires in *P. oryzae* lineages in terms of presence/absence and nucleotide
257 polymorphism, and analyzed whether it is a particularly dynamic fraction of the gene space. For this
258 aim, we used whole-genome sequencing data for 123 isolates, including 29 isolates sequenced in this
259 study and 94 publicly available genomes (Gladieux *et al.* 2018a; Pordel *et al.* 2020; Zhong *et al.* 2018),
260 of which 33 were also genotyped using our Infinium genotyping beadchip (Supplementary file 10;
261 Appendix 1). Clustering analysis indicated that this dataset covered the four lineages of *P. oryzae*, and
262 three of the clusters of lineage 1 (Laos, International and Baoshan; Appendix 1; Supplementary file

263 10). Assembly lengths for the 123 sequenced isolates varied from 36.7 to 39.6 Mb, with an average of
264 38.1 Mb (Supplementary file 10). The number of assembled contigs longer than 500bp ranged from
265 1010 to 2438, and the longest contig was 1.1 Mb (Supplementary file 10).

266 Gene prediction identified 11,684 to 12,745 genes per genome, including 1,813 to 2,070
267 genes predicted to code for effector proteins. We found a significant effect of the lineage of origin on
268 both the mean number of putative effector genes (ANOVA, $F=5.54$, $p=0.0014$), and the mean number
269 of non-effector genes (ANOVA, $F=7.95$, $p<0.001$). Multiple comparisons revealed that mean numbers
270 of putative effectors and non-effector genes were only significantly different between genomes of
271 lineage 3 and lineages 2 (Tukey's HSD test; $p<0.001$; Supplementary file 11). Orthology analysis
272 identified 14,573 groups of orthologous sequences (i.e. orthogroups), and we applied a rarefaction
273 approach to the table of orthology relationships to estimate the size of accessory and core genomes
274 in lineages while accounting for differences in sample size. For both putative effectors and the
275 remaining of the gene space, we found that the number of core and accessory genes did not plateau
276 and changed almost linearly with the number of genomes resampled, indicating that the number of
277 core genes is likely substantially smaller, and the number of accessory genes substantially higher, than
278 estimated from 123 genomes (Supplementary file 12). With pseudo-samples of size n=30 per lineage
279 (i.e. excluding lineage 4, due to small sample size), the gene content was highly variable within
280 lineages with only 61-71% of all predicted effectors, and 68-73% of the remaining gene space, being
281 conserved in lineages 1-3 (Figure 5A). Despite extensive variation in gene content within lineages,
282 clustering of isolates based on presence/absence variation for both putative effectors and the
283 remaining of the gene space was highly similar to clustering based on 3,868 SNPs (Supplementary file
284 13), indicating that variation in gene content and SNP allelic variation reflect similar genealogical
285 processes.

286 To identify putative effectors that may be involved in host specialization of rice-infecting
287 lineages of *P. oryzae*, we first identified orthogroups with distinct patterns of presence/absence

288 across lineages. Principal component analysis of presence/absence data identified 72 orthogroups of
289 putative effectors that explained 95% of the variance of the three principal components that
290 differentiate the four lineages (PC1, PC2 and PC6; Supplementary file 13; Figure 5B), and may
291 contribute to the differences in host range among lineages. Interestingly, two among them, AvrRmg8
292 (=OG0011611) and PWL3 (=OG0011928), are host range determinants in wheat-infecting *P. oryzae*
293 since they trigger immunity in wheat varieties carrying the resistance genes *Rmg8* (Anh *et al.* 2018) or
294 *PWT3* (Inoue *et al.* 2017; Kang *et al.* 1995).

295 As host specialization can also involve sequence divergence at effector proteins, we also
296 scanned the corresponding genes for signatures of diversifying selection, i.e. an excess of non-
297 synonymous nucleotide diversity (ratio of non-synonymous to synonymous nucleotide diversity π_N/π_S
298 >1). We identified 185 orthogroups with $\pi_N/\pi_S > 1$ in at least one lineage, including 164 orthogroups
299 with $\pi_N/\pi_S > 1$ in only one lineage (lineage 1: 131 orthogroups; lineage 2: 18; lineage 3: 10; lineage 4:
300 5), and twelve, seven and two orthogroups with $\pi_N/\pi_S > 1$ in two, three or four lineages, respectively
301 (Supplementary file 14). None of these orthogroups corresponded to effectors previously
302 characterized as being involved in *P. oryzae* virulence against rice or other *Poaceae* hosts.

303 **Geographic and climatic differentiation in the distribution of *P. oryzae* lineages**

304 Our tests of adaptation to host and temperature under controlled conditions showed that
305 specialization was not strict, but it remained possible that fitness differences between lineages would
306 be sufficient under natural conditions to induce separation in different ecological niches, and/or that
307 our experimental conditions did not capture the full extent of the phenotypic differences. Here, we
308 tested the hypothesis that lineages thrive in geographically and/or environmentally different
309 conditions, which would provide indirect evidence for specialization to different habitats. We tested
310 this hypothesis by collecting geographic and climatic data for all isolates, approximating to regions or
311 the nearest cities when the exact GPS position of the sampling area was not available (Supplementary
312 file 1). At a larger scale, clonal lineages 2 and 3 were both widespread, with lineage 2 found on all

313 continents, and lineage 3 on all continents except Europe. Lineage 2 was the only lineage sampled in
314 Europe (with the exception of a single isolate from lineage 1), whereas lineage 3 was more
315 widespread in intertropical regions. Lineage 4 was mostly found in South Asia (India, Bangladesh,
316 Nepal), and in the USA and Africa (Benin, Tanzania). The recombining lineage, lineage 1, was found in
317 all continents, but was mostly found in Asia (79% of all isolates and 94% of all genotypes). At a
318 smaller scale, two, or even three different lineages were sampled in the same year in the same city in
319 11 different countries from all continents. However, in only one instance did two isolates from
320 different lineages have identical GPS positions (isolates US0106 and US0107 sampled from the same
321 field). Thus, lineages had different, but partially overlapping distributions, including some overlap
322 with the distribution of the sexually reproducing lineage (lineage 1).

323 We then investigated whether differences in the geographic range of lineages were
324 associated with differences in climatic variables. By plotting sampling locations onto a map of the
325 major climate regions (Kottek *et al.* 2006) we were able to identify clear differences in the climatic
326 distributions of lineages 2 and 3, with lineage 2 mostly found in warm temperate climates and lineage
327 3 found in equatorial climates (Figure 6; Source code 1). We used the outlying mean index (OMI),
328 which measures the distance between the mean habitat conditions used by a lineage and the mean
329 habitat conditions used by the entire species, to test the hypothesis that different lineages are
330 distributed in regions with different climates. Using information for 19 climatic variables (biomes)
331 from the WorldClim bioclimatic variables database (Fick & Hijmans 2017) for all sampling locations,
332 we obtained statistically significant results, in a permutation test on OMI values, for lineages 2, 3 and
333 4 (Source code 2; permutation test: lineage 1: OMI=2.14, p=0.620; lineage 2: OMI=6.54, p<0.001;
334 lineage 3: OMI=2.08, p<0.001; lineage 4: OMI=13.73, p=0.017). The OMI analysis, in which the first
335 two axes accounted for 69% and 25% of the variability, respectively, revealed that lineage 2 was more
336 frequent in regions with a high annual temperature range (biome 7) or a high degree of seasonality
337 (biome 4); lineage 4 was associated with regions with high levels of seasonal precipitation (biomes
338 13, 16 and 18), and lineage 3 was more frequent in regions with high temperatures (biomes 1, 6, 10

339 and 11) and high levels of isothermality (biome 3), characteristic of tropical climes (Source code 2;
340 Figure 6; Supplementary file 15).
341
342 **DISCUSSION**

343 We describe here the population structure of the rice blast fungus, and the eco-evolutionary factors
344 underlying the emergence and maintenance of multiple divergent lineages in this widespread
345 pathogen. Our analysis of SNP genotyping data showed that *P. oryzae* could be subdivided into four
346 lineages. Two of the lineages previously detected in an analysis of whole-genome data, and
347 represented by only a few individuals (lineages 5 and 6; (Gladieux *et al.* 2018b)), were assigned to
348 lineage 1 in our analysis, and may correspond to a subdivision of lineage 1. Differences in the number
349 of lineages detected may reflect an ascertainment bias, given that our SNP-genotyping beadchip was
350 designed with a set of genomic sequences that include few representatives of lineages 5 and 6.
351 Alternatively, the sampling used in the previous study (Gladieux *et al.* 2018b) may not have been
352 dense enough to reveal the lack of differentiation between lineages 1, 5 and 6, making them appear
353 to be different entities. Consistent with previous findings, our analysis of the genealogical
354 relationships between isolates revealed three clonal groups connected by long branches to a central
355 recombining group. The finding of an internationally distributed cluster in lineage 1 sheds light on the
356 early evolutionary changes contributing to the emergence of clonal lineages in this pathogen. The
357 *International* cluster of lineage 1 displayed genetic footprints of recombination and sexual
358 reproduction (in the form of an excess of homoplasious variants or relatively balanced mating types),
359 but the signal of recombination detected here may be purely historical, because we also found that
360 the clonal fraction and the frequency of sterile females were very high in this cluster. The loss of
361 female fertility may, therefore, be an early and major cause of the shift towards asexual reproduction.
362 Some genotypes assigned to the international cluster of lineage 1 may, in the future, come to
363 resemble other clonal lineages in analyses of genealogical relationships, and form a long branch
364 stemming from the central lineage 1, particularly for genotypes already displaying clonal propagation,

such as genotypes 2, 18, 58, 98 and 254. However, according to the principle of competitive exclusion, competition between clonal lineages should only allow hegemonic expansion of the most competitive clonal groups, unless they separate into different ecological niches. It may be this process that ultimately leads to the fixation of a single mating type, as in clonal lineages 2-4.

We found that female sterility and intrinsic genetic incompatibilities represented strong barriers to gene flow between the clonal lineages, potentially accounting for their maintenance over time, without fusion, despite the compatibility of mating-types (at least between lineages 2 and 4 [*Mat1.1*] on the one hand, and lineage 3 [*Mat1.2*] on the other). These barriers to gene flow may have contributed to the establishment of lineage specialization (because reproductive isolation facilitates adaptation by limiting the flow of detrimental ancestral alleles in populations adapting to a new habitat), accounting for the maintenance of the clonal lineages on the same host species, when faced with competitive exclusion. Our analyses also show that, despite widely overlapping large-scale distributions, the different lineages are essentially found in different regions, with different climatic characteristics, when their distributions are observed at a finer scale. Lineage 1 was mostly sampled in Southeast Asia, lineage 4 predominated in India, and lineages 2 and 3 had global distributions, but the geographic ranges of all these lineages overlapped. However, an analysis of climatic data indicated that lineage 2 predominated in temperate climates, where temperate japonica rice is grown, lineage 3 predominated in tropical climates, in which indica rice is grown, and lineage 4 predominated in regions with high levels of seasonal precipitation in which the indica and aromatic rice types are the principal rice types grown. Despite the finding of separation in different climatic regions, our experiments revealed no strong differences between lineages in terms of sporulation and mycelial growth on synthetic media at different temperatures. This suggests that, if adaptation to temperature occurs in this pathogen, it was not measured by our experiments, either because it involves traits other than sporulation and hyphal growth, or because the *in vitro* conditions were not suitable to demonstrate differences. The host range varied across lineages, but all host ranges overlapped, indicating that host specialization was not strict, or was not fully captured by our

391 experiments. Adaptation to the host may, nevertheless, decrease gene flow between populations
392 even further. For instance, lineage 2 had a narrow host range, potentially limiting the opportunities
393 for encounters and mating with other groups.

394 Effector repertoires play a key role in pathogen specialization, in fungal pathogens in general
395 (Schulze-Lefert & Panstruga 2011), and *P. oryzae* in particular (Asuke *et al.* 2020; Couch *et al.* 2005;
396 Liao *et al.* 2016; Sweigard *et al.* 1995; Takabayashi *et al.* 2002). Previous work suggested that clonal
397 lineages of *P. oryzae* possess smaller effector repertoires (Latorre *et al.* 2020) and that the clonal
398 lineage associated with japonica rice (lineage 2) possess more effectors, and in particular Avr-
399 effectors, than other clonal lineages (Latorre *et al.* 2020; Liao *et al.* 2016) . Our analyzes on a larger
400 set of putative effectors (ca. 2000 in our study, vs 13 Avr-effectors in (Liao *et al.* 2016) and 178
401 known and candidate effectors in (Latorre *et al.* 2020)) confirm this trend, since lineage 2 associated
402 with temperate japonica exhibits significantly more putative effectors than lineage 3 associated with
403 indica. However, the number of non-effector genes is also greater in lineage 2 than in lineage 3, and
404 thus it remains possible that the larger effector complement of lineage 2 is a mere consequence of a
405 larger number of genes. Besides differences between lineages in richness of putative effector
406 repertoires, we show that patterns of presence / absence variation mirror patterns of population
407 subdivision based on SNP allelic variation, and thus that the differential sorting of both nucleotide
408 polymorphism and gene content across lineages reflect similar genealogical processes. Our
409 multivariate analyzes identify 72 effectors that most strongly contribute to the differentiation of the
410 4 lineages in terms of presence / absence. It is possible that the frequency differences observed at
411 these 72 effectors were due to chance events during bottlenecks at the onset of lineage formation,
412 or alternatively that their differential loss was important for the initial adaptation of lineages to new
413 rice populations or subspecies. Nucleotide diversity at synonymous and non-synonymous sites also
414 identify 185 effectors with signatures of diversifying selection in one or more lineage, possibly
415 mediated by coevolutionary interactions with host molecules. In the future, it will be interesting to

416 determine the molecular targets of these effectors and decipher the relationship between their
417 polymorphism and their mode of action.

418 Our results also indicate that lineage 1 may pose a major threat to rice production. Most
419 effectors with signature of diversifying selection were identified in lineage 1, highlighting the higher
420 ability of this lineage to rapidly fix advantageous mutations. Eleven of the 12 most multivirulent
421 isolates (lesion type > 2 on more than 40 varieties tested) belonged to lineage 1, and the *Yule* cluster,
422 in particular, was highly pathogenic on indica varieties. The propagation of such highly pathogenic
423 genotypes, belonging to a lineage that has both mating types and fertile females, should be
424 monitored closely. The monitoring of lineage 1 is all the more critical because this lineage has an
425 intermediate geoclimatic distribution that overlaps largely with the distributions of the other three
426 lineages, and some of the attempted crosses between clonal lineage 2-4 isolates and recombining
427 lineage 1 isolates produced viable progeny. This finding confirms the possibility of gene flow into this
428 lineage, as previously demonstrated on the basis of admixture mapping (Gladieux *et al.* 2018b), but it
429 also raises the question of the threat posed by gene flow from lineage 1, which is highly pathogenic,
430 to the other lineages.

431 In conclusion, our study of genetic and phenotypic variation within and between clonal
432 lineages, and within the recombinant lineage of *P. oryzae* suggests a scenario for the emergence of
433 widespread clonal lineages. The loss of female fertility may be a potent driver of the emergence of
434 asexually reproducing groups of clones. The reproductive isolation generated by the loss of sex and
435 the accumulation of mutations due to the absence of sexual purging would facilitate the
436 specialization of some of these clonal groups, leading to the competitive exclusion of the least
437 efficient clonal groups, and, finally, to the propagation of clonal lineages fixed for a single mating
438 type. Our results, thus, demonstrate that the spread of a pathogen across heterogeneous habitats
439 and divergent populations of a crop species can lead to niche separation and incipient speciation in
440 the pathogen.

441

MATERIALS AND METHODS

Biological material

We chose 886 *P. oryzae* isolates collected on Asian rice between 1954 and 2014 as representative of the global genetic diversity of the fungus. Isolates were selected on the basis of microsatellite data, to maximize the number of multilocus genotypes represented ((Saleh *et al.* 2014) and unpublished data), or based on geographic origin in the absence of genotypic data, to maximize the number of countries represented in the dataset.

Sixty-eight isolates were selected for experimental measurements of reproductive success, adaptation to host, and growth and sporulation experiments at different temperatures (Supplementary file 1). This subset of isolates included 10 isolates from each of the three clonal lineages (lineages 2-4), 27 isolates from the various clusters within lineage 1 [*Baoshan* (9 isolates), *International* (10), *Laos* (8), and *Yule* (11)].

Forty-four varieties were chosen as representative of the five main genetic subgroups of Asian rice (Garris *et al.* 2005): indica (Chau, Chiem chanh, DA11, De abril, IR8, JC120, Pappaku), aus (Arc 10177, Baran boro, Black gora, DA8, Dholi boro, Dular, FR13 A, JC148, Jhona 26, Jhona 149, Kalamkati, T1, Tchampa, Tepi boro), temperate japonica (Aichi asahi, Kaw luyoeng, Leung pratew, Maratelli, Nep hoa vang, Nipponbare, Sariceltik, Som Cau 70A), tropical japonica (Azucena, Binulawan, Canella de ferro, Dholi boro, Gogo lempuk, Gotak gatik, Trembese) and aromatic (Arc 10497, Basmati lamo, Dom zard, Firooz, JC1, Kaukkyisaw, N12). The Maratelli (temperate japonica) and CO39 (indica) varieties were used as susceptible controls.

Genotyping

Pyricularia oryzae isolates were genotyped at 5,657 genomic positions with an Illumina Infinium beadchip microarray carrying single-nucleotide polymorphisms identified in 49 genomes of rice- and barley-infecting *P. oryzae* isolates previously characterized (Gladieux et al., 2018b). The final dataset included 3,686 biallelic SNPs without missing data.

Whole genome sequencing

Twenty-nine isolates were sequenced using Illumina HiSeq 3000 (Supplementary file 10). Isolates were grown on rice flour-agar medium for mycelium regeneration, then in liquid rice flour medium (Adreit *et al.* 2007). Genomic DNA extraction was carried out using >100 mg of fresh mycelium from liquid culture. Fresh mycelium dried on Miracloth paper was crushed in liquid nitrogen. Nucleic acids were subsequently extracted using an extraction buffer (2 % CTAB – 1.4 M NaCl – 0.1 M Tris-HCl pH 8 – 20 mM EDTA pH 8 added before use with 1 % final of Na_2SO_3), then purified with a chloroform:isoamyl alcohol 24:1 treatment, precipited overnight in isopropanol, and rinsed with 70% ethanol. The extracted nucleic acids were further treated with Rnase A (0.2mg/mL final) and purified with another chloroform:isoamyl alcohol 24:1 treatment followed by an overnight ethanol precipitation. The concentration of extracted genomic DNA was assessed on Qubit® using the dsDNA HS Assay Kit. The purity of extracted DNA was checked by verifying that the 260/280 and 260/230 absorbance ratios measured with NanoDrop were between 1.8 and 2.0. Preparation of TruSeq nano libraries and HiSeq3000 sequencing (150 nucleotide reads, and 500 bp insert size) was carried out at GeT-PlaGe (INRAE, Toulouse, France).

Genome assembly, gene prediction, orthology analysis, effector prediction, and summary statistics of nucleotide variation

For the 123 sequenced isolates included in the dataset, low-quality reads were removed using the software CUTADAPT (Martin 2011). Reads were assembled using ABYSS 2.2.3 (Jackman *et al.* 2017;

Simpson *et al.* 2009) using different K-mer sizes and for each isolate we chose the assembled sequence with the highest N50 for further analyses. Genes were predicted with BRAKER 2.1.5 (Hoff *et al.* 2015; Hoff *et al.* 2019) using RNAseq data (Pordel *et al.* 2020) as extrinsic evidence for model refinement. Genes were also predicted with AUGUSTUS 3.4.0 (Stanke & Morgenstern 2005)(training set=*Magnaporthe grisea*) and gene models that did not overlap with gene models identified with BRAKER were added to the GFF file generated with the latter. Repeated regions were masked using REPEATMASKER 4.1.0 (http://www.repeatmasker.org/). The completeness of genome assemblies and gene predictions was assessed using BUSCO (Simão *et al.* 2015), and completeness in BUSCO genes was greater than 93.7% (Supplementary file 10). Putative effector genes were identified as genes encoding proteins predicted to be secreted by at least two methods among three [SIGNALP 4.1 (Nielsen 2017), TARGETP (Emanuelsson *et al.* 2000) and PHOBIUS (Käll *et al.* 2004)], without predicted transmembrane domain based on TMHMM analysis (Krogh *et al.* 2001), without predicted motif of retention in the endoplasmic reticulum based on PS-SCAN (Bhagwat & Aravind 2007), and without CAZy annotation based on DBSCAN V7 (Yin *et al.* 2012). Differences in numbers of putative effectors and non-effectors among lineages (ANOVA and Tukey's HSD test) were tested using using the SCIPY 1.6.0 package in PYTHON. Homology relationships among predicted genes were established using ORTHOFINDER v2.4.0 (Emms & Kelly 2015). Principal component analysis of presence/absence variation of putative effectors and non-effector proteins was carried out using R package PRCOMP. We used the get_pca_var() in PRCOMP to extract the results for variables (i.e. effectors) and identify effectors whose contribution to principal components was the greatest, defined as effectors with a contribution to loadings of PC1, PC2 and PC6 greater than 1%. To estimate the numbers of core and accessory genes, we used a rarefaction approach to account for differences in sample size across lineages. For each pseudo-sample size, we estimated the size of the core and accessory genome in each lineage using a maximum of 2000 pseudosamples. Sequences for each orthogroup were aligned and cleaned with TRANSLATORX (Abascal *et al.* 2010) using default parameters. The ratio of non-

synonymous to synonymous diversity π_N/π_S was calculated for each orthogroup for lineages with sample size greater than or equal to four using EGGLIB 3 (https://www.egglib.org).

Population subdivision and recombination

Among the 886 *P. oryzae* isolates genotyped, we identified 264 different multilocus genotypes, which were used for the analysis of population subdivision. We used the sNMF program to infer individual ancestry coefficients in *K* ancestral populations. This program is optimized for the analysis of large datasets and does not assume Hardy-Weinberg equilibrium. It is, therefore, more appropriate to deal with inbred or clonal lineages (Frichot *et al.* 2014) . We used SPLITSTREE version 4 (Huson & Bryant 2006) to visualize relationships between genotypes in a phylogenetic network, with reticulations to represent the conflicting phylogenetic signals caused by homoplasy. We also used the pairwise homoplasy index (PHI) test implemented in SPLITSTREE to test the null hypothesis of clonality. We used the sNMF software to investigate population subdivision further within the cluster for which the null hypothesis of clonality could be rejected. Weir and Cockerham's F_{ST} was calculated with the R package HIERFSTAT, by the WC84 method (Goudet 2005).

Experimental measurement of reproductive success and female fertility

Pyricularia oryzae is a heterothallic fungus with two mating types (*Mat1.1* and *Mat1.2*). Sexual reproduction between strains of opposite mating type can be observed in laboratory conditions, and results in the production of ascospores within female sexual structures called perithecia (Saleh *et al.* 2012a). On synthetic media, perithecia are formed at the contact zone between parental mycelia. Crosses were carried out on a rice flour agar medium (20 g rice flour, 2.5 g yeast extract, 15 g agar and 1 L water, supplemented with 500 000 IU penicillin G after autoclaving for 20 min at 120°C), as described by (Saleh *et al.* 2012a). We assessed reproductive success by determining the production of perithecia produced after three weeks of culture at 20°C under continuous light. Each cross was

carried out twice and we computed the average number of perithecia across repeats. Perithecia can be formed by the two interacting partners or by one partner only. Isolates forming perithecia are described as female-fertile. We measured female fertility for 220 isolates (listed in Supplementary file 1), by monitoring perithecium production in crosses involving the tester strains CH0997 (Mat1.2) and CH0999 (Mat1.1). We further assessed the presence of asci and germinating ascospores in perithecia for a subset of crosses, by excising perithecia with a scalpel and counting the number of germinated filaments for each individual ascus after overnight incubation on water agar. The subset of crosses included 10 Mat1.1 isolates (lineage 1: CH0999, CH1065, CH1076; lineage 2: CH0092, MC0016, SP0006; lineage 4: IN0017, IN0092, NP0070, CH0718) and 6 Mat1.2 isolates (lineage 1: CH0997; CH1083, CH1120; lineage 3: BR0019, CH0549, MD0929).

Pathogenicity tests

Compatibility between *P. oryzae* isolates and rice plants from the five main genetic subgroups of rice (indica, temperate japonica, tropical japonica, aus and aromatic) was measured in controlled conditions. Seventy isolates were inoculated on 46 varieties. Inoculations were performed as described by (Gallet *et al.* 2016). Conidial suspensions (25 000 conidia.mL^{-1}) in 0.5% gelatin were sprayed onto three-week-old rice seedlings (> 6 plants/variety). The inoculated plants were incubated for 16 hours in the dark at 27°C and 100% humidity and then for seven days with a day/night alternation (13 hours at 27°C/11 hours at 21°C), before scoring symptoms. Lesion type was rated from 1 to 6 (Gallet *et al.* 2016) and the symptom type was assessed visually on leaves. Each interaction was assessed in three independent experiments, and the maximum of the three scores was used in calculations.

Mycelial growth and sporulation rate at different temperatures

565 Mycelial growth and sporulation rate were measured for 41 isolates at five different temperatures
566 (10°C, 15°C, 20°C, 25°C and 30°C). For each isolate, Petri dishes containing PDA medium were
567 inoculated with mycelial plugs placed at the center, and incubated in a growth chamber with a fixed
568 temperature. Mycelium diameter was estimated as the average of two measurements made along
569 two perpendicular axes at different time points. At the end of the experiment, conidia were collected
570 by adding 5 mL of water supplemented with 0.01% Tween 20 to the Petri dish and rubbing the
571 surface of the mycelium. Conidia were counted with a hemocytometer. Three or four independent
572 experiments were performed for each isolate, at each temperature. We had initially planned to carry
573 out only three independent experiments. However, because some isolates did not grow in the first
574 experiment, and because some cultures were invaded by mites and had to be discarded in the second
575 and third experiment, a fourth experiment was eventually performed for some temperature
576 conditions.
577
578 **Statistical analyses of climatic and phenotypic data**

579 Ecological niche separation: OMI (outlying mean index), or marginality, is used to study niche
580 separation and niche breadth. OMI gives the same weight to all samplings, whether rich or poor in
581 species and individuals, and is particularly suitable in cases in which sampling is not homogeneous.
582 OMI measures the deviation between the mean environmental conditions used by one lineage and
583 the mean environmental conditions used by all lineages. OMI analysis then places the lineages in
584 environmental conditions maximizing their OMI. Environmental values, consisting of 19 biome values
585 (WorldClim bioclimatic variables (Fick & Hijmans 2017)), were retrieved for all sampling locations.
586 Climatic values were normalized before the analysis. A contingency table was generated to associate
587 the number of isolates from each lineage with each sampling location. Only isolates with a precise
588 location (known region, city or GPS position of sampling) were included. A random permutation test,
589 with 10000 permutations, was used to assess the statistical significance of marginality for each
590 lineage.

Pathotyping: Ordinal symptoms scores were analyzed with a proportional-odds model accounting for ordered categorical responses with a clm() function implemented in the ordinal R package (version 2018.8-25). The significance of the factors was analyzed by ANOVA. Pairwise comparisons of significant factors were performed after computing least-squares means with LSMEANS version 2.30-0 in R, with Tukey adjustment.

Mycelium growth: Mycelium growth was analyzed separately for each temperature, with a linear mixed-effects model. The significance of the factors was analyzed by ANOVA. Post-hoc pairwise comparisons were performed after computing least-squares means with LSMEANS version 2.30-0 in R.

Sporulation: The median number of spores calculated for the various repetitions of the experiment was analyzed with a Kruskal-Wallis test. Post-hoc pairwise comparisons were performed using Dunn's non-parametric multiple comparison test.

ACKNOWLEDGMENT

We thank Tatiana Giraud for useful suggestions. We thank all our colleagues who shared strains or participated in the collection of rice blast samples.

REFERENCES

Abascal F, Zardoya R, Telford MJ (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* **38**, W7-W13.

Abbate JL, Gladieux P, Hood ME*, et al.* (2018) Co-occurrence among three divergent plant-castrating fungi in the same Silene host species. *Molecular ecology* **27**, 3357-3370.

Adreit H, Santoso, Andriantsimialona D*, et al.* (2007) Microsatellite markers for population studies of the rice blast fungus, Magnaporthe grisea. *Molecular Ecology Notes* **7**, 667-670.

Ali S, Gladieux P, Leconte M*, et al.* (2014) Origin, migration routes and worldwide population genetic structure of the wheat yellow rust pathogen Puccinia striiformis f. sp. tritici. *PLoS pathogens* **10**, e1003903.

Anh VL, Inoue Y, Asuke S*, et al.* (2018) Rmg8 and Rmg7, wheat genes for resistance to the wheat blast fungus, recognize the same avirulence gene AVR-Rmg8. *Molecular Plant Pathology* **19**, 1252-1256.

621 Asuke S, Tanaka M, Hyon G-S, *et al.* (2020) Evolution of an Eleusine-specific subgroup of Pyricularia
622 oryzae through a gain of an avirulence gene. *Molecular Plant-Microbe Interactions* **33**, 153-
623 165.
624 Bhagwat M, Aravind L (2007) Psi-blast tutorial. In: *Comparative genomics*, pp. 177-186. Springer.
625 Bueker B, Eberlein C, Gladieux P, *et al.* (2016) Distribution and population structure of the anther
626 smut M icrobotryum silenes-acaulis parasitizing an arctic–alpine plant. *Molecular ecology* **25**,
627 811-824.
628 Burdon JJ (1993) The structure of pathogen populations in natural plant communities. *Annual review*
629 *of phytopathology* **31**, 305-323.
630 Couch BC, Fudal I, Lebrun M-H, *et al.* (2005) Origins of Host-Specific Populations of the Blast
631 Pathogen *Magnaporthe oryzae* in Crop Domestication With Subsequent Expansion of
632 Pandemic Clones on Rice and Weeds of Rice. *Genetics* **170**, 613-630.
633 Emanuelsson O, Nielsen H, Brunak S, Von Heijne G (2000) Predicting subcellular localization of
634 proteins based on their N-terminal amino acid sequence. *Journal of molecular biology* **300**,
635 1005-1016.
636 Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons
637 dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, 157.
638 Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land
639 areas. *International journal of climatology* **37**, 4302-4315.
640 Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014) Fast and efficient estimation of
641 individual ancestry coefficients. *Genetics* **196**, 973-983.
642 Gallet R, Fontaine C, Bonnot F, *et al.* (2016) Evolution of compatibility range in the rice–
643 magnaporthe oryzae system: an uneven distribution of R genes between rice subspecies.
644 *Phytopathology* **106**, 348-354.
645 Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in Oryza
646 sativa L. *Genetics* **169**, 1631-1638.
647 Gibson AK, Hood ME, Giraud T (2012) Sibling competition arena: selfing and a competition arena can
648 combine to constitute a barrier to gene flow in sympatry. *Evolution: International Journal of*
649 *Organic Evolution* **66**, 1917-1930.
650 Giraud T, Gladieux P, Gavrilets S (2010) Linking the emergence of fungal plant diseases with
651 ecological speciation. *Trends in ecology & evolution* **25**, 387-395.
652 Giraud T, Koskella B, Laine AL (2017) Introduction: microbial local adaptation: insights from natural
653 populations, genomics and experimental evolution. *Molecular ecology* **26**, 1703-1710.
654 Gladieux P, Condon B, Ravel S, *et al.* (2018a) Gene Flow between Divergent Cereal- and Grass-Specific
655 Lineages of the Rice Blast Fungus Magnaporthe oryzae. *mBio* **9**.
656 Gladieux P, Guerin F, Giraud T, *et al.* (2011) Emergence of novel fungal pathogens by ecological
657 speciation: importance of the reduced viability of immigrants. *Molecular ecology* **20**, 4521-
658 4532.
659 Gladieux P, Ravel S, Rieux A, *et al.* (2018b) Coexistence of multiple endemic and pandemic lineages of
660 the rice blast pathogen. *MBio* **9**.
661 Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular*
662 *Ecology Notes* **5**, 184-186.
663 Hardin G (1960) The competitive exclusion principle. *Science* **131**, 1292-1297.
664 Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M (2015) BRAKER1: unsupervised RNA-Seq-
665 based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767-769.
666 Hoff KJ, Lomsadze A, Borodovsky M, Stanke M (2019) Whole-genome annotation with BRAKER. In:
667 *Gene Prediction*, pp. 65-95. Springer.
668 Huson DH, Bryant D (2006) Application of Phylogenetic Networks in Evolutionary Studies. *Molecular*
669 *Biology and Evolution* **23**, 254-267.
670 Inoue Y, Vy TTP, Yoshida K, *et al.* (2017) Evolution of the wheat blast fungus through functional losses
671 in a host specificity determinant. *Science* **357**, 80-83.

672 Jackman SD, Vandervalk BP, Mohamadi H*, et al.* (2017) ABySS 2.0: resource-efficient assembly of
673 large genomes using a Bloom filter. *Genome Research* **27**, 768-777.
674 Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide
675 prediction method. *Journal of molecular biology* **338**, 1027-1036.
676 Kang S, Sweigard Ja Fau - Valent B, Valent B (1995) The PWL host specificity gene family in the blast
677 fungus Magnaporthe grisea. *Molecular Plant Microbe Interactions*.
678 Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate
679 classification updated.
680 Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein
681 topology with a hidden Markov model: application to complete genomes. *Journal of*
682 *molecular biology* **305**, 567-580.
683 Latorre SM, Reyes-Avila CS, Malmgren A*, et al.* (2020) Differential loss of effector genes in three
684 recently expanded pandemic clonal lineages of the rice blast fungus. *BMC biology* **18**, 1-15.
685 Liao J, Huang H, Meusnier I*, et al.* (2016) Pathogen effectors and plant immunity determine
686 specialization of the blast fungus to rice subspecies. *Elife* **5**, e19377.
687 Marin M, Preisig O, Wingfield BD, Kirisits T, Wingfield MJ (2009) Single sequence repeat markers
688 reflect diversity and geographic barriers in Eurasian populations of the conifer pathogen
689 Ceratocystis polonica. *Forest Pathology* **39**, 249-265.
690 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
691 *EMBnet. journal* **17**, 10-12.
692 Nielsen H (2017) Predicting secretory proteins with SignalP. In: *Protein function prediction*, pp. 59-73.
693 Springer.
694 Nosil P, Vines TH, Funk DJ (2005) Perspective: reproductive isolation caused by natural selection
695 against immigrants from divergent habitats. *Evolution; international journal of organic*
696 *evolution* **59**, 705-719.
697 Pagliaccia D, Urak RZ, Wong F*, et al.* (2018) Genetic structure of the rice blast pathogen
698 (Magnaporthe oryzae) over a decade in North Central California rice fields. *Microbial ecology*
699 **75**, 310-317.
700 Pordel A, Ravel S, Charriat F*, et al.* (2020) Tracing the origin and evolutionary history of Pyricularia
701 oryzae infecting maize and barnyard grass. *Phytopathology®*, PHYTO-09.
702 Robert S, Zapater M-F, Carlier J, Abadie C, Ravigné V (2015) Multiple introductions and admixture at
703 the origin of the continental spread of the fungal banana pathogen Mycosphaerella fijiensis
704 in Central America: a statistical test using approximate Bayesian computation. *Revue*
705 *d'écologie*.
706 Ropars J, Lo YC, Dumas E*, et al.* (2016) Fertility depression among cheese-making Penicillium
707 roqueforti strains suggests degeneration during domestication. *Evolution; international*
708 *journal of organic evolution* **70**, 2099-2109.
709 Saleh D, Milazzo J, Adreit H, Fournier E, Tharreau D (2014) South-East Asia is the center of origin,
710 diversity and dispersion of the rice blast fungus, Magnaporthe oryzae. *New Phytologist* **201**,
711 1440-1456.
712 Saleh D, Milazzo J, Adreit H, Tharreau D, Fournier E (2012a) Asexual reproduction induces a rapid and
713 permanent loss of sexual reproduction capacity in the rice fungal pathogen Magnaporthe
714 oryzae: results of in vitro experimental evolution assays. *BMC evolutionary biology* **12**, 1-16.
715 Saleh D, Xu P, Shen Y*, et al.* (2012b) Sex at the origin: an Asian population of the rice blast fungus
716 Magnaporthe oryzae reproduces sexually. *Molecular ecology* **21**, 1330-1344.
717 Schulze-Lefert P, Panstruga R (2011) A molecular evolutionary concept connecting nonhost
718 resistance, pathogen host range, and pathogen speciation. *Trends in plant science* **16**, 117-
719 125.
720 Servedio MR, Van Doorn GS, Kopp M, Frame AM, Nosil P (2011) Magic traits in speciation: 'magic' but
721 not rare? *Trends in ecology & evolution* **26**, 389-397.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212.

Simpson JT, Wong K, Jackman SD*, et al.* (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research* **19**, 1117-1123.

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* **33**, W465-W467.

Sweigard JA, Carroll AM, Kang S*, et al.* (1995) Identification, Cloning, and Characterization of *PWL2*, a Gene for Host Species Specificity in the Rice Blast Fungus. *Plant Cell* **7**, 1221-1233.

Takabayashi N, Tosa Y, Oh HS, Mayama S (2002) A Gene-for-Gene Relationship Underlying the Species-Specific Parasitism of Avena/Triticum Isolates of Magnaporthe grisea on Wheat Cultivars. *Phytopathology* **92**, 1182-1188.

Taylor JW, Fisher MC (2003) Fungal multilocus sequence typing—it's not just for bacteria. *Current opinion in microbiology* **6**, 351-356.

Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D (2006) Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 1947-1963.

Tharreau D, Fudal I, Andriantsimialona D*, et al.* (2009) World population structure and migration of the rice blast fungus, Magnaporthe oryzae. In: *Advances in genetics, genomics and control of rice blast disease*, pp. 209-215. Springer.

Walker A, Gladieux P, Decognet V*, et al.* (2015) Population structure and temporal maintenance of the multihost fungal pathogen *Botrytis cinerea*: causes and implications for disease management. *Environmental microbiology* **17**, 1261-1274.

Yang N, Ma G, Chen K, Wu X (2018) The population genetics of Alternaria tenuissima in four regions of China as determined by microsatellite markers obtained by transcriptome sequencing. *Frontiers in microbiology* **9**, 2904.

Yin Y, Mao X, Yang J*, et al.* (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **40**, W445-W451.

Yoshida K, Saunders DGO, Mitsuoka C*, et al.* (2016) Host specialization of the blast fungus Magnaporthe oryzae is associated with dynamic gain and loss of genes linked to transposable elements. *Bmc Genomics* **17**, 1.

Zhong Z, Chen M, Lin L*, et al.* (2018) Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. *The ISME journal* **12**, 1867-1878.

DATA AVAILABILITY

All raw sequencing data are deposited under accession code PRJEB42377.

The following data sets were generated:

1. European Nucleotide Archive

M Thierry, F Charriat, J Milazzo, H Adreit, S Ravel, S Cros-Arteil, S Borron, V Sella, T Kroj, R Ioos, E Fournier, D Tharreau, P Gladieux

ID PRJEB42377 Whole genome sequencing of Pyricularia oryzae fungi isolated from rice.

2. Zenodo

M Thierry, F Charriat, J Milazzo, H Adreit, S Ravel, S Cros-Arteil, S Borron, V Sella, T Kroj, R Ioos, E Fournier, D Tharreau, P Gladieux
Single-nucleotide polymorphisms, genome assemblies, genome annotations, and gene predictions of Pyricularia oryzae isolates from rice
doi 10.5281/zenodo.4561581

The following previously published data sets were used:

1. European Nucleotide Archive

A Pordel, S Ravel, F Charriat, P Gladieux, S Cros-Arteil, J Milazzo, H Adreit, M Javan-Nikkhah, A Mirzadi-Gohari, A Moumeni, D Tharreau
ID PRJEB41186 Origin and evolutionary history of Pyricularia oryzae fungal pathogens infecting maize and barnyard grass in Iran

2. European Nucleotide Archive

Z Zhong, M Chen, L Lin, Y Han, J Bao, W Tang, L Lin, Y Lin, R Somai, L Lu, W Zhang, J Chen, Y Hong, X Chen, B Wang, WC Shen, G Lu, J Norvienyeku, DJ Ebbole, Z Wang
ID PRJNA354675 Population Genomic Analysis of the Rice Blast Fungus Reveals Specific Events Associated With Expansion of Three Main Clades

ADDITIONAL FILES

Source code 1. R code used to prepare maps in Figure 6.

Source code 2. R code used for analyses based on the Outlying Mean Index, including Principal Component Analysis presented in Figure 6.

Supplementary file 1. Isolates of Pyricularia oryzae

Supplementary file 2. Assignment of genotypes to clusters identified in this study, in Gladieux et al. 2018b (A) and in Saleh et al. 2014 (B and C).

Supplementary file 3. F_{ST} estimates between clusters within lineage 1.

Supplementary file 4. Analysis of mycelial growth rates.

Supplementary file 5. Mycelial growth rates.

Supplementary file 6. Analysis of sporulation rates.

Supplementary file 7. Sporulation rates.

Supplementary file 8. Matrix of compatibility between four lineages of *P. oryzae* isolates and five subspecies of rice plants, as determined by pathogenicity tests in controlled conditions. Numbers in the matrix represent the maximum symptom score across three replicate experiments. Numbers at the right and bottom margins represent the number of compatible interactions observed (symptom score >2) for varieties and isolates, respectively.

Supplementary file 9. Analysis of pathogenicity data.

Supplementary file 10. Sequenced isolates.

Supplementary file 11. Violin plots showing the number of putative effectors and non-effector proteins detected in each lineage. Asterisks indicate significant differences in gene content (Tukey's HSD test; $p<0.001$).

Supplementary file 12. Estimating the size of the core and accessory genome using a rarefaction approach. For each lineage, genomes were resampled in <=2000 combinations of N-1 genomes (N being the sample size). (A) Non-effector genes. (B) Putative effectors.

Supplementary file 13. Principal component analysis based on the presence/absence of accessory putative effectors and non-effector proteins. Panels A and B show principal component PC1 against PC2, and PC1 against PC6, respectively, for effectors. Panels C and D show principal component PC1 against PC2, and PC1 against PC6, respectively, for non-effectors. Between parentheses is the proportion of variance represented by each principal component.

Supplementary file 14. Summary statistics of non-synonymous and synonymous polymorphism in four lineages of P. oryzae. For each lineage, only orthogroups with sample size >=4 were included in calculations. Lineage: lineage of origin of genomes used in calculations for each orthogroup. Orthogroup: orthogroup ID. N: sample size. numNS: number of non-synonymous sites. numS: number of synonymous sites. SNS: number of non-synonymous segregating sites. SS: number of synonymous segregating sites. PiNS: non-synonymous nucleotide diversity. PiS: synonymous nucleotide diversity.

Supplementary file 15. ecological niches of the four major lineages (referred to as L1 to L4) considering each of the 19 biomes individually (referred to as Bio1 to Bio19). The x-axis represents the outlying mean index (OMI), which measures the distance between the mean habitat conditions used by a lineage and the mean habitat conditions used by the entire species, to test the hypothesis that different lineages are distributed in regions with different climates.

Figure 1 – source data 1. Multilocus genotypes for neighbor-net phylogenetic inference.

Figure 1 – source data 2. Ancestry proportions in K=4 sNMF clusters.

Figure 2 – source data 1. Multilocus genotypes for neighbor-net phylogenetic inference.

Figure 2 – source data 2. Ancestry proportions in K=4 sNMF clusters.

Figure 2 – source data 3. Geographic distribution of clones (i.e. multilocus genotypes repeated multiple times).

Figure 3 – Source data 1. Production of perithecia.

Figure 3 – Source data 2. Production of germinating ascospores

Figure 4 – source data 1. Symptom scores for *P. oryzae* isolates inoculated onto rice varieties

Figure 5 – source data 1. Size of core and accessory genomes for putative effectors and non-effector proteins

Figure 5 – source data 2. Presence and absence of putative effectors in *P. oryzae* genomes

Figure 6 – source data 1. GPS position of isolates

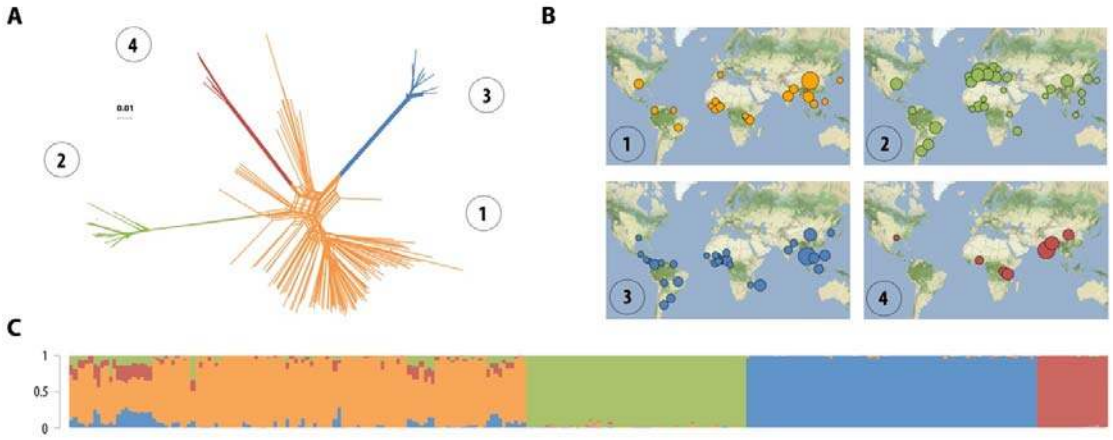

Figure 1. Rice-infecting *P. oryzae* populations are divided into four major lineages. Population subdivision was inferred from 264 distinct *P. oryzae* genotypes, representing 886 isolates, and the four lineages were represented in different colors. A: Neighbor-net phylogenetic network estimated with SPLITSTREE; reticulations indicate phylogenetic conflicts caused by homoplasy. B: Geographic distribution of the four lineages identified with SPLITSTREE and sNMF, with disk area proportional to sample size. C: Ancestry proportions in K=4 clusters, as estimated with sNMF software; each multilocus genotype is represented by a vertical bar divided into four segments indicating membership in K=4 clusters.

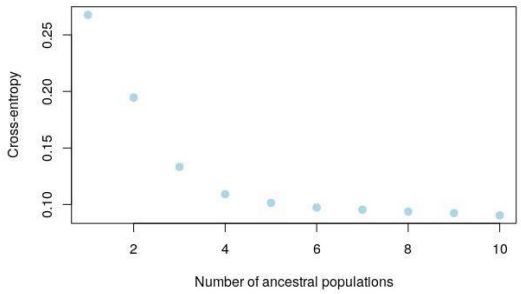

Figure 1 – figure supplement 1. Cross-entropy (CE) as a function of the number of clusters K modeled in sNMF analyses of population subdivision.

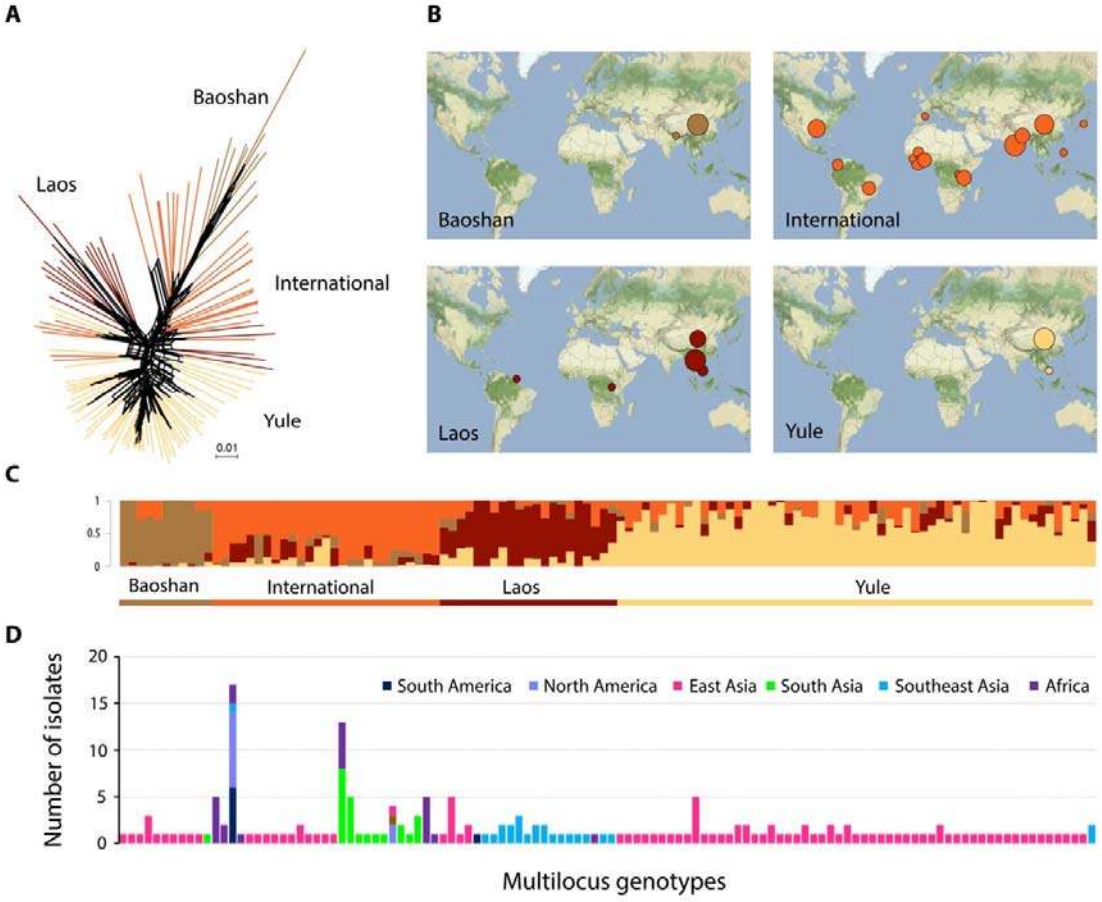

Figure 2. Population subdivision in lineage 1. A: Neighbor-net phylogenetic network estimated with SPLITSTREE; reticulations indicate phylogenetic conflicts caused by homoplasy. B: Geographic distribution of the four clusters identified with sNMF, with disk area proportional to number of isolates. C: Ancestry proportions in four clusters, as estimated with sNMF; each multilocus genotype is represented by a vertical bar divided into four segments, indicating membership in *K*=4 clusters. D: Number of isolates and their geographic origin for each multilocus genotype of lineage 1. Panels C and D share an *x*-axis (each vertical bar represents a different multilocus genotype).

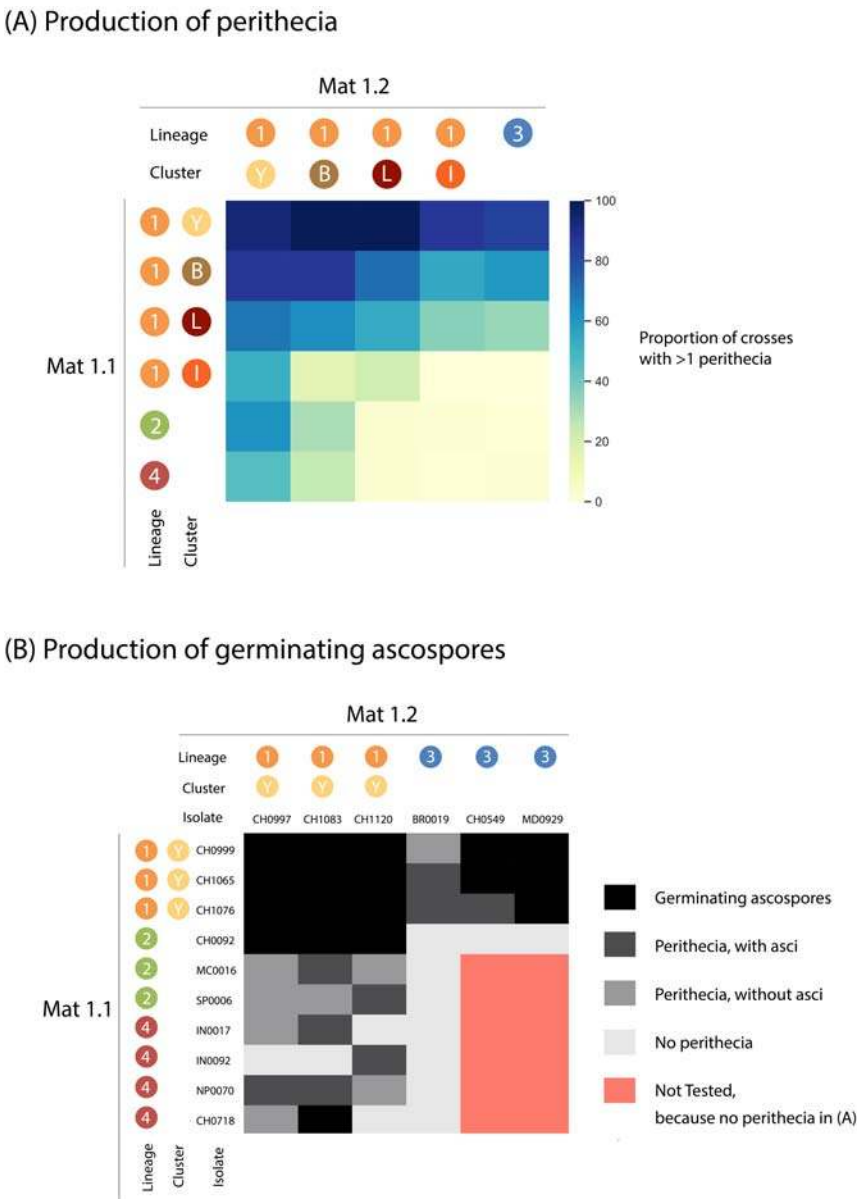

Figure 3. Success of crosses between lineages 2-4 and clusters within lineage 1 with (A) proportion of crosses producing at least one perithecium, (B) scoring of ascus formation and ascospore germination for a subset of crosses.

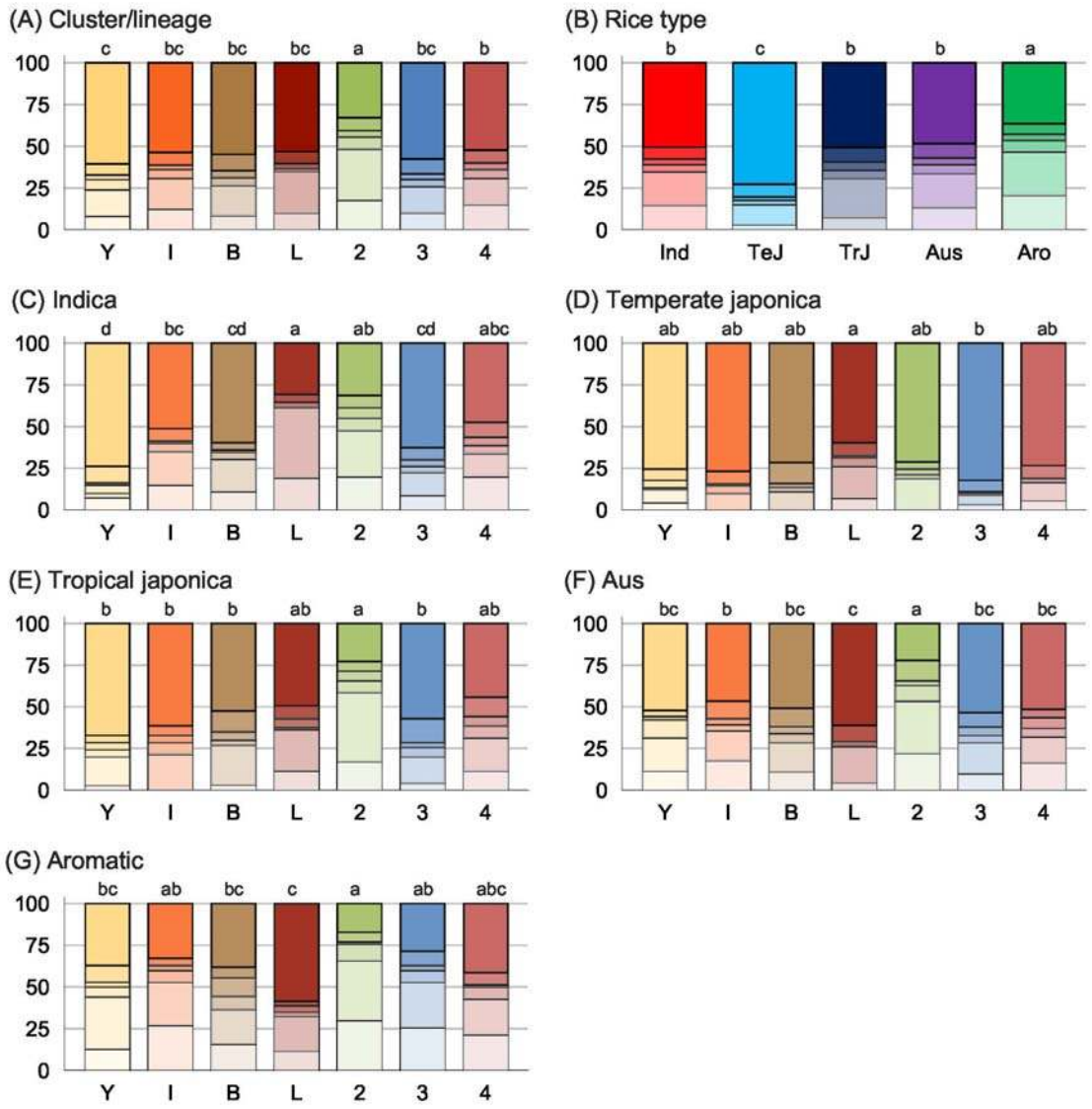

Figure 4: Compatibility between 70 *P. oryzae* isolates and 44 rice varieties, representing five types of rice. A: Proportions of symptom scores as a function of the lineage of origin of isolates, or cluster of origin for isolates from lineage 1; B: Proportions of symptom scores as a function of the type of rice; C-G: Proportions of symptom scores as a function of the lineage of origin of isolates, for each type of rice. Abbreviations: Y, *Yule*; I, *International*; B, *Baoshan*; L, *Laos*; 2, lineage 2; 3, lineage 3; 4, lineage 4; Ind, indica; TeJ, temperate japonica; TrJ, tropical japonica; Aro, aromatic. Small capitals indicate significant differences. All interactions were assessed in three independent experiments, and the highest of the three symptom scores was used in calculations.

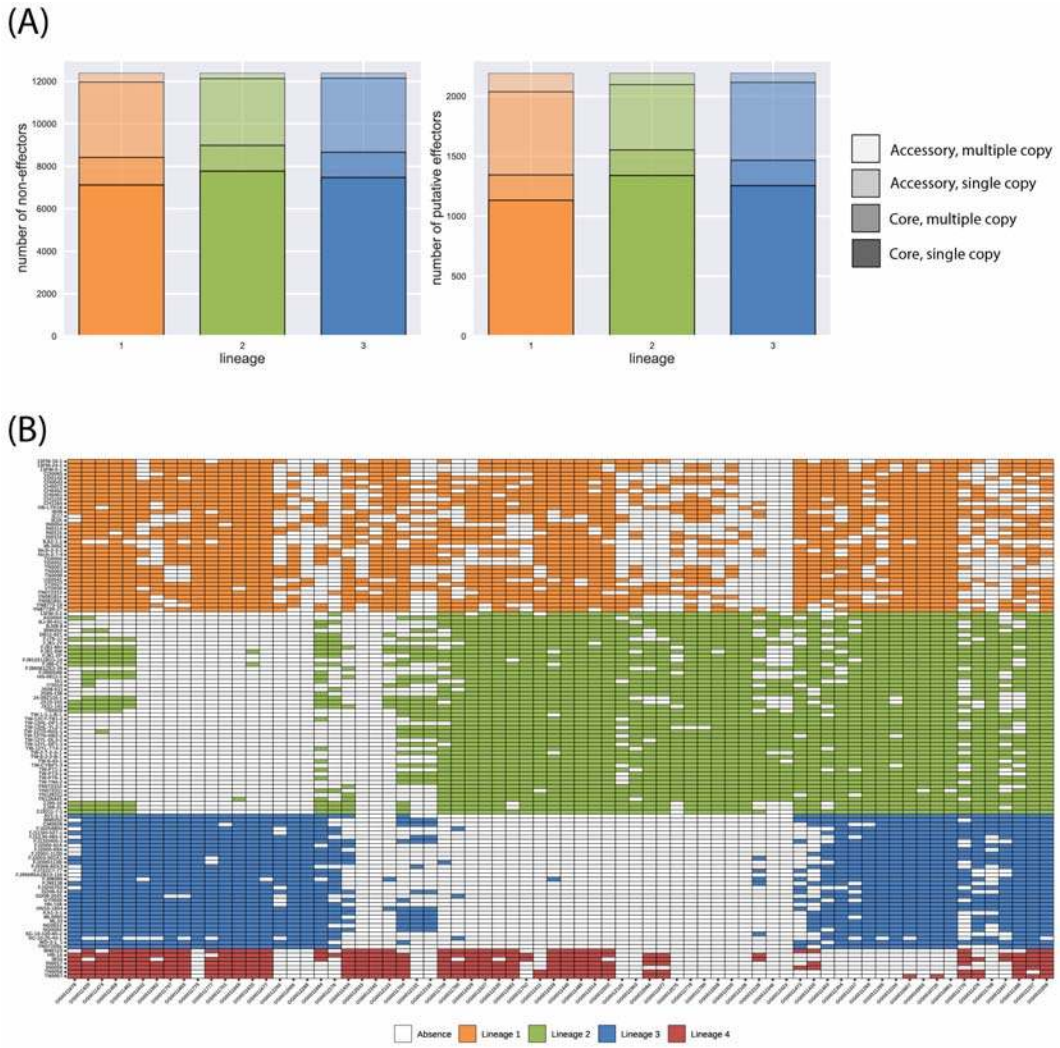

Figure 5. Pangenome analyses of 123 genomes of *P. oryzae*, representing four rice-infecting lineages. (A) Size of core and accessory genomes for putative effectors and non-effector proteins estimated using a rarefaction approach with a pseudo-sample size of n=30 genomes. Lineage 4 was not included due to small sample size. (B) Presence and absence in lineage 1-4 of the 72 putative effectors with the highest contribution to principal components 1, 2 and 6, in a principal component analysis of presence/absence data (Supplementary file 13).

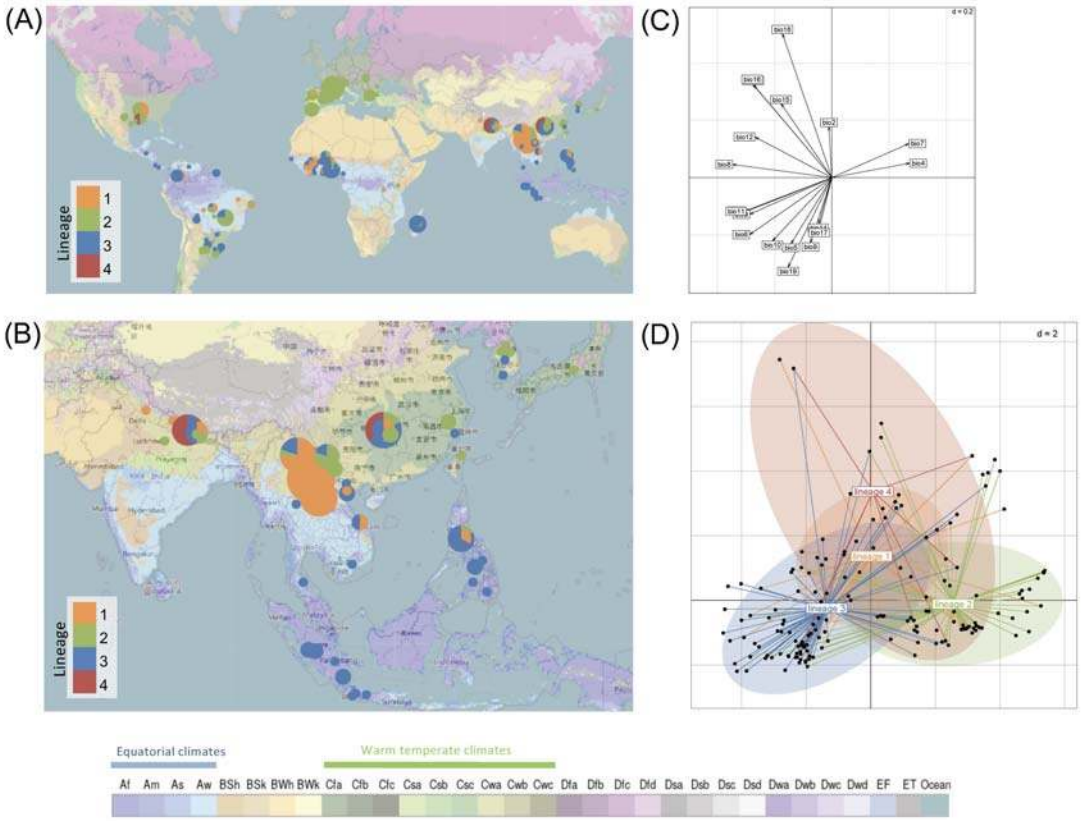

Figure 6. Geographic distribution of four lineages of *P. oryzae* and corresponding climatic data. (A, B) Pie charts representing the distribution of the four lineages in the world (A) and in South, East and Southeast Asia (B), keeping only isolates for which the sampling position was precisely known (i.e., for which the region, city or GPS position was documented). Background map is from OpenStreetMap (CC-BY-SA) and represents the updated Köppen-Geiger climate classification of main climates at world scale as described in (Kottek *et al.* 2006). (C, D) Outlying Mean Index analysis, testing the separation of ecological niches among lineages, with (C) canonical weights of the 19 environmental variables included in the analysis, and (D) site coordinates (dots) and realized niches of lineages represented as ellipses. Variable bio13 co-localizes with bio16, and variables bio1 and bio3 co-localize with bio11. The 11 temperate variables included in the analysis were (bio1) Annual Mean Temperature, (bio2) Mean Diurnal Range, (bio3) Isothermality [100*(bio2/bio7)], (bio4) Temperature Seasonality [standard deviation*100], (bio5) Max Temperature of Warmest Month, (bio6) Min Temperature of Coldest Month, (bio7) Temperature Annual Range [bio5-bio6], (bio8) Mean Temperature of Wettest Quarter, (bio9) Mean Temperature of Driest Quarter, (bio10) Mean Temperature of Warmest Quarter, (bio11) Mean Temperature of Coldest Quarter. The 8 precipitation variables included in the analysis were (bio12) Annual Precipitation, (bio13) Precipitation of Wettest Month, (bio14) Precipitation of Driest Month, (bio15) Precipitation Seasonality (Coefficient of

Variation), (bio16) Precipitation of Wettest Quarter, (bio17) Precipitation of Driest Quarter, (bio18) Precipitation of Warmest Quarter, (bio19) Precipitation of Coldest Quarter.