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

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 lineages on the same host, but the factors underlying 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 with different prevailing environmental conditions and types of rice grown, indicating niche separation. 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. Experimental crosses 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 reproductive isolation between distinct invasive lineages.

INTRODUCTION

Invasive fungal plant pathogens constitute a major threat to ecosystem health and agricultural production (Desprez-Loustau *et al.* 2007; Rizzo 2005; Strange & Scott 2005). Three decades of molecular plant pathology have revealed a seemingly inexhaustible stream of disease outbreaks or significant changes in the geographic location of diseases and host ranges of pathogens. In addition to the burden of emerging fungal infections, most major crops or domesticated trees are colonized by pathogens that are so widespread and have been present for so long, that they do not necessarily come to mind when one thinks about invasive fungi (Palm 2001; Yarwood 1970). If fungal pathogens are to spread successfully over their host species distribution, they must overcome a series of barriers to invasion related to dispersal, host availability, competition with other microbes, and abiotic conditions (Gladieux *et al.* 2015). An understanding of the factors underlying the success of invasive fungal pathogens is essential to protect world agriculture against future pandemics (Thrall *et al.* 2011).

Invasive fungal plant pathogens display a range of population structures, reflecting their highly diverse life-history strategies and invasion histories (Barrett *et al.* 2008; Gladieux *et al.* 2015; Hessenauer *et al.* 2020). Demographic events (e.g. population bottlenecks, secondary contact) and natural selection during spread across heterogeneous environments may be associated with intricate changes in demography and life history traits, or may cause such changes. Shifting to a new host is a primary life history trait change in invasive fungal pathogens, and a major route for the emergence of disease. Changes in host range are facilitated by certain features, such as mating within or on their hosts and strong selective pressure on a limited number of genes (Giraud *et al.* 2010; Gladieux *et al.* 2011a). Reproduction within, or on plant hosts favors assortative mating with respect to host use, leading to a strong association between adaptation to the host and reproductive isolation (Giraud *et al.* 2006; Servedio *et al.* 2011) and the differentiation of pathogen populations adapted to different host populations, varieties or species (Lemaire *et al.* 2016; Mercier *et al.* 2021; Peever 2007;

Stukenbrock *et al.* 2011). Such changes in pathogen host range can result from immune-escape mutations at a small number of genes encoding small secreted proteins called effectors, which enable microbes to influence the outcome of host-pathogen interactions for their own benefit, and which may be subject to surveillance by the plant immune system (Couch *et al.* 2005; Schulze-Lefert & Panstruga 2011a; Uhse & Djamei 2018). Another frequent change in the life history traits of invasive fungal pathogens is the loss of sexual reproduction (Taylor *et al.* 2015; Taylor *et al.* 1999). Invasive fungal pathogens may develop a clonal population structure over their entire introduced range or over parts of that range ((Gladieux *et al.* 2015) and references therein). The immediate causes of a the loss of sexual reproduction include a lack of compatible mating partners in the introduced range (i.e. a lack of compatible mating types) (Hessenauer *et al.* 2020), hybridization and hybrid sterility (Li *et al.* 2019; Lin & Heitman 2006), and a lack of compatible alternate hosts for sexual reproduction (Fellers *et al.* 2021). Mating type loss may be driven by the extreme population bottlenecks associated with the colonization of a new host or new area. Shifts to asexual reproduction may also result from selection against reproduction (Gladieux *et al.* 2015), particularly for pathogens of domesticated crops and trees, due to the availability of large, homogeneous host populations, releasing constraints that maintain sexual reproduction in the short term (de Vienne *et al.* 2013). These conditions may favor the asexual spread of new lineages carrying adaptive allelic combinations requiring shelter from recombination (Salcedo *et al.* 2017). Finally, changes in life history traits can also result from variations in environmental conditions, including temperature in particular, during the course of invasive spread (Frenkel *et al.* 2010; Mboup *et al.* 2012). Deciphering the complex interplay between changes in population structure and life history strategies that accompanies the spread of invasive fungi requires deep sampling the and integration of genetic, phenotypic, and environmental information throughout the native and invasive ranges.

Pyricularia oryzae (Ascomycota; syn. *Magnaporthe oryzae*) is a widespread model plant pathogen displaying population subdivision. *Pyricularia oryzae* is best known as the pathogen causing rice blast, which remains a major disease of Asian rice (*Oryza sativa*), but it is also a major constraint

on the production of wheat (*Triticum*), finger millet (*Eleusine coracana*) and Italian millet (*Setaria italica*), a significant disease of ryegrass and an emerging pathogen of maize (*Zea maydis*) (Farman *et al.* 2017; Inoue *et al.* 2017; Islam *et al.* 2016; Milazzo *et al.* 2019; Pordel *et al.* 2021; Tanaka *et al.* 2009). An in-depth genomic analysis of a large worldwide collection of *P. oryzae* isolates from 12 host plants revealed the existence of 10 distinct *P. oryzae* lineages, each associated mostly with a single main cereal crop or grass host (Gladieux *et al.* 2018a). Sequence divergence between lineages was only slight, of the order of about 1% (Gladieux *et al.* 2018a), and analyses of gene flow and admixture provided evidence that the different host-adapted lineages were connected by relatively recent genetic exchanges, and therefore corresponded to a single phylogenomic species (Langner *et al.* 2018; Valent *et al.* 2019). The reproductive biology of *P. oryzae* is typical of plant pathogenic Ascomycetes. *Pyricularia oryzae* has a haplontic life cycle. Thus, the multicellular state is haploid, and the breeding system is heterothallic, with mating occurring only between haploid individuals of opposite mating types. *Pyricularia oryzae* is hermaphrodite, but successful mating requires that at least one of the partners to be capable of producing female structures (i.e. "female-fertile").

The lineage of *P. oryzae* that infects rice is the most widely studied at the population level, due to its agricultural importance. Analyses of populations from commercial fields and trap nurseries based on molecular markers and mating assays have consistently shown that non-Asian populations of the rice-infecting lineage harbor a single mating-type and are clonal, with sexual reproduction never observed in the field (Tharreau *et al.* 2009; Zeigler 1998). Populations with female-fertile strains, a coexistence of both mating types and signatures of recombination have been observed exclusively in Asia (Kumar *et al.* 1999; Saleh *et al.* 2014; Saleh *et al.* 2012b; Tharreau *et al.* 2009; Zeigler 1998). More recent population genomic studies of the rice-infecting lineage have revealed genetic subdivision into four main lineages, estimated to have diverged approximately 1000 years ago and displaying contrasting modes of reproduction (Gladieux *et al.* 2018b; Latorre *et al.* 2020; Zhong *et al.* 2018). Only one of the four lineages that infects rice (lineage 1, prevailing in East and Southeast Asia) displays a genome-wide signal of recombination, a balanced ratio of mating type alleles and a

104 high frequency of fertile females, consistent with sexual reproduction (Gladieux *et al.* 2018b; Saleh *et*
105 *al.* 2014). All the other lineages (lineages 2, 3 and 4, present on several continents) have clonal
106 population structures, highly biased frequencies of mating types and only a few fertile females,
107 suggesting that reproduction in these lineages is strictly asexual (Gladieux *et al.* 2018b; Saleh *et al.*
108 2014). The recombining lineage has a nucleotide diversity four times greater than that of the clonal
109 lineages (Gladieux *et al.* 2018c). In theory, gene flow remains possible between the majority of the
110 lineages, because there are rare fertile strains and some lineages have fixed opposite mating types.
111 However, while resequencing studies have revealed signatures of recent shared ancestry between
112 clonal lineages and the recombinant lineage from South-West Asia, this is not the case between the
113 different clonal lineages despite their wide and overlapping distributions, suggesting a lack of recent
114 gene flow between lineages of opposite mating types (Gladieux *et al.* 2018b). The loss of female
115 fertility is probably a key factor in the emergence of clonal lineages, inducing a shift towards strictly
116 asexual reproduction, and protecting populations adapting to new conditions from maladaptive gene
117 flow from ancestral, recombining populations (Giraud *et al.* 2010). This raises questions about the
118 nature of the factors underlying the emergence of different rice-infecting lineages in *P. oryzae*, and
119 contributing to their maintenance, because, according to the competitive exclusion principle,
120 different lineages should not be able to co-exist on the same host (Abbate *et al.* 2018; Giraud *et al.*
121 2017; Hardin 1960). Previous studies have helped to cement the status of *P. oryzae* as a model for
122 studies of the population biology of fungal pathogens, but most efforts to understand the population
123 structure of the pathogen have been unable to provide a large-scale overview of the distribution of
124 rice-infecting lineages and of the underlying phenotypic differences, because the number of genetic
125 markers was limited (Saleh *et al.* 2014; Tharreau *et al.* 2009), the number of isolates was relatively
126 small (Gladieux *et al.* 2018b; Zhong *et al.* 2018) and because few phenotypic traits were scored (Saleh
127 *et al.* 2014; Tharreau *et al.* 2009; Zhong *et al.* 2018).

128 We provide here a detailed genomic and phenotypic overview of natural variation in rice-
129 infecting *P. oryzae* isolates sampled across all rice-growing areas of the world. With a view to

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

142 **RESULTS**

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

145 We characterized the global genetic structure of the rice blast pathogen, using an Illumina
146 genotyping beadchip to score 5,657 SNPs distributed throughout the genome of 886 *P. oryzae*
147 isolates collected from cultivated Asian rice in 51 countries (Supplementary file 1). The final dataset
148 included 3,686 SNPs after the removal of positions with missing data, which identify 264 distinct
149 multilocus genotypes out of 886 isolates. The genotyping error rate was below 0.62%
150 (Supplementary file 1). Clustering analyses based on sparse nonnegative matrix factorization
151 algorithms, as implemented in the sNMF program (Frichot et al., 2014), revealed four clusters,
152 hereafter referred to as "lineages" (Figure 1C). The model with K=4 clusters was identified as the best
153 model on the basis of the cross-entropy criterion, as models with K>4 induced only a small decrease
154 in cross-entropy, suggesting that K=4 captured the deepest subdivisions in the dataset (Figure 1 –
155 figure supplement 1). The neighbor phylogenetic network inferred with SPLITSTREE also supported the
156 subdivision into four widely distributed lineages, with long branches separating three peripheral
157 lineages branching off within a central lineage (Figure 1A; 1B). Principal Component Analysis (PCA)
158 revealed four groups coinciding with the four lineages identified with phylogenetic network and
159 clustering analyses. A comparison with previous findings revealed that the central lineage in the
160 Neighbor-net network and PCA corresponded to the combination of recombining lineage 1 and
161 lineages 5 and 6, represented by two and one individuals, respectively, in a previous study (Gladieux
162 *et al.* 2018b) (Supplementary file 2). This central lineage is referred to hereafter as lineage 1. The
163 three peripheral lineages in the Neighbor-net network and PCA corresponded to the previously
164 described lineages 2, 3 and 4 (Gladieux *et al.* 2018b) with lineages 2 and 3 similar to lineages B and C,
165 respectively, identified on the basis of microsatellite data in a previous study (Saleh *et al.*
166 2014)(Supplementary file 2). Genetic differentiation between the four lineages was strong and
167 significant (Weir and Cockerham's F_{ST}> 0.54), indicating strong barriers to gene flow between the

168 lineages. All genotypes from lineages 2 to 4 had high proportions of membership *q* in a single cluster
169 (q > 0.89), whereas shared ancestry with the other three lineages was detected in lineage 1, with
170 32% of genotypes having q > 0.10 in lineages 2-4 (Figure 1C). Admixture may account for the lower
171 F_{ST} observed in comparisons between lineage 1 and the other lineages.

172 We detected no footprints of recombination in lineages 2, 3 and 4 with three different tests
173 (pairwise homoplasy index (Bruen *et al.* 2006), maximum χ^2 (Maynard Smith 1992) and neighbour
174 similarity score (Jakobsen & Easteal 1996)) and a null hypothesis of clonality (Table 1), confirming
175 previous findings obtained with smaller datasets (Gladieux *et al.* 2018b; Latorre *et al.* 2020; Saleh *et*
176 *al.* 2014; Zhong *et al.* 2018). The alignment of genotypes from lineages 2, 3 and 4 in the PCA (Figure
177 1C), their lack of linkage disequilibrium decay with increasing physical distance (Supplementary file
178 3), and their low proportions of homoplastic mutations (Table 1), provided additional evidence for an
179 absence of recombination in these lineages, contrasting with lineage 1.

180 Using whole-genome data for isolates from lineages of *P. oryzae* infecting other grasses and
181 cereals, we showed that our set of Infinium SNPs could be used to identify isolates not belonging to
182 the rice-infecting lineage, although it was unable to differentiate between the different host-specific
183 lineages of *P. oryzae* (Appendix 1). Clustering analysis identified to two isolates from West Africa
184 (BN0019 and BF0072, clonal group 18 [Supplementary file 1]) corresponding to a lineage different
185 from the rice lineage and possibly resulting from ancient admixture between other lineages
186 (Appendix 1). However, on inclusion of these divergent isolates in our dataset, we detected no fixed
187 differences with respect to the rest of the isolates from lineage 1, and these divergent isolates did
188 not form a separate group in clustering analyses. They are not, therefore, likely to have affected on
189 our main conclusions on population subdivision and mode of reproduction.

190

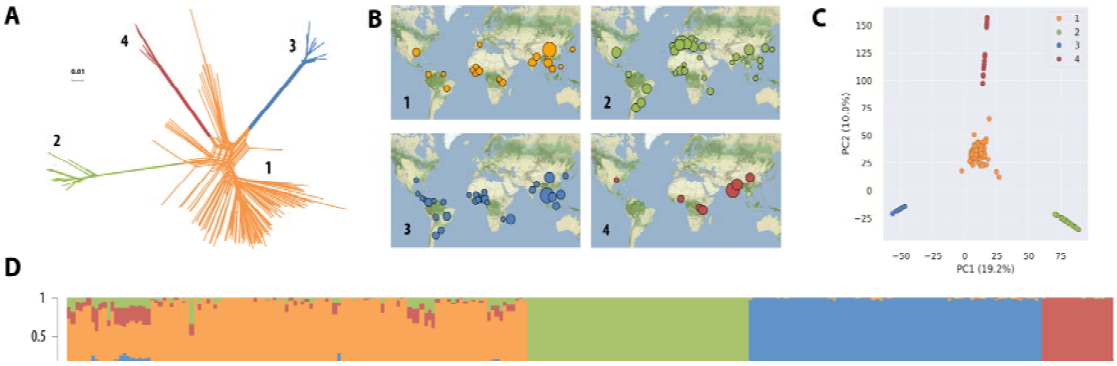

Figure 1. Rice-infecting *P. oryzae* populations are divided into four major lineages. Population subdivision was inferred from 264 distinct *P. oryzae* genotypes, representing 886 isolates, and the four lineages were represented in different colors. **(A)** Neighbor-net phylogenetic network estimated with SPLITSTREE; reticulations indicate phylogenetic conflicts caused by homoplasy. **(B)** Geographic distribution of the four lineages identified with SPLITSTREE (A) and sNMF (D), with disk area proportional to sample size. **(C)** Principal Component Analysis with colors indicating lineages identified with SPLITSTREE (A) and sNMF (D). **(D)** 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.

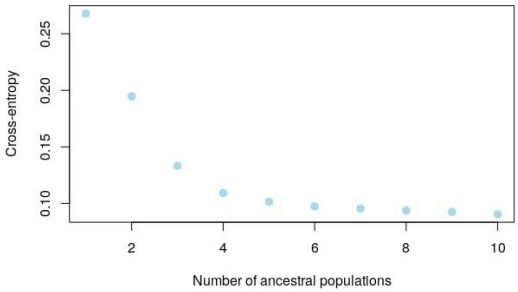

Figure 1 – figure supplement 1. Cross-entropy (CE) as a function of the number of clusters K modeled in sNMF analyses of population subdivision.

Figure 1 – source data 1.

Multilocus genotypes for neighbor-net phylogenetic inference, principal component analysis and recombination analyses.

Figure 1 – source data 2.

Ancestry proportions in K=4 sNMF clusters.

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 on all continents in which the pathogen was sampled (Europe: one isolate; North America:

211 10, Central and South America: 7, Africa: 22) (Figure 2). Clustering analysis on this single lineage with
212 sNMF detected four clusters with different geographic distributions (Figure 2). Estimates of
213 differentiation were lower between the clusters within lineage 1 (F_{ST} < 0.49) than between the main
214 lineages (F_{ST} > 0.54), consistent with a longer history of restricted gene flow between the main
215 lineages and a more recent history of admixture between clusters within lineage 1 (Supplementary
216 file 3; Figure 1C).

217 Two clusters within lineage 1 (referred to hereafter as *Baoshan* and *Yule*) consisted mostly of
218 isolates sampled from two different sites in the Yunnan province of China (Baoshan and Yule,
219 respectively, located about 370 km apart). The third cluster consisted mostly of isolates from Laos
220 and South China (referred to hereafter as *Laos*), and the fourth brought together 95% of the isolates
221 from lineage 1 collected outside Asia (referred to hereafter as *International*). In Asia, the *International*
222 cluster was found mostly in the Yunnan province of China, India and Nepal (Figure 2). PHI tests for
223 recombination rejected the null hypothesis of clonality for all four clusters (Table 1). Admixture was
224 widespread in lineage 1 (Figure 2), with most of the isolates (78%) displaying membership
225 proportions with q > 0.10 for two or more clusters (Figure 2). Only five genotypes were detected in
226 multiple countries (genotype ID [number of countries]: 2 [6], 18 [2], 58 [2], 98 [3] and 254 [3],
227 corresponding to 41 isolates in total). All these genotypes belonged to the *International* cluster.
228

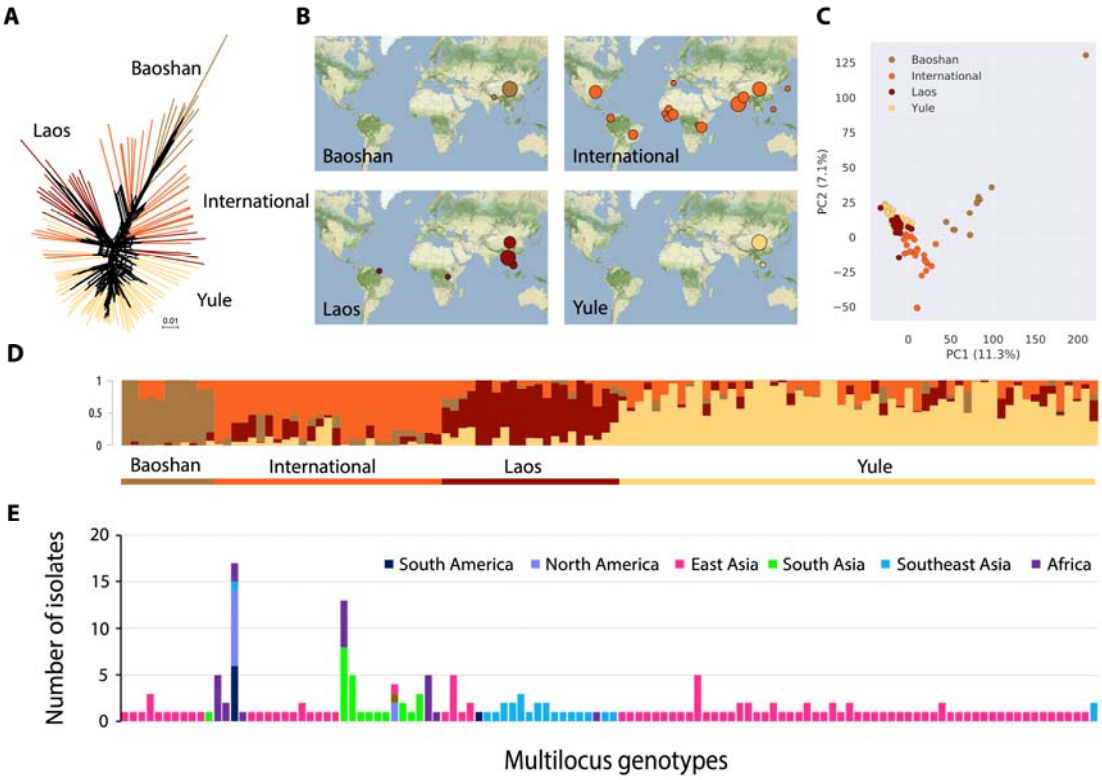

Figure 2. Population subdivision in lineage 1. **(A)** Neighbor-net phylogenetic network estimated with SPLITSTREE; reticulations indicate phylogenetic conflicts caused by homoplasy. **(B)** Principal Component Analysis with colors indicating clusters identified with SPLITSTREE (A) and sNMF (D). **(C)** Geographic distribution of the four clusters identified with sNMF (D), with disk area proportional to number of isolates. **(D)** 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. **(E)** Number of isolates and their geographic origin for each multilocus genotype of lineage 1. Panels D and E share *x*-axis (each vertical bar represents a different multilocus genotype).

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).

Table 1. Tests for recombination (null hypothesis: clonality), distribution of mating types, fertile females, and proportion of homoplastic polymorphisms. Analyses were conducted for four lineages, and for clusters within lineage 1.

Lineage/*Cluster*	PHI (*p*-value)	Max χ^2 (*p*-value)	NSS (*p*-value)	Homoplastic sites (%)	Mat1.1/Mat1.2	Female-fertile isolates (%)
1	<0.0001	<0.0001	<0.0001	1388/2214 (63)	53/47	43
Baoshan	<0.0001	<0.0001	<0.0001	NC	69/31	67
International	<0.0001	<0.0001	<0.0001	NC	55/45	4
Laos	<0.0001	<0.0001	<0.0001	NC	68/32	37
Yule	<0.0001	<0.0001	<0.0001	NC	40/60	79
2	0.25	0.32	0.72	1/391 (0)	97/3	0
3	0.18	0.35	0.88	0/408 (0)	3/97	7
4	0.22	0.30	0.51	0 (218)	97/3	0

NC, not computable because clusters within lineage 1 are polyphyletic in the phylogeny used to estimate the proportion of homoplastic sites. Tests for recombination: Pairwise homoplasy index (PHI; (Bruen *et al.* 2006)), Maximum χ^2 (Max χ^2, (Maynard Smith 1992)) and Neighbour Similarity Score (NSS; (Jakobsen & Easteal 1996)). Source datafiles: Figure 1 – source data 1 for PHI, Max χ^2, NSS; Table 1 – source data 1, Table 1 – source data 2 and Table 1 – source data 3 for the proportion of homoplastic sites. Information about mating types and female fertility is available in Supplementary file 1.

Table 1 – source data 1.

RAxML genealogy for homoplasy analysis, estimated from Figure 1 – source data 1.

Table 1 – source data 2

Pairs of nodes included in homoplasy analysis, as determined from Table 1 – source data 1.

Table 1 – source data 3

Result of 'Trace all characters' analysis conducted with MESQUITE.

Reproductive barriers caused by female sterility and postmating genetic incompatibilities

Our analysis of genotyping data revealed the existence of a recombinant population from which several clonal lineages have more recently emerged. This raised the question as to why these populations became clonal, with no further gene flow. We addressed this question by determining the capacity of the different lineages to engage in sexual reproduction, by characterizing the

266 distribution of mating types and the ability to produce female structures (i.e. female fertility). Using
267 *in vitro* crosses with tester strains, we found that lineages 2, 3 and 4 were composed almost
268 exclusively of a single mating type (97% of lineage 2 and 4 isolates tested carried the *Mat1.1* allele;
269 97% of lineage 3 isolates carried the *Mat1.2* allele), and had only small proportions (0-7%) of isolates
270 with female fertility (Table 1). By contrast to these clonal lineages, the mating type ratio was more
271 balanced in lineage 1 (52% of *Mat1.1*; Table 1), with *Mat1.1*/*Mat1.2* ratios ranging from 40/60 in the
272 *Yule* cluster to 69/31 in the *Baoshan* cluster. Female fertility rates were higher in three of the clusters
273 in lineage 1 (*Yule*: 79%; *Baoshan*: 67%; *Laos*: 37%; Table 1), but that in the International cluster (4%
274 female fertility) was as low as those in the clonal clusters (Table 1). These observations indicate that
275 there is a potential for sexual reproduction between lineages with compatible mating types and,
276 albeit rare, fertile females, but not between lineages 2 and 4, due to the low female fertility rates
277 and highly biased mating type ratios.

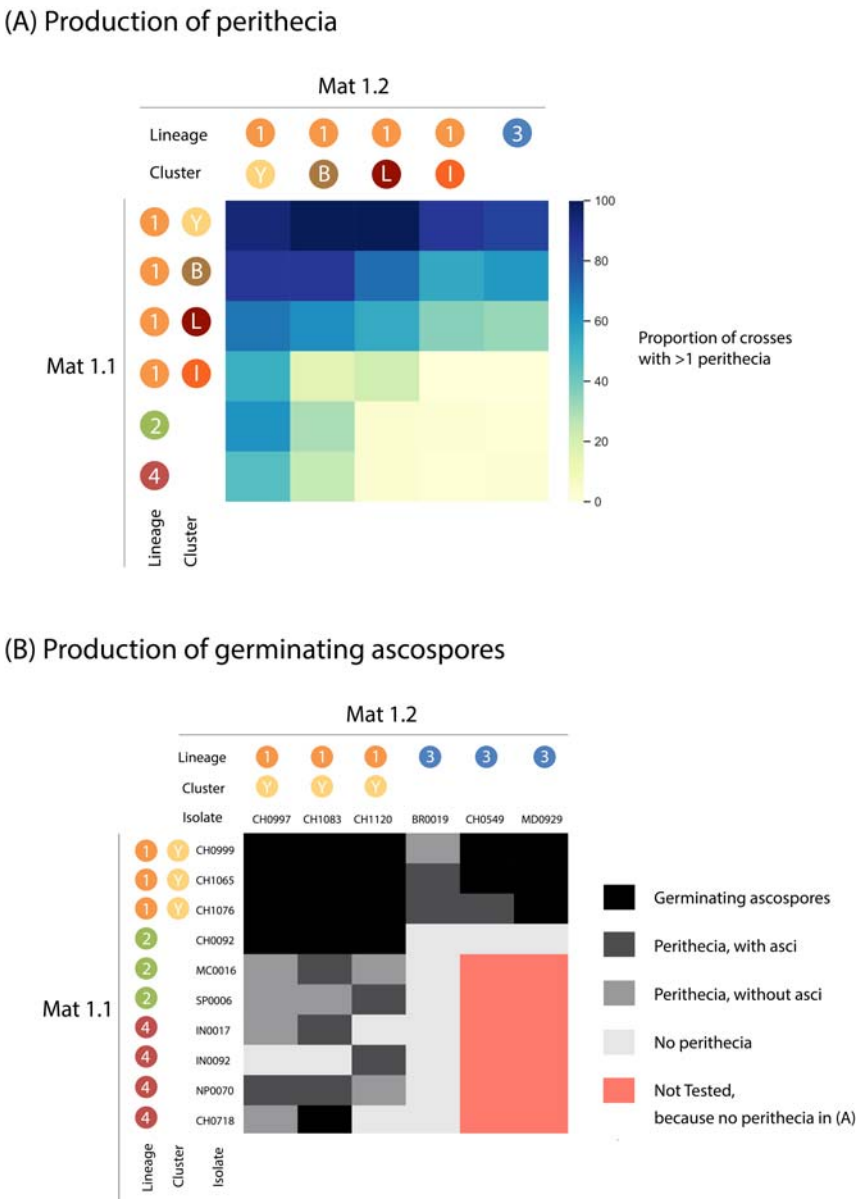

Figure 3. Success of crosses between lineages 2-4 and clusters within lineage 1 with (A) proportion of crosses producing at least one perithecium, (B) scoring of ascus formation and ascospore germination for a subset of crosses.

Figure 3 – Source data 1

Production of perithecia

Figure 3 – Source data 2

Production of germinating ascospores

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

289 crosses between randomly selected isolates of opposite mating types from all four lineages (Figure 3).
290 This experiment revealed a marked heterogeneity in the rate of perithecium formation across
291 lineages. Three clusters in lineage 1 had the highest rates of perithecia formation, with isolates in the
292 *Yule* cluster, in particular, forming perithecia in 93% of crosses with isolates from the same cluster,
293 and in more than 46% of crosses with isolates from other lineages (Figure 3A). Isolates from lineages
294 2, 3 and 4 could only be crossed with isolates from other lineages, given their highly biased mating
295 type ratios, and the percentage of these crosses producing perithecia was highly variable (ranging
296 from 0% to 83%, depending on the lineages involved; Figure 3A). The rate of perithecium formation
297 was similar in crosses involving the *International* cluster of lineage 1 and in crosses involving clonal
298 lineages 2-4, and none of the crosses attempted between isolates from the *International* cluster led
299 to perithecium formation (Figure 3A). Perithecium dissection for a subset of crosses involving some of
300 the most fertile isolates revealed that most inter-lineage crosses produced perithecia that did not
301 contain asci or contained asci with low ascospore germination rates (Figure 3B). Whereas 100% of
302 crosses between isolates of the *Yule* cluster produced numerous germinating ascospores, ascospore
303 germination rates were only 33%, 56% and 7% in *Yule* x lineage 2, *Yule* x lineage 3 and *Yule* x lineage 4
304 crosses, respectively. Together, these results indicate that the clonal structure of lineages 2-4 is
305 caused by highly biased mating types and female sterility, and that the three clonal lineages and the
306 *International* cluster of lineage 1 are isolated from each other by strong pre- and early-postmating
307 barriers, including breeding system isolation (differences in mating type and female sterility), and a
308 combination of gametic incompatibility, zygotic mortality or hybrid non-viability. However, three of
309 the clusters in lineage 1 had biological features consistent with an ability to reproduce sexually,
310 suggesting possible sexual compatibility with clonal lineages 2-4 and the *International* cluster.
311

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

313 Given the very broad distribution of *P. oryzae* and the strong environmental heterogeneity with
314 which it is confronted in terms of the nature of its hosts and the climate in which it thrives, we

evaluated the variation of pathogen fitness over different types of rice and temperature conditions. We measured growth rate and sporulation rate of representative isolates cultured at different temperatures to test the hypothesis of adaptation to temperature. Experiments were conducted with 41 isolates (lineage 1 [yule cluster]: 11; lineage 2: 10; lineage3: 10; lineage 4: 10). For all lineages, mycelial growth rate increased with incubation temperature, although this trend was more visible from 10°C to 15°C (increased mean mycelial growth of +2.22 mm/day) than from 25°C to 30°C (+0.05 mm/day) (growth curves and a full statistical treatment of the data are presented in Supplementary file 5; data are reported in Supplementary file 6). Fitting a linear mixed-effects model with incubation time, experimental replicate, and lineage of origin as fixed effects and isolate as a random effect, revealed a significant lineage effect at each incubation temperature [10°C: $F(1,364)=7988$, $p<0.001$; 15°C: $F(1,419)=33161$, $p<0.001$; 20°C: $F(1,542)=40335$, $p<0.001$; 25°C: $F(1,413)=30156$, $p<0.001$; 30°C: $F(1,870)=52681$, $p<0.001$]. Comparing least-squares means resulting from linear mixed-effects models at each temperature revealed a significantly lower growth rate of lineage 4 at 10°C relative to other lineages and a significantly higher growth rate of lineage 1 at 15°C, 20°C, 25°C and 30°C relative to other lineages. Almost no sporulation was detected 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 a full statistical treatment of the data are presented in Supplementary file 6; data are reported in Supplementary file 8). For all lineages, sporulation rates increased with temperature from 15 to 25 °C and dropped at 30°C, although isolates were cultured for only 7 days only at this temperature, versus 10 days at other temperatures, because by this stage, the mycelium had already 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 of lineages based on Dunn's test 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.

340 Together, measurements of mycelial growth and sporulation at different temperatures revealed
341 differences in performance between lineages, but no clear pattern of temperature specialization.
342 We assessed the importance of adaptation to the host, by challenging 45 rice varieties
343 representative of the five main genetic groups of Asian rice (i.e. the temperate japonica, tropical
344 japonica, aus, aromatic and indica subgroups of *Oryza sativa*) with 70 isolates representative of the
345 four lineages of *P. oryzae* and the four clusters within lineage 1 (Supplementary file 9). Interaction
346 phenotypes were assessed qualitatively by scoring resistance (from full to partial resistance: scores 1
347 to 3) or disease symptoms (from weak to full susceptibility: scores 4 to 6) and analysed by fitting a
348 proportional-odds model and performing an analysis of variance (a full statistical treatment of the
349 data is presented in Supplementary file 10). This analysis revealed significant differences between
350 groups of isolates ($\chi^2(6)=100$, $p<0.001$), and between rice genetic groups ($\chi^2(4)=161$, $p<0.001$), and a
351 significant interaction between these two variables ($\chi^2(24)=97$, $p<0.001$). The finding of a significant
352 interaction between groups of isolates (lineages or clusters) and rice types indicates that the effect of
353 the group of isolates on the proportion of compatible interactions differed between rice types,
354 suggesting adaptation to the host. This is also supported by the specific behaviour of certain lineages.
355 Isolates from lineage 2 had much lower symptom scores than the other lineages on all rice types
356 except temperate japonica, and the isolates of the *Yule* cluster were particularly virulent on indica
357 varieties (Figure 4A; Supplementary file 9). In comparisons of rice genetic groups, significantly higher
358 symptom scores were observed on temperate japonica rice than on the other types of rice, whereas
359 the varieties of the aromatic genetic group were significantly more resistant to rice blast (Figure 4B).
360 Together, these experiments therefore revealed significant differences in host range between
361 lineages, but this specialization to the host was not strict, because there was an overlap between
362 host ranges (Figure 4C-G).

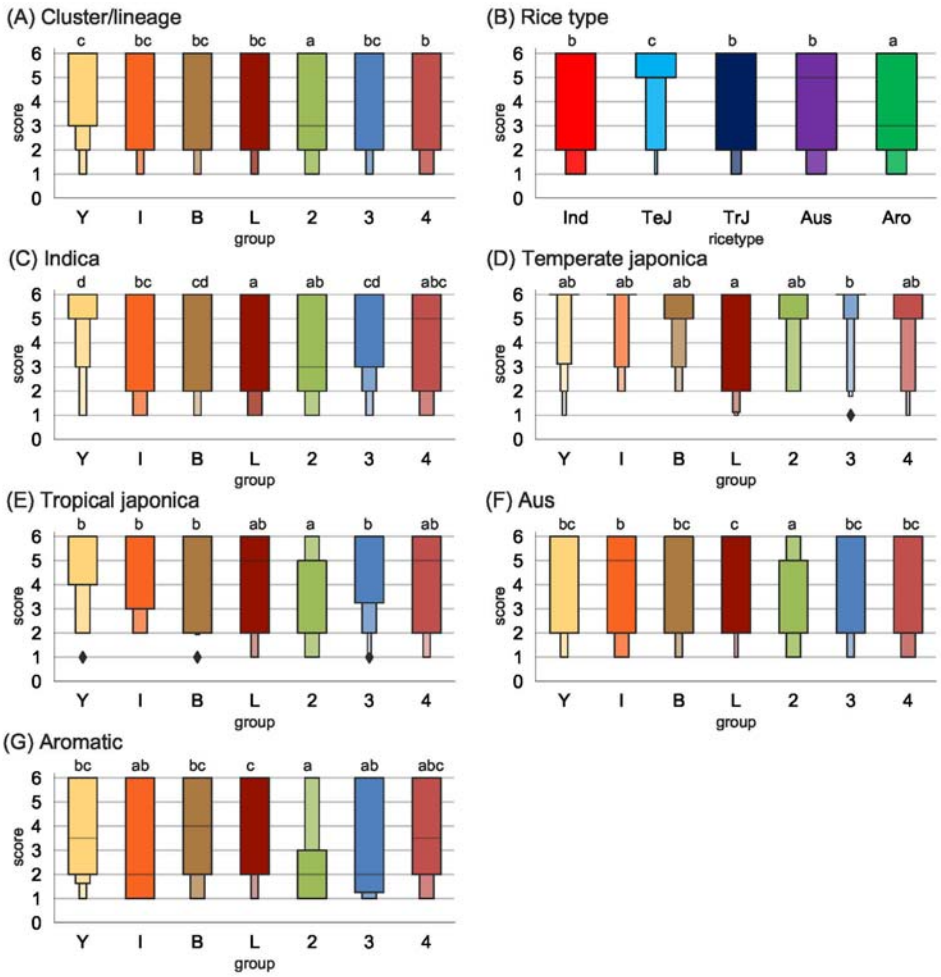

Figure 4: Compatibility between 70 *P. oryzae* isolates and 44 rice varieties, representing five types of rice. Compatibility was measured as symptom scores estimated using pathogenicity tests in controlled conditions. A: Symptom scores as a function of the lineage of origin of isolates, or cluster of origin for isolates from lineage 1; B: Symptom scores as a function of the type of rice; C-G: 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 between rice varieties and *P. oryzae* isolates were assessed in three independent experiments, and the highest of the three symptom scores was used in calculations. Boxen plots were drawn using function boxenplot() with PYTHON package SEABORN 0.11.1. Number of boxes was controlled using option k_depth='trustworthy'. Grey horizontal lines represent the median.

Figure 4 – source data 1

Symptom scores for *P. oryzae* isolates inoculated onto rice varieties

376 **Differences in the number and genetic variability of putative effector genes between *P. oryzae***
377 **lineages**

378 Effectors are the most prominent class of fungal virulence factors. They act by manipulating host
379 cellular pathways and are key determinants of the host range and fitness on compatible hosts (Liao *et*
380 *al.* 2016; Yoshida *et al.* 2016). We therefore determined the differences in effector repertoires
381 between *P. oryzae* lineages in terms of presence/absence and nucleotide polymorphisms, and
382 analyzed whether this was a particularly dynamic fraction of the gene space. To this end, we used
383 whole-genome sequencing data for 123 isolates, including 29 isolates sequenced in this study and 94
384 publicly available genomes (Gladieux *et al.* 2018a; Pordel *et al.* 2020; Zhong *et al.* 2018), 33 of which
385 were also genotyped with our Infinium genotyping beadchip (Supplementary file 11; Appendix 2).
386 Clustering analysis indicated that this dataset covered the four lineages of *P. oryzae*, and three of the
387 clusters of lineage 1 (Laos, International and Baoshan; Appendix 2; Supplementary file 11). Assembly
388 lengths for the 123 sequenced isolates ranged from 36.7 to 39.6 Mb, with a mean value of 38.1 Mb
389 (Supplementary file 11). The number of assembled contigs longer than 500 bp ranged from 1010 to
390 2438, and the longest contig was 1.1 Mb long (Supplementary file 11).

391 Gene prediction identified 11,684 to 12,745 genes per genome, including 1,813 to 2,070
392 genes predicted to encode effector proteins. We found a significant effect of the lineage of origin on
393 both the mean number of putative effector genes (ANOVA, $F=5.54$, $p=0.0014$), and the mean number
394 of non-effector genes (ANOVA, $F=7.95$, $p<0.001$). Multiple comparisons revealed that mean numbers
395 of putative effectors and non-effector genes were only significantly different between the genomes of
396 lineage 2 and lineages 3 (Tukey's HSD test; $p<0.001$; Figure 5A; Figure 5B). An orthology analysis
397 identified 14,573 groups of orthologous sequences (i.e. orthogroups), and we applied a rarefaction
398 approach to the table of orthology relationships to estimate the size of accessory and core genomes
399 in lineages while accounting for differences in sample size. For both putative effectors and the
400 remaining of the gene space, we found that the number of core and accessory genes did not reach a
401 plateau, instead displaying an almost linear relationship with the number of genomes resampled,

402 indicating that the number of core genes was probably much smaller, and the number of accessory
403 genes substantially higher than estimated from 123 genomes (Supplementary file 12). With
404 pseudosamples of size n=30 per lineage (i.e. excluding lineage 4, due to its small sample size), the
405 gene content was highly variable within lineages with only 61-71% of all predicted effectors, and 68-
406 73% of the remaining of the gene space being conserved in lineages 1-3 (Supplementary file 13).
407 Despite extensive variation of the gene content within lineages, the clustering of isolates based on
408 presence/absence variation for both putative effectors and the remaining of the gene space was
409 highly similar to that based on 3,868 SNPs (Figure 5D; Figure 5E), indicating that variation in gene
410 content and SNP allelic variation reflected similar genealogical processes.

411 For the identification of putative effectors potentially involved in the host specialization of
412 rice-infecting lineages of *P. oryzae*, we first identified orthogroups with different patterns of
413 presence/absence across lineages. Principal component analysis on presence/absence data identified
414 72 orthogroups of putative effectors accounting for 95% of the variance of the three principal
415 components differentiating the four lineages (Figure 5C; Figure 5D; Figure 5E), and potentially
416 contributing to the differences in host range between lineages.

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

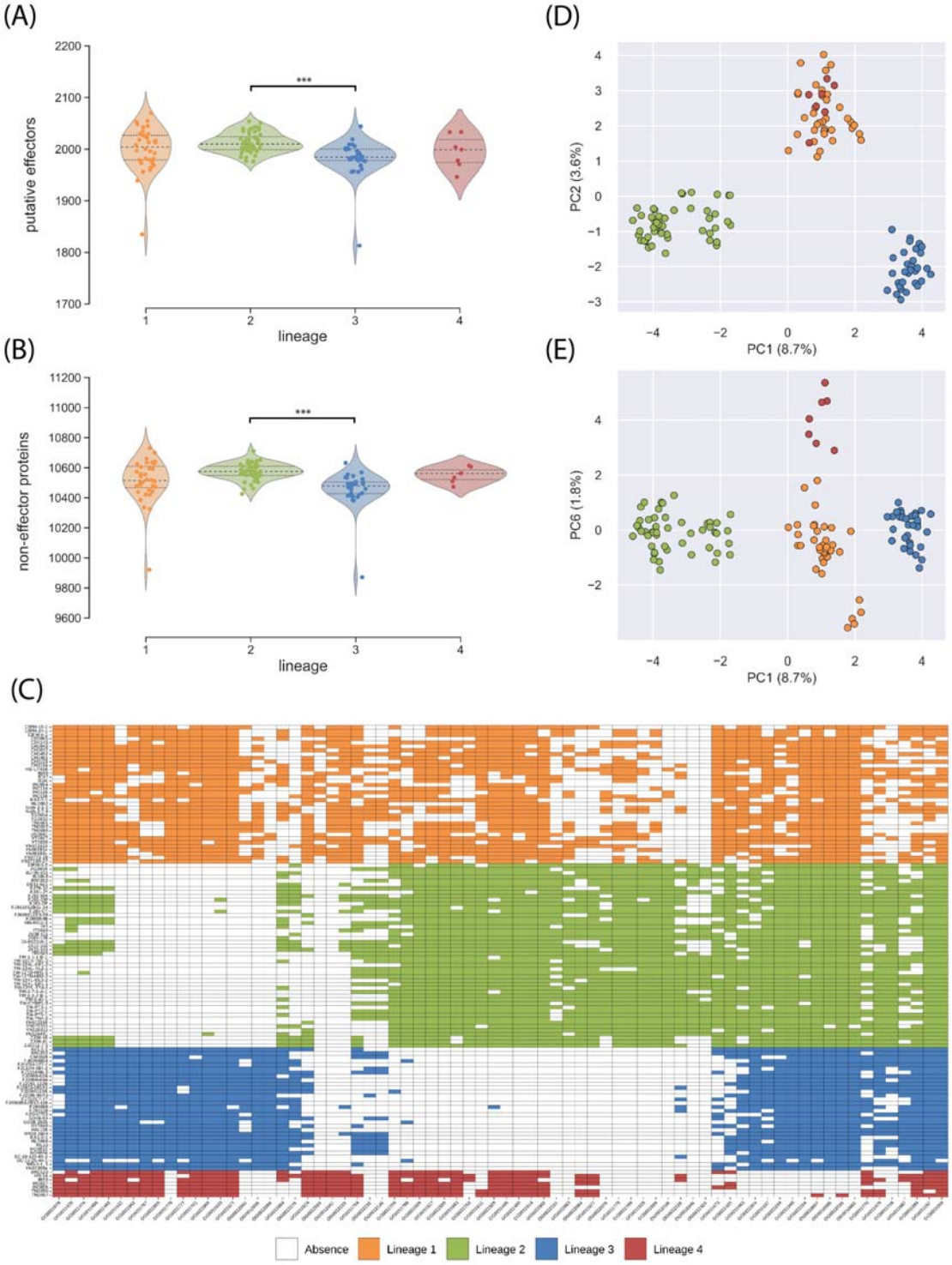

Figure 5. Pangenome analyses of 123 genomes of *P. oryzae*, representing four rice-infecting lineages. (A) and (B) Violin plots showing the number of putative effectors and non-effector proteins detected in each lineage, respectively; asterisks indicate significant differences in gene content (Tukey's HSD test; $p<0.001$). (C) and (D) principal components PC1 against PC2, and PC1 against PC6, respectively, in a principal component analysis of presence/absence of accessory putative effectors and non-effector proteins; between parentheses is the proportion of variance represented by each principal

431 component. (E) Presence and absence in lineage 1-4 of the 72 putative effectors with the highest contribution to principal
432 components 1, 2 and 6, in a principal component analysis of presence/absence data.
433 Figure 5 – source data 1
434 Size of core and accessory genomes for putative effectors and non-effector proteins
435 Figure 5 – source data 2
436 Presence and absence of putative effectors in *P. oryzae* genomes
437 Figure 5 – source data 3
438 Presence and absence of non-effector genes in *P. oryzae* genomes
439 Figure 5 – source data 4
440 Orthology table, with one orthogroup per line and one isolate per column
441

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

443 Our tests of adaptation to host and temperature under controlled conditions showed that
444 specialization was not strict, but it remained possible that fitness differences between lineages would
445 be sufficient under natural conditions to induce separation in different ecological niches, and/or that
446 our experimental conditions did not capture the full extent of the phenotypic differences. Here, we
447 tested the hypothesis that lineages thrive in geographically and/or environmentally different
448 conditions, which would provide indirect evidence for specialization to different habitats. We tested
449 this hypothesis by collecting geographic and climatic data for all isolates, using the GPS position of
450 the nearest city or the geographic center of their region of origin when the exact GPS position of
451 isolates was not available (Supplementary file 1). At a larger scale, clonal lineages 2 and 3 were both
452 widespread, with lineage 2 found on all continents, and lineage 3 on all continents except Europe.
453 Lineage 2 was the only lineage sampled in Europe (with the exception of a single isolate from lineage
454 1), whereas lineage 3 was more widespread in intertropical regions. Lineage 4 was mostly found in
455 South Asia (India, Bangladesh, Nepal), and in the USA and Africa (Benin, Tanzania). The recombining
456 lineage, lineage 1, was found on all continents, but mostly in Asia (79% of all isolates and 94% of all
457 genotypes). At a smaller scale, two, or even three different lineages were sampled in the same year in
458 the same city in 11 different countries from all continents. However, in only one instance did two
459 isolates from different lineages have identical GPS positions (isolates US0106 and US0107 sampled

from the same field). Thus, lineages had different, but partially overlapping distributions, including some overlap with the distribution of the sexually reproducing lineage (lineage 1).

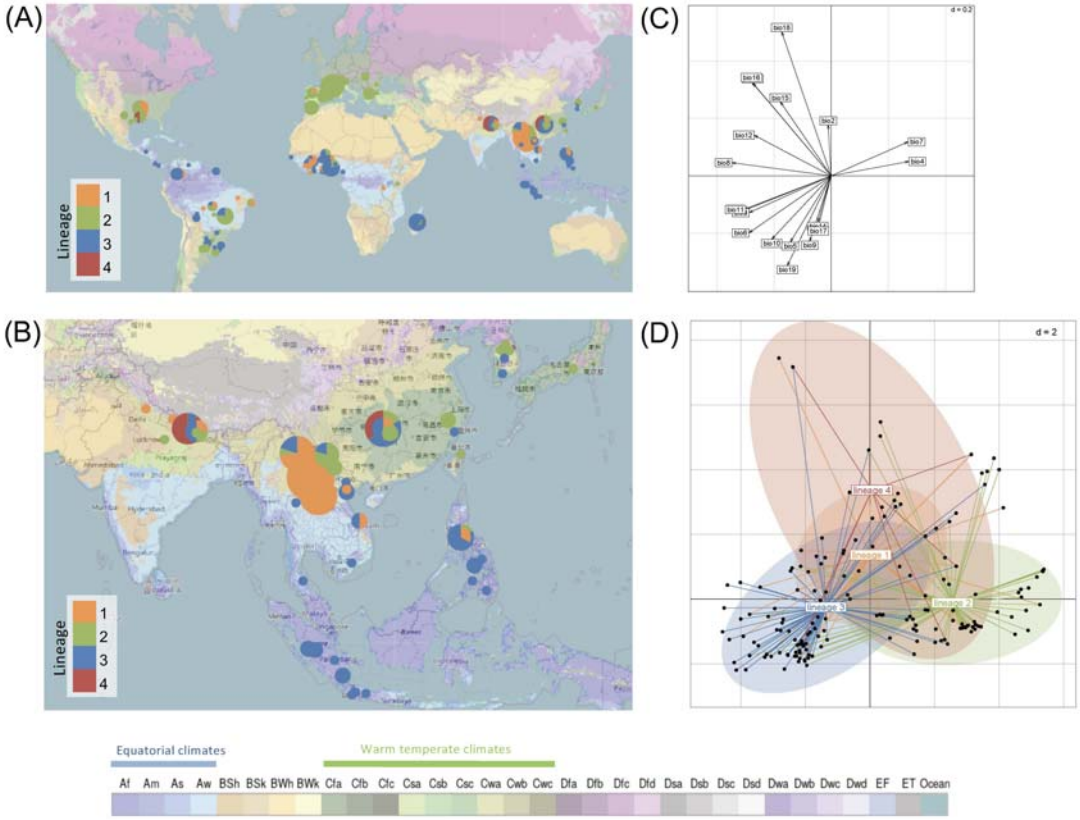

Figure 6. Geographic distribution of four lineages of *P. oryzae* and corresponding climatic data. (A, B) Pie charts representing the distribution of the four lineages in the world (A) and in South, East and Southeast Asia (B), keeping only isolates for which the sampling position was precisely known (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 are listed in the Methods section.

Figure 6 – source data 1

GPS position of isolates

We then investigated whether differences in the geographic range of lineages were associated with differences in climatic variables. By plotting sampling locations onto a map of the major climate regions (Kottek *et al.* 2006) we were able to identify clear differences in the climatic distributions of lineages 2 and 3, with lineage 2 mostly found in warm temperate climates and lineage 3 found in equatorial climates (Figure 6; Source code 1). We used 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. Using information for 19 climatic variables (biomes) from the WorldClim bioclimatic variables database (Fick & Hijmans 2017) for all sampling locations, we obtained statistically significant results, in a permutation test on OMI values, for lineages 2, 3 and 4 (Source code 2; permutation test: lineage 1: OMI=2.14, p=0.620; lineage 2: OMI=6.54, p<0.001; lineage 3: OMI=2.08, p<0.001; lineage 4: OMI=13.73, p=0.017). The OMI analysis, in which the first two axes accounted for 69% and 25% of the variability, respectively, revealed that lineage 2 was more frequent in regions with a high annual temperature range (biome 7) or a high degree of seasonality (biome 4); lineage 4 was associated with regions with high levels of seasonal precipitation (biomes 13, 16 and 18), and lineage 3 was more frequent in regions with high temperatures (biomes 1, 6, 10 and 11) and high levels of isothermality (biome 3), characteristic of tropical climes (Source code 2; Figure 6; Supplementary file 15).

DISCUSSION

This study provides a detailed overview of the worldwide distribution of genetic and phenotypic diversity in *P. oryzae*, and of the eco-evolutionary factors underlying the emergence and maintenance of multiple divergent lineages in this widespread pathogen. We detected three clonal lineages and one recombining lineage with broad and largely overlapping distributions, corresponding to the four groups previously identified on the basis of fewer markers or fewer isolates (Gladieux *et al.* 2018c; Latorre *et al.* 2020; Saleh *et al.* 2014; Tharreau *et al.* 2009; Zhong *et al.* 2018). Our findings provide

strong support for previous findings, and extend these findings by revealing that the lack of recombination and recent admixture in widespread lineages persists with deeper sampling, and by providing evidence for the subdivision of the recombining lineage into three Asian clusters and one international cluster. This deep sampling of the natural genetic variationalso reveals that, despite distributions displaying a broad overlap over large scales, rice blast lineages thrive in areas differing in terms of the prevailing environmental conditions and types of rice grown. We also provide experimental evidence that host-specialization, female sterility and early post-mating genetic incompatibilities act as strong barriers to gene flow between lineages. These results highlight the ability of pathogen spread over heterogeneous habitats and of divergent populations of a crop species to create niche separation and barriers to gene flow between different invasive lineages.

Niche separation and barriers to gene flow between clonal lineages

Many invasive plant pathogens consist of multiple divergent lineages coexisting on the same host, over large areas, raising questions about the maintenance of different lineages in the face of potential gene flow and competitive exclusion. We show 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.

527 However, an analysis of climatic data indicated that lineage 2 predominated in temperate climates,
528 where temperate japonica rice is grown, lineage 3 predominated in tropical climates, in which indica
529 rice is grown, and lineage 4 predominated in regions with high levels of seasonal precipitation in
530 which the indica and aromatic rice types are the principal rice types grown. Despite the finding of
531 separation in different climatic regions, our experiments revealed no strong differences between
532 lineages in terms of sporulation and mycelial growth on synthetic media at different temperatures.
533 This suggests that, if adaptation to temperature occurs in this pathogen, it was not measured by our
534 experiments, either because it involves traits other than sporulation and hyphal growth, or because
535 the *in vitro* conditions were not suitable to demonstrate differences. The host range varied across
536 lineages, but all host ranges overlapped, indicating that host specialization was not strict, or was not
537 fully captured by our experiments. Nevertheless, adaptation to the host or small differences in
538 temperature optima may, nonetheless, further increase niche separation and reduce gene flow
539 between lineages, via pre-mating barriers, such as immigrant non-viability (i.e., lower rates of contact
540 between potential mates due to the mortality of immigrants; (Gladieux *et al.* 2011a, b; Nosil *et al.*
541 2005)), and post-mating barriers, such as ecological hybrid non-viability (i.e., lower survival of ill-
542 adapted hybrid offspring).
543
544 **Pestification can foster lineage divergence**
545 There is a direct connection between disease emergence and speciation, because the emergence of
546 new pathogens can be facilitated, caused by or associated with strong divergent selection in
547 organisms mating within or on their hosts, and by major changes in demographics and geographic
548 isolation (Giraud *et al.* 2006; Hansen 1987; Restrepo *et al.* 2014; Stukenbrock 2013). There is also a
549 connection between disease emergence and mode of reproduction, because disease emergence
550 should be facilitated by multiple asexual cycles, corresponding to multiple cycles of selection for local
551 adaptation without migration introducing locally deleterious immigrant alleles and recombination
552 breaking down locally advantageous allelic combinations (Giraud *et al.* 2006). However, theoretically,

553 in the long term, asexual organisms should accumulate a greater load of deleterious mutations and
554 be less able to fix advantageous mutations, resulting in higher rates of extinction and lower rates of
555 speciation (Orr 2000; Otto & Lenormand 2002). The pattern of population structure revealed in *P.*
556 *oryzae* is seems to be intermediate between observations for invasive fungi with fully clonal
557 population structures — such as the pathogens causing wheat brown rust (Fellers *et al.* 2020),
558 boxwood blight (LeBlanc *et al.* 2020) or verticillium wilt (Short *et al.* 2015; Short *et al.* 2014) — and
559 invasive fungi with barriers maintaining invasive lineages that are permeable and allow hybridization
560 and introgression to occur, such as the pathogens causing *Dothistroma* needle blight (Capron *et al.*
561 2021), Dutch elm disease (Hessenauer *et al.* 2020), soybean anthracnose *Colletotrichum truncatum*
562 (Rogério *et al.* 2021), and coffee rust (Ramírez-Camejo *et al.* 2021; Silva *et al.* 2018). The global
563 structure of *P. oryzae* is more reminiscent of that of the pathogens causing chestnut blight (Dutech *et*
564 *al.* 2012; Stauber *et al.* 2021), wheat yellow rust (Ali *et al.* 2014; Bueno-Sancho *et al.* 2017), and hop
565 powdery mildew (Weldon *et al.* 2021), for which clonal invasive lineages coexist with recombining
566 lineages, although the proximal causes for the loss of sexual reproduction may differ between these
567 models and rice blast. Another marked difference from most of the abovementioned pathogens is
568 that the clonal lineages of rice blast diverged about a thousand years ago (Gladieux *et al.* 2018c;
569 Latorre *et al.* 2020; Zhong *et al.* 2018), corresponding to several thousand asexual generations.
570 Recombining lineage 1 may, at some point, become a source of adaptive introgression, reinvigorating
571 the genomes of clonal lineages to reduce the mutational load, introducing beneficial variation, and
572 leading to marked changes in population structure. Thus, although our work shows that pestification
573 can drive the emergence of reproductive barriers, we cannot predict whether these lineages will
574 continue to exist in their current form for the foreseeable future.
575

576 **Loss of female sterility as the proximal cause of shifts to clonality.**

577 The loss of sexual reproduction, caused by demographic events or selection against sex, is a frequent
578 change in the life history traits of invasive fungal pathogens (Gladieux *et al.* 2015; Hessenauer *et al.*

579 2021). However, the underlying changes and driving factors remain largely unexplored. In *P. oryzae*,
580 the release of constraints maintaining sex in the short term, such as the need for sexual structures
581 for winter survival or long-distance dispersal, is probably not responsible for the loss of sex, because
582 the likelihood of overwintering on seeds and the rate of human-mediated transportation would not
583 be expected to differ between clonal lineages 2-4 and recombining lineage 1. The protection of
584 locally advantageous allelic combinations is a more plausible explanation for the loss of sex in
585 lineages 2-4, because, as shown here, these lineages are found in, and adapted to, different hosts
586 and environmental conditions. This leaves open the question of the maintenance of sex in lineage 1
587 open, but it is possible the diversity of cultivated and wild rice diversity of rice is higher in the
588 geographical range of this lineage (Saleh *et al.* 2012b). It is difficult to draw firm conclusions about
589 the evolutionary causes for selection against sexual reproduction, but the finding of an
590 internationally distributed cluster in lineage 1 sheds light on the possible sequence of proximal
591 events underlying the emergence of clonal lineages in this pathogen. The *International* cluster of
592 lineage 1 displayed genetic footprints of recombination and sexual reproduction (in the form of an
593 excess of homoplastic variants or relatively balanced mating types), but the signal of recombination
594 detected here may be purely historical, because we also found that the clonal fraction and the
595 frequency of sterile females were very high in this cluster. The loss of female fertility may, therefore,
596 be an early and major cause of the shift towards asexual reproduction. Some genotypes assigned to
597 the international cluster of lineage 1 may, in the future, come to resemble other clonal lineages in
598 analyses of genealogical relationships, and form a long branch stemming from the central lineage 1,
599 particularly for genotypes already displaying clonal propagation, such as genotypes 2, 18, 58, 98 and
600 254. However, according to the principle of competitive exclusion, competition between clonal
601 lineages should only allow hegemonic expansion of the most competitive clonal groups, unless they
602 separate into different ecological niches. The asexual spread of the International cluster in a separate
603 ecological niche may ultimately lead to the fixation of a single mating type, as in clonal lineages 2-4.
604 This model could also apply to other fungal pathogens with similar life history strategies.

605

606 **Differences in effector repertoires between lineages**

607 Effector repertoires play a key role in pathogen specialization, in fungal pathogens in general
608 (Schulze-Lefert & Panstruga 2011b), and *P. oryzae* in particular (Asuke *et al.* 2020; Couch *et al.* 2005;
609 Liao *et al.* 2016; Sweigard *et al.* 1995; Takabayashi *et al.* 2002). Previous work suggested that clonal
610 lineages of *P. oryzae* possess smaller effector repertoires (Latorre *et al.* 2020) and that the clonal
611 lineage associated with japonica rice (lineage 2) possesses more effectors, and, in particular, Avr-
612 effectors, than other clonal lineages (Latorre *et al.* 2020; Liao *et al.* 2016). Our analyzes on a larger
613 set of putative effectors (ca. 2000 in our study, vs. 13 Avr-effectors in (Liao *et al.* 2016) and 178
614 known and candidate effectors in (Latorre *et al.* 2020)) confirm this trend, because lineage 2
615 associated with temperate japonica has significantly more putative effectors than lineage 3
616 associated with indica. However, the number of non-effector genes is also greater in lineage 2 than in
617 lineage 3, and thus it remains possible that the larger effector complement of lineage 2 is a mere
618 consequence of a larger number of genes. We also found that patterns of presence / absence
619 variation mirrored patterns of population subdivision based on SNP allelic variation, and, thus, that
620 the differential sorting of both nucleotide polymorphism and gene content across lineages reflects
621 similar genealogical processes. Our multivariate analyses identified 72 effectors making the greatest
622 contribution to the differentiation of the four lineages in terms of presence / absence. It is possible
623 that the frequency differences for these 72 effectors were due to chance events during bottlenecks
624 at the onset of lineage formation, or, alternatively, that their differential loss was important for the
625 initial adaptation of lineages to new rice populations or subspecies. Interestingly, two of these 72
626 effectors, AvrRmg8 (=OG0011611) and PWL3 (=OG0011928), are host range determinants in wheat-
627 infecting *P. oryzae* because they trigger immunity in wheat varieties carrying the resistance genes
628 *Rmg8* (Anh *et al.* 2018) or *PWT3* (Inoue *et al.* 2017; Kang *et al.* 1995). Nucleotide diversity at
629 synonymous and non-synonymous sites also identified 185 effectors with signatures of diversifying
630 selection in one or more lineages, possibly mediated by coevolutionary interactions with host

631 molecules. In the future, it will be interesting to determine the molecular targets of these effectors
632 and to decipher the relationship between their polymorphism and their mode of action.
633
634 **Lineage 1 as a threat to global rice production**
635 Our results also indicate that lineage 1 may pose a major threat to rice production. Most of the
636 effectors with signatures of diversifying selection were identified in lineage 1, highlighting the greater
637 ability of this lineage to fix advantageous mutations rapidly. Eleven of the 12 most multivirulent
638 isolates (lesion type > 2 on more than 40 varieties tested) belonged to lineage 1, and the *Yule* cluster,
639 in particular, was highly pathogenic on indica varieties. The propagation of such highly pathogenic
640 genotypes, belonging to a lineage that has both mating types and fertile females, should be
641 monitored closely. The monitoring of lineage 1 is all the more critical because this lineage has an
642 intermediate geoclimatic distribution that overlaps largely with the distributions of the other three
643 lineages, and some of the attempted crosses between clonal lineage 2-4 isolates and recombining
644 lineage 1 isolates produced viable progeny. This finding confirms the possibility of gene flow into this
645 lineage, as previously demonstrated on the basis of admixture mapping (Gladieux *et al.* 2018b;
646 Latorre *et al.* 2020), but it also raises the question of the threat posed by gene flow from lineage 1,
647 which is highly pathogenic, to the other lineages.
648
649 **Concluding remarks**
650 Our study of genetic and phenotypic variation within and between clonal lineages, and within the
651 recombinant lineage of *P. oryzae* suggests a scenario for the emergence of widespread clonal
652 lineages. The loss of female fertility may be a potent driver of the emergence of asexually
653 reproducing groups of clones. The reproductive isolation generated by the loss of sex and the
654 accumulation of mutations due to the absence of sexual purging would facilitate the specialization of
655 some of these clonal groups, leading to the competitive exclusion of the least efficient clonal groups,
656 and, finally, to the propagation of clonal lineages fixed for a single mating type. Our results, thus,

657 demonstrate that the spread of a pathogen across heterogeneous habitats and divergent populations
658 of a crop species can lead to niche separation and reproductive isolation between different invasive
659 lineages in the pathogen.
660

661 **MATERIALS AND METHODS**

662 **Biological material**

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

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

673 Forty-four varieties were chosen as representative of the five main genetic subgroups of
674 Asian rice (Garris *et al.* 2005): indica (Chau, Chiem chanh, DA11, De abril, IR8, JC120, Pappaku), aus
675 (Arc 10177, Baran boro, Black gora, DA8, Dholi boro, Dular, FR13 A, JC148, Jhona 26, Jhona 149,
676 Kalamkati, T1, Tchampa, Tepi boro), temperate japonica (Aichi asahi, Kaw luyoeng, Leung pratew,
677 Maratelli, Nep hoa vang, Nipponbare, Sariceltik, Som Cau 70A), tropical japonica (Azucena,
678 Binulawan, Canella de ferro, Dholi boro, Gogo lempuk, Gotak gatik, Trembese) and aromatic (Arc
679 10497, Basmati lamo, Dom zard, Firooz, JC1, Kaukkyisaw, N12). The Maratelli (temperate japonica)
680 and CO39 (indica) varieties were used as susceptible controls.
681

682 **Genotyping**

683 *Pyricularia oryzae* isolates were genotyped at 5,657 genomic positions with an Illumina Infinium
684 beadchip microarray carrying single-nucleotide polymorphisms identified in 49 previously
685 characterized genomes of rice- and barley-infecting *P. oryzae* isolates (Gladieux et al., 2018b). The
686 final dataset included 3,686 biallelic SNPs without missing data.
687

688 **Whole genome sequencing**

689 We used a set of 123 whole-genome sequences to investigate differences in effector content
690 between clusters and lineages, including 94 publicly available genomes and 29 new genomes that we
691 selected to increase sample size for the International and Laos clusters of lineage 1, and for lineage 4
692 (Supplementary file 11). The 29 isolates were sequenced using Illumina HiSeq 3000. Isolates were
693 grown on rice flour-agar medium for mycelium regeneration, then in liquid rice flour medium (Adreit
694 *et al.* 2007). Genomic DNA extraction was carried out with >100 mg of fresh mycelium from liquid
695 culture. Fresh mycelium dried on Miracloth paper was crushed in liquid nitrogen. Nucleic acids were
696 subsequently extracted using an extraction buffer (2 % CTAB – 1.4 M NaCl – 0.1 M Tris-HCl pH 8 – 20
697 mM EDTA pH 8 added before use with a final concentration of 1 % Na_2SO_3), then purified with a
698 chloroform:isoamyl alcohol (24:1) treatment, precipitated overnight in isopropanol, and rinsed with
699 70% ethanol. The extracted nucleic acids were further treated with RNase A (0.2 mg/mL final
700 concentration) and purified with another chloroform:isoamyl alcohol (24:1) treatment followed by
701 overnight precipitation in ethanol. The concentration of extracted genomic DNA was assessed on
702 Qubit® using the dsDNA HS Assay Kit. The purity of the extracted DNA was checked by verifying that
703 the 260/280 and 260/230 absorbance ratios measured with a NanoDrop spectrophotometer were
704 between 1.8 and 2.0. Preparation of TruSeq nano library preparation and HiSeq3000 sequencing (150
705 nucleotide reads, and 500 bp insert size) were performed at GeT-PlaGe (INRAE, Toulouse, France).
706

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

709 For the 123 sequenced isolates included in the dataset, low-quality reads were removed with
710 CUTADAPT software (Martin 2011). Reads were assembled with ABYSS 2.2.3 (Jackman *et al.* 2017;
711 Simpson *et al.* 2009), using different K-mer sizes. For each isolate, we chose the assembled sequence
712 with the highest N50 for further analyses. Genes were predicted with BRAKER 2.1.5 (Hoff *et al.* 2015;
713 Hoff *et al.* 2019) using RNAseq data (Pordel *et al.* 2020) as extrinsic evidence for model refinement.
714 Genes were also predicted with AUGUSTUS 3.4.0 (Stanke & Morgenstern 2005)(training
715 set=*Magnaporthe grisea*) and gene models that did not overlap with the gene models identified with
716 BRAKER were added to the GFF file generated with the latter. Repeated regions were masked with
717 REPEATMASKER 4.1.0 (http://www.repeatmasker.org/). The completeness of genome assemblies and
718 gene predictions was assessed with BUSCO (Simão *et al.* 2015), and completeness in BUSCO genes
719 was greater than 93.7% (Supplementary file 11). Putative effector genes were identified as genes
720 encoding proteins predicted to be secreted by at least two of three methods [SIGNALP 4.1 (Nielsen
721 2017), TARGETP (Emanuelsson *et al.* 2000) and PHOBIUS (Käll *et al.* 2004)], with no predicted
722 transmembrane domain based on TMHMM analysis (Krogh *et al.* 2001), no predicted motif of
723 retention in the endoplasmic reticulum based on PS-SCAN (Bhagwat & Aravind 2007), and no CAZy
724 annotation based on DBSCAN V7 (Yin *et al.* 2012). Differences in the numbers of putative effectors and
725 non-effectors between lineages (ANOVA and Tukey's HSD test) were assessed with the SCIPY 1.6.0
726 package in PYTHON. Homology relationships betweenpredicted genes were established with
727 ORTHOFINDER v2.4.0 (Emms & Kelly 2015). Principal component analysis of presence/absence
728 variation for putative effectors and non-effector proteins was performed with the R package PRCOMP.
729 We used the get_pca_var() in PRCOMP to extract the results for variables (i.e. effectors) and to identify
730 the effectors making the largest contributions to the principal components, defined as effectors with
731 a contribution to loadings of PC1, PC2 and PC6 greater than 1%. For estimates of the numbers of core
732 and accessory genes, we used a rarefaction approach to account for differences in sample size across
733 lineages. For each pseudo-sample size, we estimated the size of the core and accessory genome in
734 each lineage using a maximum of 2000 pseudosamples. Sequences for each orthogroup were aligned

735 and cleaned with TRANSLATORX (Abascal *et al.* 2010), using default parameters. The ratio of non-
736 synonymous to synonymous diversity π_N/π_S was calculated for each orthogroup, for lineages with a
737 sample size of at least four, with EGGLIB 3 (https://www.egglib.org).
738

739 **Population subdivision and recombination**

740 The dataset used for population genetic analysis had no missing data. Clones were therefore
741 identified as isolates with identical genotypes at 3,686 sites. Among the 886 *P. oryzae* isolates
742 genotyped, we identified 264 different multilocus genotypes (i.e., "clones"), which were used for the
743 analysis of population subdivision. We used the sNMF program to infer individual ancestry
744 coefficients in *K* ancestral populations. This program is optimized for the analysis of large datasets
745 and does not assume Hardy-Weinberg equilibrium. It is, therefore, more appropriate to deal with
746 inbred or clonal lineages (Frichot *et al.* 2014) . We used SPLITSTREE version 4 (Huson & Bryant 2006) to
747 visualize relationships between genotypes in a phylogenetic network, with reticulations to represent
748 the conflicting phylogenetic signals caused by homoplasy. PCA was performed as implemented in the
749 PYTHON package SCIKIT-ALLEL v. 1.3.3. (https://zenodo.org/record/4759368#.YL4R5DaiHwQ). Weir and
750 Cockerham's F_{ST} was calculated with the R package HIERFSTAT, by the WC84 method (Goudet 2005).
751 We tested the null hypothesis of clonality using the pairwise homoplasy index (PHI; (Bruen *et*
752 *al.* 2006)), maximum χ^2 (Max χ^2, (Maynard Smith 1992)) and neighbour similarity score (NSS;
753 (Jakobsen & Easteal 1996)), as implemented in the PHIPACK program
754 (https://www.maths.otago.ac.nz/~dbryant/software.html). Homoplastic sites have sequence
755 identities that are not inherited from a common ancestor, but instead derived from independent
756 events in different branches, such as recurrent mutations, sequencing errors or recombination.
757 Homoplastic sites were identified by concatenating genotypes at all sites, inferring a genealogy with
758 the GTR+GAMMA model in RAXML v.8.2.12 (Stamatakis 2014), and mapping mutations onto the
759 genealogy with the 'Trace All Characters' function of MESQUITE under maximum parsimony (Maddison
760 & Maddison 2019). In the resulting matrix of ancestral states for all nodes, the number of mutations

761 occurring at each site was counted and sites displaying multiple mutations across the genealogy were
762 considered to be homoplastic. We used PopLDDecay (Zhang *et al.* 2019) to measure linkage
763 disequilibrium (r^2) as a function of distance between pairs of SNPs. PopLDDecay was configured with a
764 maximum distance between SNPs of 400 kbp, and a minor allele frequency of 0.005.
765

766 **Experimental measurement of reproductive success and female fertility**

767 *Pyricularia oryzae* is a heterothallic fungus with two mating types (*Mat1.1* and *Mat1.2*). Sexual
768 reproduction between strains of opposite mating type can be observed in laboratory conditions, and
769 results in the production of ascospores within female sexual structures called perithecia (Saleh *et al.*
770 2012a). Crosses were carried out on a rice flour agar medium (20 g rice flour, 2.5 g yeast extract, 15 g
771 agar and 1 L water, supplemented with 500 000 IU penicillin G after autoclaving for 20 min at 120°C),
772 as described by (Saleh *et al.* 2012a). We assessed reproductive success by determining the
773 production of perithecia produced after three weeks of culture at 20°C under continuous light. Each
774 cross was performed twice and we determined the mean number of perithecia across repeats. We
775 further assessed the presence of asci and germinating ascospores in perithecia for a subset of
776 crosses, by excising perithecia with a scalpel and counting the number of germinated filaments for
777 each individual ascus after overnight incubation on water agar. The subset of crosses included 10
778 Mat1.1 isolates (lineage 1: CH0999, CH1065, CH1076; lineage 2: CH0092, MC0016, SP0006; lineage 4:
779 IN0017, IN0092, NP0070, CH0718) and 6 Mat1.2 isolates (lineage 1: CH0997; CH1083, CH1120;
780 lineage 3: BR0019, CH0549, MD0929).

781 We measured female fertility for 210 isolates (listed in Supplementary file 1), by monitoring
782 perithecium production in crosses involving the tester strains CH0997 (Mat1.2) and CH0999 (Mat1.1).
783 On synthetic media, perithecia are formed at the zone of contact between parental mycelia. Isolates
784 forming perithecia are described as female-fertile. Perithecia can be formed by the two interacting
785 partners or by one partner only. If the two partners are female-fertile, two parallel lines of perithecia

786 are observed at the contact zone between mycelia. If only one partner is female-fertile, only one line
787 of perithecia is observed, at the edge of the mycelium of the fertile isolate.
788
789 **Pathogenicity tests**
790 Compatibility between *P. oryzae* isolates and rice plants from the five main genetic subgroups of rice
791 (indica, temperate japonica, tropical japonica, aus and aromatic) was measured in controlled
792 conditions. We inoculated on 46 varieties (see section Biological Material), with 70 isolates.
793 Inoculations were performed as described by (Gallet *et al.* 2016). Conidial suspensions (25 000
794 conidia.mL^{-1}) in 0.5% gelatin were sprayed onto three-week-old rice seedlings (> 6 plants/variety).
795 The inoculated plants were incubated for 16 hours in the dark at 27°C and 100% humidity and then
796 for seven days with a day/night alternation (13 hours at 27°C/11 hours at 21°C), before the scoring of
797 symptoms. Lesion type was rated from 1 to 6 (Gallet *et al.* 2016) and the symptom type was assessed
798 visually on leaves. Each interaction was assessed in three independent experiments. Symptoms
799 scores are ordinal-scale variables (i.e. variables that can be ranked but are not evenly spaced). We
800 therefore analyzed the data with a proportional-odds model with the clm() function implemented in
801 the ORDINAL version 2018.8-25 in R, including rice type and pathogen lineage as main effects, and the
802 rice type – pathogen lineage interaction. A comprehensive mixed model including all experimental
803 replicates, with isolate as a random effect, was also tested, but did not display convergence. We
804 therefore used the maximum of the three scores in calculations, as this is what most accurately
805 represented the ability of an isolate to exploit a given host genotype. Analyses were also conducted
806 with median values, and gave similar results (not shown). We used an analysis of deviance to
807 evaluate whether effect and interaction terms were statistically significant. Pairwise comparisons of
808 significant factors were performed after computing least-squares means with LSMEANS version 2.30-0
809 in R, with Tukey adjustment. The R commands and their outputs are provided in Supplementary file
810 9.
811

812 **Mycelial growth and sporulation rate at different temperatures**

813 Mycelial growth and sporulation rate were measured for 41 isolates at five different temperatures
814 (10°C, 15°C, 20°C, 25°C and 30°C). For each isolate, Petri dishes containing PDA medium were
815 inoculated with mycelial plugs placed at the center, and incubated in a growth chamber at a fixed
816 temperature. Mycelium diameter was estimated as the mean of two measurements made along two
817 perpendicular axes at different time points. At the end of the experiment, conidia were collected by
818 adding 5 mL of water supplemented with 0.01% Tween 20 to the Petri dish and rubbing the surface of
819 the mycelium. Conidia were counted with a hemocytometer. Three or four independent experiments
820 were performed for each isolate, at each temperature. We had initially planned to carry out only
821 three independent experiments. However, a fourth experiment was eventually performed for some
822 temperature conditions because some isolates did not grow in the first experiment, and because
823 some cultures were invaded by mites and had to be discarded in the second and third experiments.

824 Mycelium growth was analyzed independently for each temperature with linear mixed-
825 effects models. The variable used in statistical analysis was the square root of the mean mycelial
826 diameter measured along two perpendicular axes on a Petri plate. Post-hoc pairwise comparisons
827 were performed after computing least-squares means with LSMEANS version 2.30-0 in R, with Tukey
828 adjustment. The R commands and their outputs are provided in Supplementary file 4.

829 Sporulation rates were analyzed with a Kruskal-Wallis test with RSTATIX version 0.7.0 in R. The
830 variable used in statistical analysis was the median number of spores calculated across replicates.
831 Post-hoc pairwise comparisons were performed with Dunn's non-parametric multiple comparison
832 test. The R commands and their outputs are provided in Supplementary file 7.

833

834 **Statistical analyses of climatic data**

835 The outlying mean index (OMI), or marginality, is used to study niche separation and niche breadth.
836 The OMI approach evaluates whether a particular lineage is mostly associated with some climatic
837 conditions or can be sampled with the same probability throughout the range of the species. The

838 approach is based on a multivariate analysis (here a PCA) of climatic data associated with the
839 locations included in the dataset, regardless of the identity of the lineages sampled or the number of
840 isolates sampled. An index of marginality or specialization (the outlying mean index) is then
841 calculated for each lineage represented in the dataset. This index measures the distance between
842 the average habitat of a given lineage and the average habitat conditions of the area studied
843 (corresponding to the distribution of a hypothetical species uniformly distributed under all
844 conditions). Lineages are then placed in environmental conditions that maximize their OMIs, and
845 permutation tests are used to test whether the distribution of a lineage differs from random. OMI
846 gives the same weight to all sampling locations, whether rich or poor in individuals or lineages, and is
847 particularly suitable in cases in which sampling is not homogeneous. Environmental values, consisting
848 of 19 biome values (WorldClim bioclimatic variables (Fick & Hijmans 2017)), were retrieved for all
849 sampling locations. The 11 temperature variables included in the analysis were (bio1) annual mean
850 temperature, (bio2) mean diurnal range, (bio3) isothermality [100*(bio2/bio7)], (bio4) temperature
851 seasonality [standard deviation*100], (bio5) maximum temperature of the warmest month, (bio6)
852 minimum temperature of the coldest month, (bio7) annual temperature range [bio5-bio6], (bio8)
853 mean temperature of the wettest quarter, (bio9) mean temperature of the driest quarter, (bio10)
854 mean temperature of the warmest quarter, (bio11) mean temperature of the coldest quarter. The
855 eight precipitation variables included in the analysis were (bio12) annual precipitation, (bio13)
856 precipitation of the wettest month, (bio14) precipitation of the driest month, (bio15) precipitation
857 seasonality (coefficient of variation), (bio16) precipitation of the wettest quarter, (bio17)
858 precipitation of the driest quarter, (bio18) precipitation of the warmest quarter, (bio19) precipitation
859 of the coldest quarter. The 11 climatic variables were normalized before analysis. A contingency table
860 was generated to associate the number of isolates from each lineage with each sampling location.
861 Only isolates from our dataset with a precisely known geographic origin (region, city or GPS position)
862 were included in the analysis. A random permutation test with 10,000 permutations was used to
863 assess the statistical significance of marginality for each lineage.

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 Puccinia striiformis f. sp tritici. *Plos Pathogens* **10**.

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.

Asuke S, Tanaka M, Hyon G-S*, et al.* (2020) Evolution of an Eleusine-specific subgroup of Pyricularia oryzae through a gain of an avirulence gene. *Molecular Plant-Microbe Interactions* **33**, 153-165.

Barrett LG, Thrall PH, Burdon JJ, Linde CC (2008) Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends in Ecology & Evolution* **23**, 678-685.

Bhagwat M, Aravind L (2007) Psi-blast tutorial. In: *Comparative genomics*, pp. 177-186. Springer.

Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665-2681.

Bueno-Sancho V, Persoons A, Hubbard A*, et al.* (2017) Pathogenomic Analysis of Wheat Yellow Rust Lineages Detects Seasonal Variation and Host Specificity. *Genome Biology and Evolution* **9**, 3282-3296.

Capron A, Feau N, Heinzelmann R*, et al.* (2021) Signatures of post-glacial genetic isolation and human-driven migration in the Dothistroma needle blight pathogen in western Canada. *Phytopathology®* **111**, 116-127.

Couch BC, Fudal I, Lebrun M-H*, et al.* (2005) Origins of Host-Specific Populations of the Blast Pathogen *Magnaporthe oryzae* in Crop Domestication With Subsequent Expansion of Pandemic Clones on Rice and Weeds of Rice. *Genetics* **170**, 613-630.

de Vienne DM, Giraud T, Gouyon P-H (2013) Lineage selection and the maintenance of sex. *Plos One* **8**, e66906.

Desprez-Loustau ML, Robin C, Buée M*, et al.* (2007) The fungal dimension of biological invasions. *Trends in Ecology & Evolution* **22**, 472-480.

Dutech C, Barres B, Bridier J, *et al.* (2012) The chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. *Molecular Ecology* **21**, 3931-3946.

Emanuelsson O, Nielsen H, Brunak S, Von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of molecular biology* **300**, 1005-1016.

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, 157.

Farman M, Peterson G, Chen L, *et al.* (2017) The Lolium pathotype of Magnaporthe oryzae recovered from a single blasted wheat plant in the United States. *Plant Disease* **101**, 684-692.

Fellers JP, Sakthikumar S, He F, *et al.* (2020) Whole-genome sequencing of multiple isolates of Puccinia triticina reveals asexual lineages evolving by recurrent mutations. *bioRxiv*.

Fellers JP, Sakthikumar S, He F, *et al.* (2021) Whole-genome sequencing of multiple isolates of Puccinia triticina reveals asexual lineages evolving by recurrent mutations. *G3 Genes|Genomes|Genetics*.

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology* **37**, 4302-4315.

Frenkel O, Peever TL, Chilvers MI, *et al.* (2010) Ecological Genetic Divergence of the Fungal Pathogen Didymella rabiei on Sympatric Wild and Domesticated Cicer spp. (Chickpea). *Applied and Environmental Microbiology* **76**, 30-39.

Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014) Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, 973-983.

Gallet R, Fontaine C, Bonnot F, *et al.* (2016) Evolution of compatibility range in the rice–magnaporthe oryzae system: an uneven distribution of R genes between rice subspecies. *Phytopathology* **106**, 348-354.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in Oryza sativa L. *Genetics* **169**, 1631-1638.

Giraud T, Gladieux P, Gavrilets S (2010) Linking the emergence of fungal plant diseases with ecological speciation. *Trends in ecology & evolution* **25**, 387-395.

Giraud T, Koskella B, Laine AL (2017) Introduction: microbial local adaptation: insights from natural populations, genomics and experimental evolution. *Molecular ecology* **26**, 1703-1710.

Giraud T, Villareal LMMA, Austerlitz F, Le Gac M, Lavigne C (2006) Importance of the Life Cycle in Sympatric Host Race Formation and Speciation of Pathogens. *Phytopathology* **96**, 280-287.

Gladieux P, Condon B, Ravel S, *et al.* (2018a) Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus Magnaporthe oryzae. *mBio* **9**.

Gladieux P, Feurtey A, Hood ME, *et al.* (2015) The population biology of fungal invasions. *Molecular Ecology*, 1969-1986.

Gladieux P, Guerin F, Giraud T, *et al.* (2011) Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Molecular Ecology* **20**, 4521-4532.

Gladieux P, Ravel S, Rieux A, *et al.* (2018b) Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. *MBio* **9**.

Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* **5**, 184-186.

Hansen EM (1987) Speciation in plant pathogenic fungi - the influence of agricultural practice. *Canadian Journal of Plant Pathology* **9**, 403-410.

Hardin G (1960) The competitive exclusion principle. *Science* **131**, 1292-1297.

Hessenauer P, Feau N, Gill U, *et al.* (2021) Evolution and adaptation of forest and crop pathogens in the Anthropocene. *Phytopathology®* **111**, 49-67.

Hessenauer P, Fijarczyk A, Martin H, *et al.* (2020) Hybridization and introgression drive genome evolution of Dutch elm disease pathogens. *Nature ecology & evolution* **4**, 626-638.

957 Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M (2015) BRAKER1: unsupervised RNA-Seq-
958 based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767-769.
959 Hoff KJ, Lomsadze A, Borodovsky M, Stanke M (2019) Whole-genome annotation with BRAKER. In:
960 *Gene Prediction*, pp. 65-95. Springer.
961 Huson DH, Bryant D (2006) Application of Phylogenetic Networks in Evolutionary Studies. *Molecular*
962 *Biology and Evolution* **23**, 254-267.
963 Inoue Y, Vy TTP, Yoshida K*, et al.* (2017) Evolution of the wheat blast fungus through functional losses
964 in a host specificity determinant. *Science* **357**, 80-83.
965 Islam MT, Croll D, Gladieux P*, et al.* (2016) Emergence of wheat blast in Bangladesh was caused by a
966 South American lineage of Magnaporthe oryzae. *BMC biology* **14**, 84.
967 Jackman SD, Vandervalk BP, Mohamadi H*, et al.* (2017) ABySS 2.0: resource-efficient assembly of
968 large genomes using a Bloom filter. *Genome Research* **27**, 768-777.
969 Jakobsen IB, Easteal S (1996) A program for calculating and displaying compatibility matrices as an
970 aid in determining reticulate evolution in molecular sequences. *Bioinformatics (Oxford,*
971 *England)* **12**, 291-295.
972 Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide
973 prediction method. *Journal of molecular biology* **338**, 1027-1036.
974 Kang S, Sweigard Ja Fau - Valent B, Valent B (1995) The PWL host specificity gene family in the blast
975 fungus Magnaporthe grisea. *Molecular Plant Microbe Interactions*.
976 Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate
977 classification updated.
978 Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein
979 topology with a hidden Markov model: application to complete genomes. *Journal of*
980 *molecular biology* **305**, 567-580.
981 Kumar J, Nelson RJ, Zeigler RS (1999) Population structure and dynamics of Magnaporthe grisea in
982 the Indian Himalayas. *Genetics* **152**, 971-984.
983 Langner T, Białas A, Kamoun S (2018) The blast fungus decoded: Genomes in flux. *mBio* **9**, e00571-
984 00518.
985 Latorre SM, Reyes-Avila CS, Malmgren A*, et al.* (2020) Differential loss of effector genes in three
986 recently expanded pandemic clonal lineages of the rice blast fungus. *BMC biology* **18**, 1-15.
987 LeBlanc N, Cubeta MA, Crouch JA (2020) Population Genomics Trace Clonal Diversification and
988 Intercontinental Migration of an Emerging Fungal Pathogen of Boxwood. *Phytopathology®*
989 **111**, 184-193.
990 Lemaire C, De Gracia M, Leroy T*, et al.* (2016) Emergence of new virulent populations of apple scab
991 from nonagricultural disease reservoirs. *New Phytologist* **209**, 1220-1229.
992 Li F, Upadhyaya NM, Sperschneider J*, et al.* (2019) Emergence of the Ug99 lineage of the wheat stem
993 rust pathogen through somatic hybridisation. *Nature Communications* **10**, 5068.
994 Liao J, Huang H, Meusnier I*, et al.* (2016) Pathogen effectors and plant immunity determine
995 specialization of the blast fungus to rice subspecies. *eLife* **5**, e19377.
996 Lin X, Heitman J (2006) The biology of the Cryptococcus neoformans species complex. *Annu. Rev.*
997 *Microbiol.* **60**, 69-105.
998 Maddison WP, Maddison DR (2019) Mesquite: a modular system for evolutionary analysis. Version
999 3.61
1000 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
1001 *EMBnet. journal* **17**, 10-12.
1002 Maynard Smith J (1992) Analyzing the mosaic structure of genes. *Journal of molecular evolution* **34**,
1003 126-129.
1004 Mboup M, Bahri B, Leconte M*, et al.* (2012) Genetic structure and local adaptation of European
1005 wheat yellow rust populations: the role of temperature-specific adaptation. *Evolutionary*
1006 *Applications* **5**, 341-352.

1007 Mercier A, Simon A, Lapalu N, *et al.* (2021) Population genomics reveals molecular determinants of
1008 specialization to tomato in the polyphagous fungal pathogen Botrytis cinerea in France.
1009 *Phytopathology*.
1010 Milazzo J, Pordel A, Ravel S, Tharreau D (2019) First scientific report of Pyricularia oryzae causing gray
1011 leaf spot disease on perennial ryegrass (Lolium perenne) in France. *Plant Disease* **103**, 1024-
1012 1024.
1013 Nielsen H (2017) Predicting secretory proteins with SignalP. In: *Protein function prediction*, pp. 59-73.
1014 Springer.
1015 Nosil P, Vines TH, Funk DJ (2005) Perspective: reproductive isolation caused by natural selection
1016 against immigrants from divergent habitats. *Evolution; international journal of organic*
1017 *evolution* **59**, 705-719.
1018 Orr HA (2000) The rate of adaptation in asexuals. *Genetics* **155**, 961-968.
1019 Otto SP, Lenormand T (2002) Resolving the paradox of sex and recombination. *Nature Reviews*
1020 *Genetics* **3**, 252-261.
1021 Palm ME (2001) Systematics and the Impact of Invasive Fungi on Agriculture in the United
1022 States. *Bioscience* **51**, 141-147.
1023 Peever TL (2007) Role of host specificity in the speciation of Ascochyta pathogens of cool season food
1024 legumes. *European Journal of Plant Pathology* **119**, 119-126.
1025 Pordel A, Ravel S, Charriat F, *et al.* (2020) Tracing the origin and evolutionary history of Pyricularia
1026 oryzae infecting maize and barnyard grass. *Phytopathology®*, PHYTO-09.
1027 Pordel A, Ravel S, Charriat F, *et al.* (2021) Tracing the origin and evolutionary history of Pyricularia
1028 oryzae infecting maize and barnyard grass. *Phytopathology®* **111**, 128-136.
1029 Ramírez-Camejo LA, Eamvijarn A, Díaz-Valderrama JR, *et al.* (2021) Global analysis of Hemileia
1030 vastatrix populations shows clonal reproduction for the coffee leaf rust pathogen throughout
1031 most of its range. *Phytopathology®*.
1032 Restrepo S, Tabima JF, Mideros MF, Grünwald NJ, Matute DR (2014) Speciation in fungal and
1033 oomycete plant pathogens. *Annual review of phytopathology* **52**, 289-316.
1034 Rizzo DM (2005) Exotic species and fungi: Interactions with fungal, plant and animal communities. In:
1035 *The Fungal Community: Its Organization and Role in the Ecosystem*, pp. 857-880. CRC Press,
1036 Boca Raton.
1037 Rogério F, Van Oosterhout C, Ciampi-Guillardi M, *et al.* (2021) Means, motive, and opportunity for
1038 biological invasions: genetic introgression in a fungal pathogen. *bioRxiv*.
1039 Salcedo A, Rutter W, Wang S, *et al.* (2017) Variation in the AvrSr35 gene determines Sr35 resistance
1040 against wheat stem rust race Ug99. *Science* **358**, 1604-1606.
1041 Saleh D, Milazzo J, Adreit H, Fournier E, Tharreau D (2014) South-East Asia is the center of origin,
1042 diversity and dispersion of the rice blast fungus, Magnaporthe oryzae. *New Phytologist* **201**,
1043 1440-1456.
1044 Saleh D, Milazzo J, Adreit H, Tharreau D, Fournier E (2012a) Asexual reproduction induces a rapid and
1045 permanent loss of sexual reproduction capacity in the rice fungal pathogen Magnaporthe
1046 oryzae: results of in vitro experimental evolution assays. *BMC evolutionary biology* **12**, 1-16.
1047 Saleh D, Xu P, Shen Y, *et al.* (2012b) Sex at the origin: an Asian population of the rice blast fungus
1048 Magnaporthe oryzae reproduces sexually. *Molecular Ecology* **21**, 1330-1344.
1049 Schulze-Lefert P, Panstruga R (2011a) A molecular evolutionary concept connecting nonhost
1050 resistance, pathogen host range, and pathogen speciation. *Trends in Plant Science* **16**, 117-
1051 125.
1052 Schulze-Lefert P, Panstruga R (2011b) A molecular evolutionary concept connecting nonhost
1053 resistance, pathogen host range, and pathogen speciation. *Trends in plant science* **16**, 117-
1054 125.
1055 Servedio MR, Doorn GSV, Kopp M, Frame AM, Nosil P (2011) Magic traits in speciation: 'magic' but
1056 not rare? *Trends in Ecology & Evolution* **26**, 389-397.

1057 Short DPG, Gurung S, Gladieux P, *et al.* (2015) Globally invading populations of the fungal plant
1058 pathogen V erticillium dahliae are dominated by multiple divergent lineages. *Environmental*
1059 *Microbiology* **17**, 2824-2840.
1060 Short DPG, Gurung S, Hu X, Inderbitzin P, Subbarao KV (2014) Maintenance of sex-related genes and
1061 the co-occurrence of both mating types in Verticillium dahliae. *Plos One* **9**, e112145.
1062 Silva DN, Várzea V, Paulo OS, Batista D (2018) Population genomic footprints of host adaptation,
1063 introgression and recombination in coffee leaf rust. *Molecular Plant Pathology* **19**, 1742-
1064 1753.
1065 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing
1066 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*
1067 **31**, 3210-3212.
1068 Simpson JT, Wong K, Jackman SD, *et al.* (2009) ABySS: a parallel assembler for short read sequence
1069 data. *Genome Research* **19**, 1117-1123.
1070 Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1071 phylogenies. *Bioinformatics (Oxford, England)* **30**, 1312-1313.
1072 Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that
1073 allows user-defined constraints. *Nucleic Acids Research* **33**, W465-W467.
1074 Stauber L, Badet T, Feurtey A, Prospero S, Croll D (2021) Emergence and diversification of a highly
1075 invasive chestnut pathogen lineage across southeastern Europe. *eLife* **10**, e56279.
1076 Strange RN, Scott PR (2005) Plant Disease: A Threat to Global Food Security. *Annual review of*
1077 *phytopathology* **43**, 83-116.
1078 Stukenbrock EH (2013) Evolution, selection and isolation: a genomic view of speciation in fungal plant
1079 pathogens. *New Phytologist* **199**, 895-907.
1080 Stukenbrock EH, Bataillon T, Dutheil JY, *et al.* (2011) The making of a new pathogen: insights from
1081 comparative population genomics of the domesticated wheat pathogen Mycosphaerella
1082 graminicola and its wild sister species. *Genome Research* **21**, 2157-2166.
1083 Sweigard JA, Carroll AM, Kang S, *et al.* (1995) Identification, Cloning, and Characterization of
1084 *PWL2*, a Gene for Host Species Specificity in the Rice Blast Fungus. *Plant Cell* **7**, 1221-
1085 1233.
1086 Takabayashi N, Tosa Y, Oh HS, Mayama S (2002) A Gene-for-Gene Relationship Underlying the
1087 Species-Specific Parasitism of Avena/Triticum Isolates of Magnaporthe grisea on Wheat
1088 Cultivars. *Phytopathology* **92**, 1182-1188.
1089 Tanaka M, Nakayashiki H, Tosa Y (2009) Population structure of Eleusine isolates of Pyricularia oryzae
1090 and its evolutionary implications. *Journal of General Plant Pathology* **75**, 173-180.
1091 Taylor JW, Hann-Soden C, Branco S, Sylvain I, Ellison CE (2015) Clonal reproduction in fungi.
1092 *Proceedings of the National Academy of Sciences* **112**, 8901-8908.
1093 Taylor JW, Jacobson D, Fisher M (1999) The Evolution of Asexual Fungi: Reproduction,
1094 Speciation and Classification. *Annual review of phytopathology* **37**, 197-246.
1095 Tharreau D, Fudal I, Andriantsimialona D, *et al.* (2009) World population structure and migration of
1096 the rice blast fungus, Magnaporthe oryzae. In: *Advances in genetics, genomics and control of*
1097 *rice blast disease*, pp. 209-215. Springer.
1098 Thrall PH, Oakeshott JG, Fitt G, *et al.* (2011) Evolution in agriculture: the application of evolutionary
1099 approaches to the management of biotic interactions in agro-ecosystems. *Evolutionary*
1100 *Applications* **4**, 200-215.
1101 Uhse S, Djamei A (2018) Effectors of plant-colonizing fungi and beyond. *Plos Pathogens* **14**,
1102 e1006992.
1103 Valent B, Farman M, Tosa Y, *et al.* (2019) Pyricularia graminis-tritici is not the correct species name
1104 for the wheat blast fungus: response to Ceresini et al.(MPP 20: 2). *Molecular Plant Pathology*
1105 **20**, 173.

Weldon WA, Knaus BJ, Grünwald NJ, *et al.* (2021) Transcriptome-derived amplicon sequencing markers elucidate the US Podosphaera macularis population structure across feral and commercial plantings of Humulus lupulus. *Phytopathology®* **111**, 194-203.
Yarwood CE (1970) Man-Made Plant Diseases. *Science* **168**, 218 - 220.
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.
Zeigler RS (1998) Recombination in *Magnaporthe Grisea*. *Annual review of phytopathology* **36**, 249-275.
Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L (2019) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics (Oxford, England)* **35**, 1786-1788.
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 codes PRJEB42377 and PRJEB46618.

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. 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 PRJEB46618 Whole genome sequencing of Pyricularia oryzae fungi from Benin and Burkina Faso
3. 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
4. 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 in isolates Pyricularia oryzae isolates from rice and other hosts
doi: 10.5281/zenodo.5500197

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

3. National Center for Biotechnology Information

ID PRJNA320483 Genome-based Molecular Diagnostics for the Blast Fungus

4. National Center for Biotechnology Information

ID PRJNA417903 Resequencing of a global collection of *Pyricularia oryzae* isolates

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 and analysis of genotyping technical replicates

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. Linkage disequilibrium as a function of physical distance.

Supplementary file 4. F_{ST} estimates between clusters within lineage 1.

Supplementary file 5. Analysis of mycelial growth rates.

Supplementary file 6. Mycelial growth rates.

Supplementary file 7. Analysis of sporulation rates.

Supplementary file 8. Sporulation rates.

Supplementary file 9. 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 10. Analysis of pathogenicity data.

Supplementary file 11. Sequenced isolates.

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. Size of core and accessory genomes for non-effector proteins (A) and putative effectors (B) estimated using a rarefaction approach with a pseudo-sample size of n=30 genomes. Lineage 4 was not included due to small sample size.

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.