PLOS PATHOGENS

RESEARCH ARTICLE

Maintenance of divergent lineages of the Rice Blast Fungus *Pyricularia oryzae* through niche separation, loss of sex and post-mating genetic incompatibilities

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]*, Pierre Gladieux[1]*

1 PHIM Plant Health Institute, Univ Montpellier, INRAE, CIRAD, Institut Agro, IRD, Montpellier, France, **2** CIRAD, UMR PHIM, Montpellier, France, **3** ANSES Plant Health Laboratory, Mycology Unit, Malzéville, France

* didier.tharreau@cirad.fr (DT); pierre.gladieux@inrae.fr (PG)

Abstract

Many species of fungal plant pathogens coexist as multiple lineages on the same host, but the factors underlying the origin and maintenance of population structure remain largely unknown. The rice blast fungus *Pyricularia oryzae* is a widespread model plant pathogen displaying population subdivision. However, most studies of natural variation in *P. oryzae* have been limited in genomic or geographic resolution, and host adaptation is the only factor that has been investigated extensively as a contributor to population subdivision. In an effort to complement previous studies, we analyzed genetic and phenotypic diversity in isolates of the rice blast fungus covering a broad geographical range. Using single-nucleotide polymorphism genotyping data for 886 isolates sampled from 152 sites in 51 countries, we showed that population subdivision of *P. oryzae* in one recombining and three clonal lineages with broad distributions persisted with deeper sampling. We also extended previous findings by showing further population subdivision of the recombining lineage into one international and three Asian clusters, and by providing evidence that the three clonal lineages of *P. oryzae* were found in areas with different prevailing environmental conditions, indicating niche separation. Pathogenicity tests and bioinformatic analyses using an extended set of isolates and rice varieties indicated that partial specialization to rice subgroups contributed to niche separation between lineages, and differences in repertoires of putative virulence effectors were consistent with differences in host range. Experimental crosses revealed that female sterility and early post-mating genetic incompatibilities acted as strong additional barriers to gene flow between clonal lineages. Our results demonstrate that the spread of a fungal pathogen across heterogeneous habitats and divergent populations of a crop species can lead to niche separation and reproductive isolation between distinct, widely distributed, lineages.

OPEN ACCESS

Citation: Thierry M, Charriat F, Milazzo J, Adreit H, Ravel S, Cros-Arteil S, et al. (2022) Maintenance of divergent lineages of the Rice Blast Fungus *Pyricularia oryzae* through niche separation, loss of sex and post-mating genetic incompatibilities. PLoS Pathog 18(7): e1010687. https://doi.org/10.1371/journal.ppat.1010687

Editor: Eva H. Stukenbrock, Max-Planck-Institut fur Evolutionsbiologie, GERMANY

Received: December 9, 2021

Accepted: June 17, 2022

Published: July 25, 2022

Data Availability Statement: All raw sequencing data are deposited in European Nucleotide Archive under accession codes PRJEB42377 (https://www.ebi.ac.uk/ena/browser/view/PRJEB42377) and PRJEB46618 (https://www.ebi.ac.uk/ena/browser/view/PRJEB46618). Published data sets with accession codes PRJNA417903, PRJEB41186 and PRJNA354675 were used. Single-nucleotide polymorphisms, genome assemblies, genome annotations, and gene predictions were deposited in Zenodo (doi: 10.5281/zenodo.4561581).

Funding: This work was funded by the Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) (to TK), the Agence Nationale de la Recherche (ANR) (Grant ANR-18-CE20-0016 to PG), the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) (to DT), the Agence Nationale de Sécurité Sanitaire de l'Alimentation (Anses) (to RI), the CGIAR Research Program on Rice (to DT). This work was also partially funded by Bayer Crop Science Singapore (to DT), whose role in the study was limited to the collection and selection of 211 isolates, and genotyping of all 886 isolates. Bayer Crop Science Singapore had no additional role in the design, collection and analysis of data, decision to publish, or preparation of the manuscript. Other funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

The spread of infectious plant diseases associated with introduced fungal pathogens can have devastating effects on agrosystems, ecosystems, and human livelihoods. In addition to the burden of emerging fungal infections, most major crops or domesticated trees are colonized by introduced pathogens that are widespread and have been present for extended periods of time. An understanding of the factors underlying the colonization success of fungal pathogens is essential to limit the risk of future pandemics. Here, we use genotyping, genome sequencing and phenotypic assays of a large collection of isolates of the rice blast fungus *Pyricularia oryzae* to investigate the factors underlying population differentiation and differences in life history strategies in this major plant pathogen. We found four lineages of *P. oryzae* that displayed significant differences in host range and repertoires of virulence genes, and thrived in areas with different types of environmental conditions. We also found that sexual compatibility between lineages is limited by the lack of fertile female structures and early, post-mating, reproductive barriers. Our study shows that niche separation, loss of sex and post-mating genetic incompatibilities can contribute to the maintenance of population structure in widely distributed fungal crop pathogens.

Introduction

Introduced fungal plant pathogens constitute a major threat to ecosystem health and agricultural production [1–3]. Research in plant pathology has revealed a seemingly inexhaustible stream of disease outbreaks or significant changes in the geographic location of diseases and host ranges of pathogens [4,5]. In addition to the burden of emerging fungal infections, most major crops or domesticated trees are colonized by introduced 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 [6,7]. 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 [8]. An understanding of the factors underlying the colonization success of widespread fungal pathogens is essential to protect world agriculture against future pandemics [9,10].

Introduced fungal plant pathogens display a range of population structures, reflecting their highly diverse life-history strategies and invasion histories [8,11,12]. Demographic events (e.g. population bottlenecks, secondary contact) and natural selection during spread across heterogeneous environments may be associated with intricate changes in population dynamics and life history traits, or may cause such changes. Shifting to a new host is a primary life history trait change in introduced 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 [13,14]. 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 [15,16] and the differentiation of pathogen populations adapted to different host populations, varieties or species [17–20]. Such changes in pathogen host range can result from immune-escape mutations. Typically, they occur in 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 [21–23]. Another frequent change in the life history traits of introduced fungal pathogens is the loss of sexual reproduction [24,25]. Introduced fungal pathogens may develop a clonal population structure over

their entire introduced range or over parts of that range ([8] and references therein). The immediate causes of the loss of sexual reproduction include a lack of compatible mating partners in the introduced range (i.e. a lack of compatible mating types) [11], hybridization and hybrid sterility [26,27], and a lack of compatible alternate hosts for sexual reproduction [28]. 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 [8], 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 [29]. These conditions may favor the asexual spread of new lineages carrying adaptive allelic combinations requiring shelter from recombination [30]. Finally, changes in life history traits can also result from variations in environmental conditions, including temperature in particular, during the course of pathogen spread [31,32]. Deciphering the complex interplay between changes in population structure and life history strategies during the spread of plant pathogenic fungi requires deep sampling and the integration of genetic, phenotypic, and environmental information throughout the native and introduced 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 (*Lolium*) and an emerging pathogen of maize (*Zea maydis*) [33–38]. 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 [39]. Sequence divergence between lineages was low, of the order of about 1% [39], 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 [40,41]. 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 hermaphroditic [42], but successful mating also 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 [42,43]. Populations with female-fertile strains, a coexistence of both mating types and signatures of recombination have been observed exclusively in Asia [42–46]. 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 [47–49]. 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 high frequency of fertile females, consistent with sexual reproduction [44]. All the other lineages (lineages 2, 3 and 4, present on several continents) have clonal population structures, highly biased frequencies of mating types and rare fertile females, suggesting that reproduction in these lineages is strictly asexual [44,49]. The recombining lineage has a nucleotide diversity that is four times greater than that of the clonal lineages [49]. In theory, gene flow remains possible

between the majority of the lineages, because there are rare fertile strains and some lineages have fixed opposite mating types. This raises questions about the nature of the factors underlying the emergence of different rice-infecting lineages in *P. oryzae*, and contributing to their maintenance in the face of potential gene flow. Previous studies have helped to cement the status of *P. oryzae* as a model for studies of the population biology of fungal pathogens. However, most efforts to understand the population structure of the pathogen have been unable to provide a large-scale overview of the distribution of rice-infecting lineages and of the underlying phenotypic differences, because the number of genetic markers was limited [43,44], the number of isolates was relatively small [47,49] and because host-pathogen interaction phenotypes were the only phenotypes that were scored [43,44,47]. The contribution of differences in the geographic and climatic distribution of rice blast lineages and genetic incompatibilities between them, in particular, remains unknown.

To further explore the factors underlying the origin and maintenance of population structure in rice-infecting *P. oryzae*, we analyzed genomic and phenotypic variation in isolates sampled across all rice-growing areas of the world. Our aims were to re-evaluate population subdivision and the reproductive mode using a large collection of samples, to further explore host specialization using a larger number of isolates and rice varieties, to examine differences in the geographic and climatic distribution of lineages, and to measure their interfertility. For this, 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 effector proteins using whole-genome resequencing data for a representative set of isolates.

Results

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

We characterized the global genetic structure of the rice blast pathogen, using an Illumina genotyping beadchip to score 5,657 single nucleotide polymorphisms (SNPs) distributed throughout the genome of 886 *P. oryzae* isolates collected from cultivated Asian rice in 152 sites in 51 countries (S1 Table). The final dataset included 3,686 SNPs after the removal of positions with missing data, which identify 264 distinct multilocus genotypes in the 886 isolates. The genotyping error rate was below 0.62% (S1 Table). Clustering analyses based on sparse nonnegative matrix factorization algorithms, as implemented in the sNMF program [50], revealed four clusters, hereafter referred to as "lineages" (Fig 1C and S1 Data and S2 Data). The model with $K = 4$ clusters was identified as the best model on the basis of the cross-entropy criterion. Models with $K>4$ induced only a small decrease in cross-entropy, suggesting that $K = 4$ captured the deepest subdivisions in the dataset (S1 Fig). The neighbor phylogenetic network inferred with SPLITSTREE also supported the subdivision into four widely distributed lineages, with long branches separating three peripheral lineages branching off within a central lineage (Fig 1A and 1B and S1 Data). A comparison with previous findings revealed that the central lineage in the Neighbor-net network corresponds to the combination of recombining lineage 1 with lineages 5 and 6, represented by two and one individuals, respectively, in a previous study [49] (S2 Fig). This central lineage is referred to hereafter as lineage 1. The three peripheral lineages in the Neighbor-net network corresponded to the previously described lineages 2, 3 and 4 [49]. Lineages 2 and 3 are similar to lineages B and C, respectively, identified with microsatellite data in a previous study [44] (S2 Fig). Genetic differentiation between the four lineages was strong (Weir and Cockerham's $F_{ST}> 0.56$), indicating strong barriers to gene flow between the

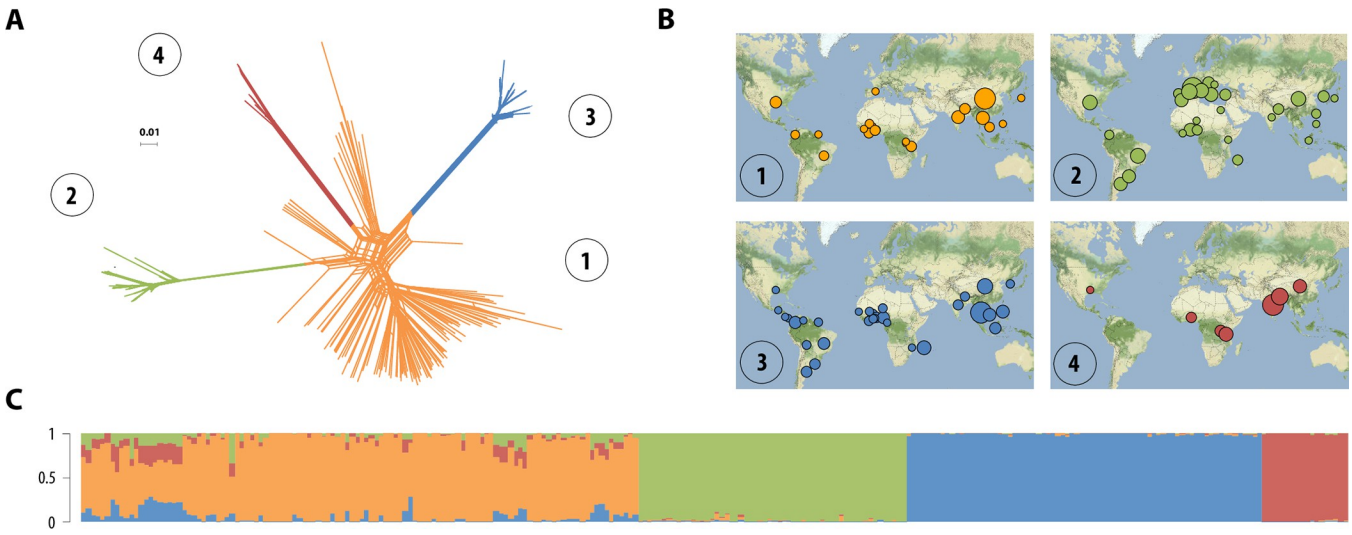

Fig 1. Rice-infecting *P. oryzae* populations are divided into four major lineages. Population subdivision was inferred from 264 distinct *P. oryzae* genotypes, representing 886 isolates, and the four lineages were represented in different colors. **(A)** Neighbor-net phylogenetic network estimated with SPLITSTREE; reticulations indicate phylogenetic conflicts caused by homoplasy. **(B)** Geographic distribution of the four lineages identified with SPLITSTREE (A) and sNMF (C), with disk area proportional to sample size. **(C)** Ancestry proportions in $K = 4$ clusters, as estimated with sNMF software; each multilocus genotype is represented by a vertical bar divided into four segments indicating individual ancestry coefficients in $K = 4$ clusters. Map tiles by Stamen Design, under CC BY 3.0, and data by OpenStreetMap, under ODbL (http://maps.stamen.com/#terrain/).

https://doi.org/10.1371/journal.ppat.1010687.g001

lineages (S1 Text). All genotypes from lineages 2, 3 and 4 had high individual ancestry coefficients in a single cluster ($q > 0.89$, q representing the proportion of an individual genome that originate from a given ancestral gene pool), whereas shared ancestry with the other three lineages was detected in lineage 1, with 32% of the genotypes having $q > 0.10$ in lineages 2, 3 and 4 (Fig 1C and S1 and S2 Data). Admixture may account for the lower F_{ST} observed in comparisons between lineage 1 and the other lineages.

We detected no signatures of recombination in lineages 2, 3 and 4 with three different tests (pairwise homoplasy index [51], maximum χ^2 [52] and neighbour similarity score [53]) and a null hypothesis of clonality (Table 1), confirming previous findings obtained with smaller datasets [44,47–49]. The lack of linkage disequilibrium decay with increasing physical distance in lineages 2, 3 and 4 (S3 Fig), and their low proportions of homoplastic mutations (Table 1), provided additional evidence for an absence of recombination in these lineages, contrasting with lineage 1. Consistent with patterns of linkage disequilibrium, the clonal fraction (1—

Table 1. Tests for recombination (null hypothesis: clonality), clonal fraction ((sample size N—number of genotypes G)/sample size N), distribution of mating types Mat1.1/Mat1.2, proportion of females fertile isolates, and proportion of homoplastic polymorphisms. Analyses were conducted for four lineages, and for clusters within lineage 1.

Lineage/ *Cluster*	Clonal fraction (%)	PHI (*p*-value)	Max χ^2 (*p*-value)	NSS (*p*-value)	Homoplastic sites (%)	Mat1.1/ Mat1.2	Proportion of female-fertile isolates (%)
1	73/189 (39)	<0.0001	<0.0001	<0.0001	1388/2214 (63)	53/47	62/144 (43)
Baoshan	2/13 (15)	<0.0001	<0.0001	<0.0001	NC	69/31	6/9 (67)
International	47/76 (62)	<0.0001	<0.0001	<0.0001	NC	55/45	2/51 (4)
Laos	11/32 (35)	<0.0001	<0.0001	<0.0001	NC	68/32	10/27 (37)
Yule	12/68 (18)	<0.0001	<0.0001	<0.0001	NC	40/60	44/58 (79)
2	166/222 (75)	0.25	0.32	0.72	1/391 (0)	97/3	0/20 (0)
3	344/418 (82)	0.18	0.35	0.88	0/408 (0)	3/97	2/30 (7)
4	39/57 (68)	0.22	0.30	0.51	0 (218)	97/3	0/15 (0)

https://doi.org/10.1371/journal.ppat.1010687.t001

number of observed multilocus genotypes over the number of isolates) was lower in lineage 1 (39%) than in lineages 2, 3 and 4 (75, 82 and 68%, respectively; Table 1).

Using whole-genome data for isolates from lineages of *P. oryzae* infecting other grasses and cereals, we showed that our set of Infinium SNPs could be used to identify isolates not belonging to the rice-infecting lineage, although it was unable to differentiate between the different host-specific lineages of *P. oryzae* (S2 Text). Clustering analysis identified two isolates from West Africa (BN0019 and BF0072, clonal group 18 [S1 Table]) corresponding to a lineage different from the rice lineage and possibly resulting from ancient admixture between other lineages (S2 Text). However, on inclusion of these divergent isolates in our dataset, we detected no fixed differences with respect to the rest of the isolates from lineage 1, and these divergent isolates did not form a separate group in clustering analyses. They are not, therefore, likely to have affected on our main conclusions on population subdivision and mode of reproduction.

Geographic structure and admixture within recombining lineage 1

Lineage 1 was the only lineage for which there was population genetic evidence of recombination (Table 1 and S6–S8 Data). 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: 10, Central and South America: 7, Africa: 22) (Fig 2A and 2B and S3 and S4 Data). Clustering analysis on this single lineage with sNMF detected four clusters with different geographic distributions (Fig 2D and S4 Data). Estimates of differentiation were lower between the clusters within lineage 1 ($F_{ST} < 0.51$) than between the main lineages ($F_{ST} > 0.56$), consistent with a longer history of restricted gene flow between the main lineages and a more recent history of admixture between clusters within lineage 1 (S1 Text).

Two clusters within lineage 1 (referred to hereafter as *Baoshan* and *Yule*) consisted mostly of isolates sampled from two different sites in the Yunnan province of China (Baoshan and Yule, respectively, located about 370 km apart). The *Baoshan* cluster also included one isolate from Nepal and one isolate from Hunan; the *Yule* cluster also included two isolates from Vietnam. The third cluster consisted mostly of isolates from Laos (referred to hereafter as *Laos*), but also included one isolate from Rwanda, three isolates from Yunnan, two isolates from Vietnam, and one isolate from French Guyana. This latter isolate is the widely used laboratory strain GUY11, collected from fields cultivated by the Hmong diaspora. The fourth cluster brought together 95% of the isolates from lineage 1 collected outside Asia (referred to hereafter as *International*). In Asia, the *International* cluster was found mostly in the Yunnan province of China, India and Nepal (Fig 2B and S1 Table). The *Yule*, *Laos* and *International* clusters coexisted in the same field (Yule village), the same year (2009). PHI tests for recombination rejected the null hypothesis of clonality for all four clusters (Table 1 and S6–S8 Data). Shared ancestry was widespread in lineage 1, with most of the isolates (78%) displaying individual ancestry coefficients (q) greater than 0.10 for two or more clusters (Fig 2D and S4 Data). Only five genotypes were detected in multiple countries (genotype ID [number of countries]: 2 [6], 18 [2], 58 [2], 98 [3] and 254 [3], corresponding to 41 isolates in total; Fig 2E and S5 Data). All these clonal genotypes belonged to the *International* cluster, which included 76 isolates in total. The clonal fraction (i.e. 1- number of genotypes over the sample size) was larger in the *International* cluster (62%) than in *Baoshan*, *Laos* and *Yule* clusters (15, 35 and 18%, respectively; Table 1).

Reproductive barriers caused by female sterility and postmating genetic incompatibilities

To elucidate how the clonal lineages have emerged from the more ancient recombining populations in lineage 1, we determined the capacity of the different lineages to engage in sexual

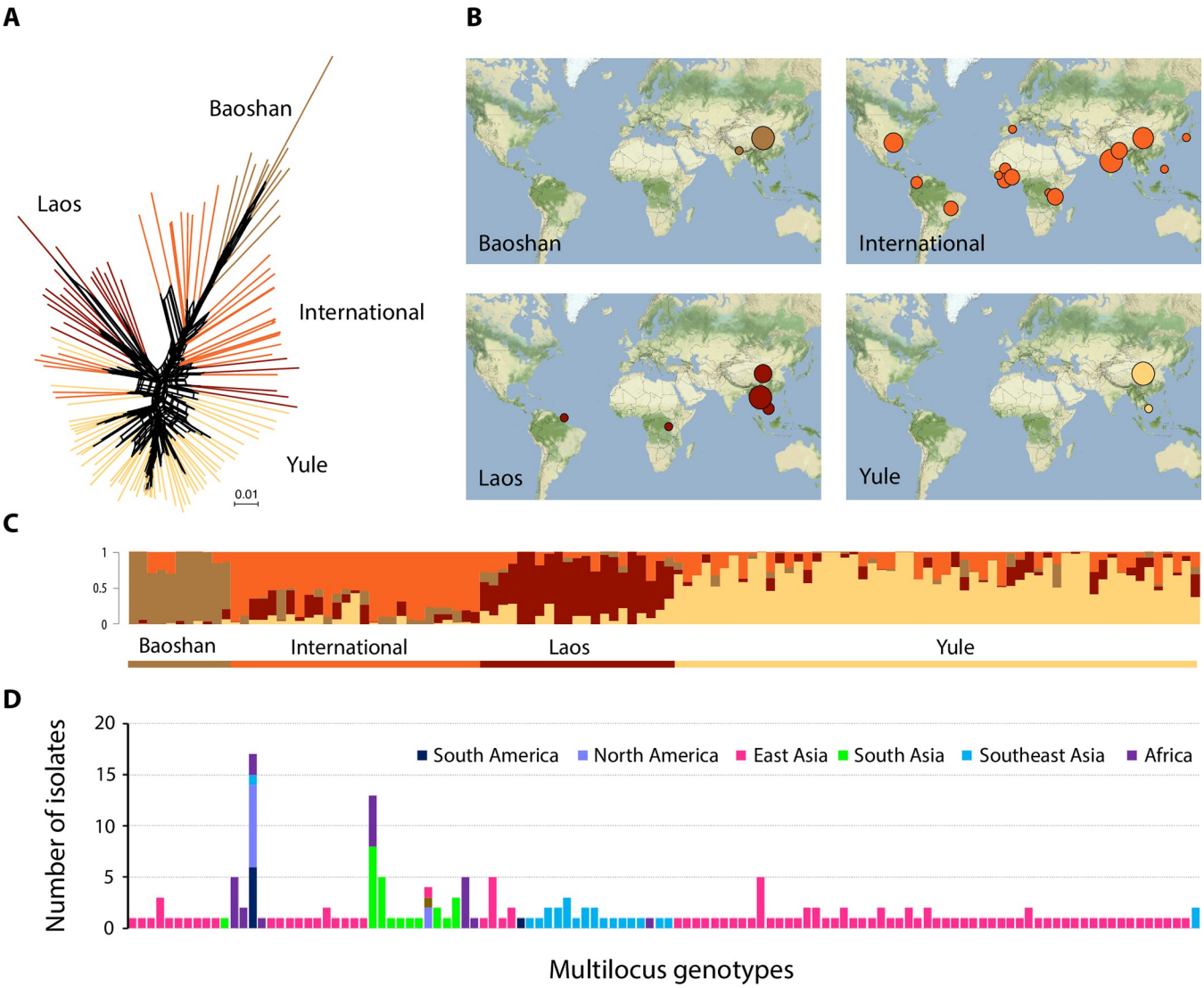

Fig 2. Population subdivision in lineage 1. (A) Neighbor-net phylogenetic network estimated with SPLITSTREE; reticulations indicate phylogenetic conflicts caused by homoplasy. **(B)** Geographic distribution of the four clusters identified with sNMF (C), with disk area proportional to number of isolates. **(C)** Ancestry proportions in four clusters, as estimated with sNMF; each multilocus genotype is represented by a vertical bar divided into four segments, indicating individual ancestry coefficients in $K = 4$ clusters. **(D)** Number of isolates and their geographic origin for each multilocus genotype of lineage 1. Panels C and D share *x*-axis (each vertical bar represents a different multilocus genotype). Map tiles by Stamen Design, under CC BY 3.0, and data by OpenStreetMap, under ODbL (http://maps.stamen.com/#terrain/).

https://doi.org/10.1371/journal.ppat.1010687.g002

reproduction. Using *in vitro* crosses with *Mat1.1* and *Mat1.2* tester strains, we analyzed the distribution of mating types and determined the ability to produce female structures (i.e. female fertility). We found that lineages 2, 3 and 4 were composed almost exclusively of a single mating type: 97% of lineage 2 and 4 isolates tested carried the *Mat1.1* allele, while 97% of lineage 3 isolates carried the *Mat1.2* allele. Only a small proportion (0–7%) of isolates from these lineages showed female fertility (Table 1). The mating type ratio was more balanced in lineage 1 (52% of *Mat1.1*; Table 1), with *Mat1.1*/*Mat1.2* ratios ranging from 40/60 in the *Yule* cluster to 69/31 in the *Baoshan* cluster. Female fertility rates were also high in most of the clusters in lineage 1 (*Yule*: 79%; *Baoshan*: 67%; *Laos*: 37%; Table 1). Only in the *International* cluster female fertility was as low as those in the clonal lineages, with 4% fertile females (Table 1).

These observations suggest that despite the generally low rate of female fertility, sexual reproduction between most lineages is possible, except between lineages 2 and 4, that have the same highly biased mating type ratio.

We further assessed the likelihood of sexual reproduction within and between lineages, by evaluating the formation of sexual structures (perithecia), the production of asci (i.e. meiotic octads within the perithecia) and the germination of ascospores (i.e. meiospores in meiotic octads) in crosses between randomly selected isolates of opposite mating types from all four lineages (Fig 3). This experiment revealed a marked heterogeneity in the rate of perithecium formation across lineages. Clusters in lineage 1 had the highest rates of perithecium formation, with isolates in the *Yule* cluster, in particular, forming perithecia in 93% of intra-lineage crosses, and in more than 46% of inter-lineage crosses (Fig 3A and S9 Data). Due to their highly biased mating type ratios, isolates from lineages 2, 3 and 4 could only be crossed with isolates from other lineages. The proportion of these inter-lineage crosses producing perithecia was highly variable and ranged, depending on the lineages involved, from 0% to 83% (Fig 3A and S9 Data). In inter-lineage crosses involving the *International* cluster of lineage 1, the rate of perithecium formation was similar as in inter-lineage crosses involving clonal lineages 2, 3 and 4. None of the intra-lineage crosses attempted between isolates from the *International* cluster led to perithecium formation (Fig 3A and S9 Data).

Perithecium dissection for a subset of crosses involving some of the most fertile isolates revealed that most inter-lineage crosses produced perithecia that did not contain asci or contained asci with low ascospore germination rates (Fig 3B and S10 Data). While 100% of intra-lineage crosses between isolates of the *Yule* cluster produced numerous germinating ascospores, only 33%, 56% and 7% of *Yule* x lineage 2, *Yule* x lineage 3 and *Yule* x lineage 4 crosses produced germinating ascospores, respectively.

Together, our crossing experiments indicate that the three clonal lineages and the *International* cluster of lineage 1 are isolated from each other by strong pre- and early-postmating barriers, including breeding system isolation (differences in mating type and female sterility), and a combination of gametic incompatibility, zygotic mortality or hybrid non-viability. However, three of the clusters in lineage 1 had biological features consistent with an ability to reproduce sexually, suggesting possible sexual compatibility with clonal lineages 2, 3 and 4, and the *International* cluster.

Specialization to temperature conditions and rice subgroups

Given the very broad distribution of *P. oryzae* and the strong environmental heterogeneity it encounters in terms of the nature of its hosts and the climate in which it thrives, we evaluated the variation of fitness over different types of rice host genotypes and temperature conditions. We measured growth rate and sporulation of 41 representative isolates cultured at different temperatures to test the hypothesis of adaptation to temperature (lineage 1 [yule cluster]: 11; lineage 2: 10; lineage3: 10; lineage 4: 10). For all lineages, mycelial growth rate increased with incubation temperature. 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 S6 Text; data are reported in S11 Data). 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-value<0.001; 15˚C: $F(1, 419) = 33161$, p-value<0.001; 20˚C: $F(1, 542) = 40335$, p-value<0.001; 25˚C: $F(1, 413) = 30156$, p-value<0.001; 30˚C: $F(1, 870) = 52681$, p-value<0.001]. Comparing least-squares means resulting from linear mixed-effects models at each temperature revealed a significantly lower growth

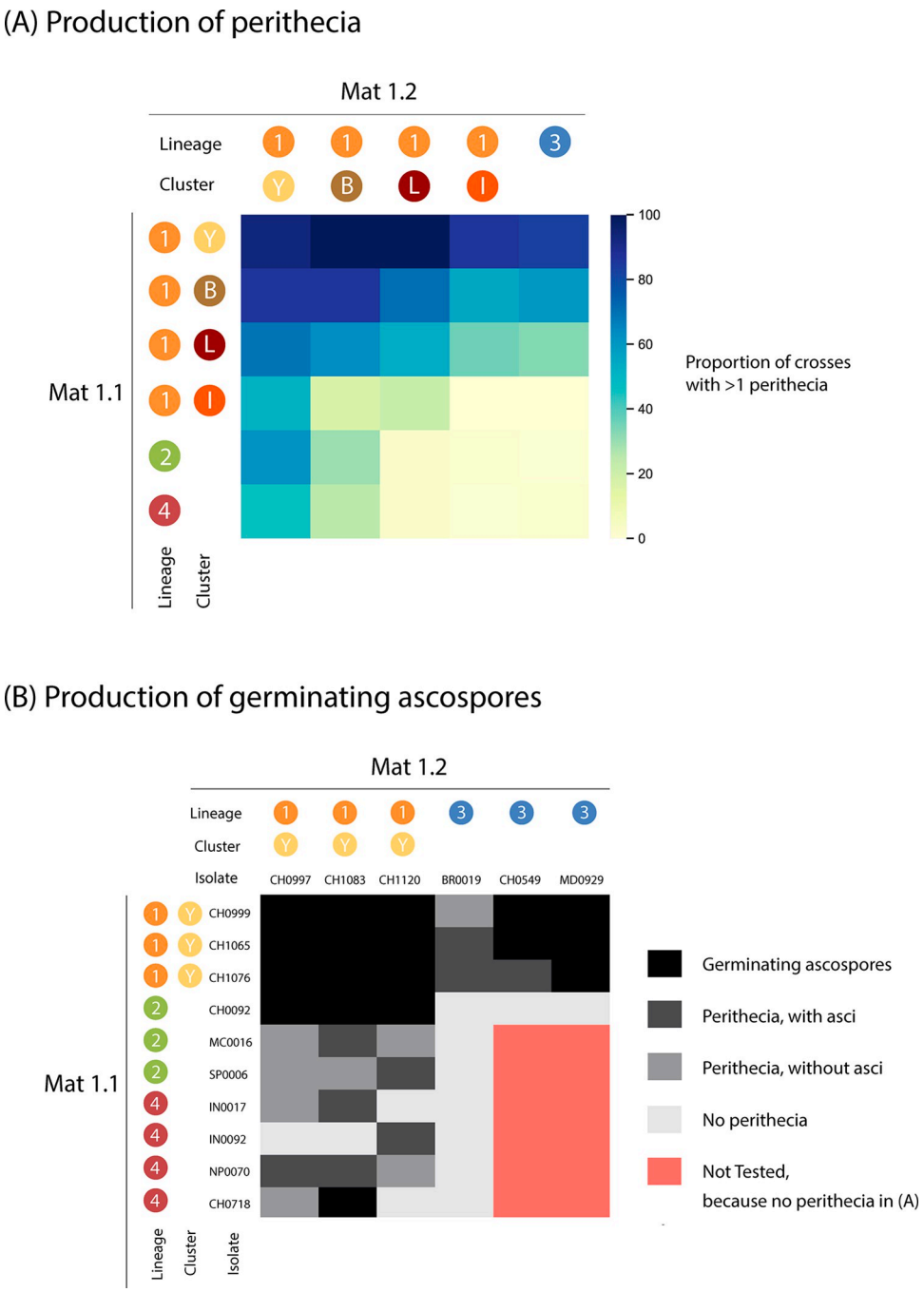

Fig 3. Success of crosses between lineages 2–4 and clusters within lineage 1 with (A) proportion of crosses producing at least one perithecium, (B) scoring of ascus formation and ascospore germination for a subset of crosses.

https://doi.org/10.1371/journal.ppat.1010687.g003

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 S7 Text; data are reported in S12 Data). For all lineages, sporulation (weighted by mycelium colony size) increased with

temperature from 15 to 25˚C and dropped at 30˚C. Significant lineage effects were observed at 10˚C, 15˚C and 30˚C (Kruskal-Wallis tests; 10˚C: H(3) = 9.27, p = 0.026; 15˚C: H(3) = 17.5, p-value<0.001; 30˚C: H(3) = 9.60, p = 0.022). 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. Together, measurements of mycelial growth and sporulation at different temperatures revealed differences in performance between lineages, but no clear pattern of temperature specialization.

We assessed the importance of adaptation to the host, by challenging 45 rice varieties representative of the five main genetic groups of Asian rice (i.e. the temperate japonica, tropical japonica, aus, aromatic and indica subgroups of *Oryza sativa*) with 70 isolates representative of the four lineages of *P. oryzae* and the four clusters within lineage 1 (S2 Table and S13 Data). Interaction phenotypes were assessed qualitatively by scoring resistance (from full to partial resistance: scores 1 to 3) or disease symptoms (from weak to full susceptibility: scores 4 to 6), which is standard in rice blast phenotyping and documented by images [54–57]. Resistance/susceptibility scores were analyzed by fitting a proportional-odds model and performing an analysis of variance that revealed significant differences between groups of isolates ($\chi^2(6)$ = 100, p-value<0.001), between rice genetic groups ($\chi^2(4)$ = 161, p-value<0.001), and a significant interaction between these two variables ($\chi^2(24)$ = 97, p-value<0.001) (S8 Text). The finding of a significant interaction between groups of isolates (lineages or clusters) and rice types indicates that the effect of the group of isolates on the proportion of compatible interactions differed between rice types, suggesting adaptation to the host. This is also supported by the specific behaviour of certain lineages. Isolates from lineage 2 had much lower symptom scores than the other lineages on all rice types except temperate japonica, and the isolates of the *Yule* cluster were particularly virulent on indica varieties (Fig 4A and S2 Table). In comparisons of rice genetic groups, significantly higher symptom scores were observed on temperate japonica rice than on the other types of rice, whereas the varieties of the aromatic genetic group were significantly more resistant to rice blast (Fig 4B). Together, these experiments therefore revealed significant differences in host range between lineages. However, this specialization to the host was not strict, because there was an overlap between host ranges (Fig 4C–4G).

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

Effectors, defined as secreted proteins that modulate host responses, are the most prominent class of fungal virulence factors. They act by manipulating host cellular pathways and are thus key determinants of the host range and fitness on compatible hosts [58,59]. We therefore determined the differences in effector repertoires between *P. oryzae* lineages in terms of presence/absence and nucleotide polymorphisms, and analyzed whether this was a particularly dynamic fraction of the gene space. To this end, we used whole-genome sequencing data for 123 isolates, including 29 isolates sequenced in this study and 94 publicly available genomes [39,47,60], 33 of which were also genotyped with our Infinium genotyping beadchip (S3 Table and S3 Text). Clustering analysis indicated that this dataset covered the four lineages of *P. oryzae*, and three of the clusters of lineage 1 (Laos, International and Baoshan; S3 Text and S3 Table). Assembly lengths for the 123 sequenced isolates ranged from 36.7 to 39.6 Mb, with a mean value of 38.1 Mb (S3 Table). The number of assembled contigs longer than 500 bp ranged from 1010 to 2438, and the longest contig was 1.1 Mb long (S3 Table).

Gene prediction identified 11,684 to 12,745 genes per genome, including 1,813 to 2,070 genes predicted to encode effector proteins. We found a significant effect of the lineage of origin on both the mean number of putative effector genes (ANOVA, F = 5.54, p = 0.0014), and

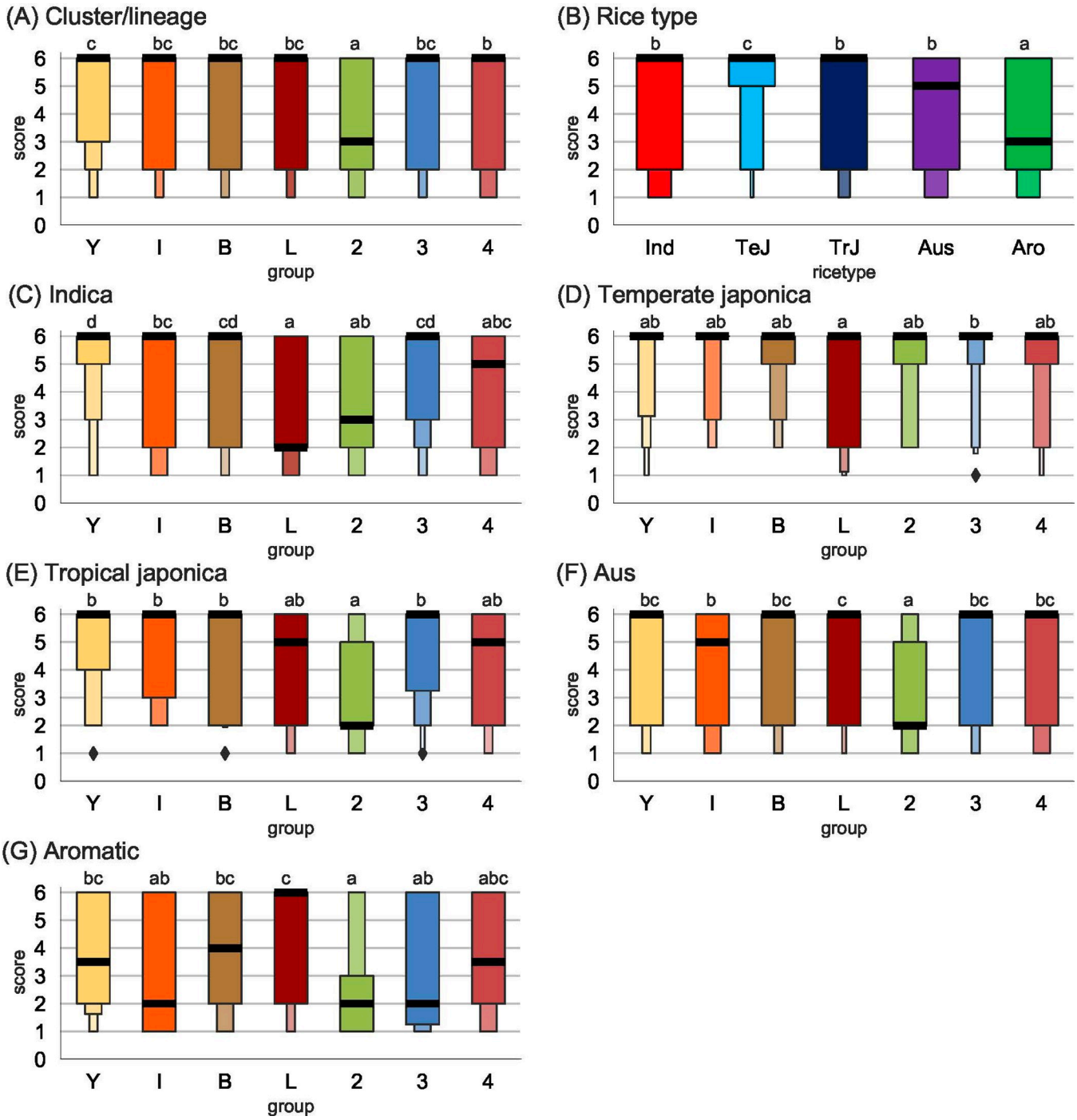

Fig 4. Compatibility between 70 *P. oryzae* isolates and 45 rice varieties, representing five types of rice. Compatibility was measured as symptom scores [54,55] estimated using pathogenicity tests in controlled conditions. A: Symptom scores as a function of the lineage of origin of isolates, or cluster 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. Shared small capitals indicate non-significant differences (pairwise comparisons after computing least-square means, with Tukey adjustment). All interactions between rice varieties and *P. oryzae* isolates were assessed in three independent experiments, and the median of the three symptom scores was used in calculations. Boxen plots were drawn using function boxenplot() with PYTHON package SEABORN 0.11.1. Starting with the median as centerline, each successive level (i.e., each successive box) outward contains half of the remaining data. Number of boxes was controlled using option k_depth = 'trustworthy'. Black horizontal lines represent the median.

https://doi.org/10.1371/journal.ppat.1010687.g004

the mean number of non-effector genes (ANOVA, F = 7.95, p-value<0.001). Multiple comparisons revealed that mean numbers of putative effectors and non-effector genes were only significantly different between the genomes of lineage 2 and lineages 3 (Tukey's HSD test; p-value<0.001; Fig 5A and 5B and S14 Data).

To estimate the size of accessory and core genomes in the different lineages, we performed an orthology analysis that identified 14,573 groups of orthologous sequences (i.e. orthogroups; S17 Data), and we applied a rarefaction approach to the table of orthology relationships to estimate the size of accessory and core genomes in lineages while accounting for differences in sample size. For both putative effectors and the remaining of the gene space, we found that the number of core and accessory genes did not reach a plateau. Instead, they displayed an almost linear relationship with the number of genomes resampled. This indicates that the number of core genes was probably much smaller, and the number of accessory genes substantially higher than estimated from 123 genomes (S4 Fig). With pseudosamples of size n = 30 per lineage (i.e. excluding lineage 4, due to its small sample size), the gene content was highly variable within lineages with only 61–71% of all predicted effectors, and 68–73% of the remaining of the gene space being conserved in lineages 1–3 (S5 Fig). Despite extensive variation of the gene content within lineages, the clustering of isolates based on presence/absence variation for both putative effectors and the remaining of the gene space was highly similar to that based on 3,868 SNPs (Fig 5D and 5E and S15 and S16 Data), indicating that variation in gene content and SNP allelic variation reflected similar genealogical processes.

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

Host specialization can also involve sequence divergence for effector proteins. We therefore also scanned the corresponding genes for signatures of higher rates of evolution, i.e. an excess of non-synonymous nucleotide diversity (ratio of non-synonymous to synonymous nucleotide diversity $\pi_N/\pi_S > 1$), and signatures of divergent selection (i.e. high relative divergence, F_{ST}). We identified 185 orthogroups with $\pi_N/\pi_S > 1$ in at least one lineage, including 164 orthogroups with $\pi_N/\pi_S > 1$ in only one lineage (lineage 1: 131 orthogroups; lineage 2: 18; lineage 3: 10; lineage 4: 5), and twelve, seven and two orthogroups with $\pi_N/\pi_S > 1$ in two, three or four lineages, respectively (S4 Table). There was no significant difference in inter-lineage F_{ST} between effector and non-effector genes (Mann-Whitney U-tests, p-value>0.05; S1 Text). We identified 291 orthogroups in the top 10% of inter-lineage of F_{ST} in at least one pair of lineages, including 42 orthogroups with $\pi_N/\pi_S > 1$ in at least one lineage (S5 Table). None of the orthogroups with $\pi_N/\pi_S > 1$ or high inter-lineage F_{ST} corresponded to effectors previously characterized as being involved in *P. oryzae* virulence on rice or other *Poaceae* hosts [61].

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

Our tests of adaptation to host and temperature under controlled conditions showed that specialization was not strict, but it remained possible that fitness differences between lineages would be sufficient under natural conditions to induce separation in different ecological niches, and/or that our experimental conditions did not capture the full extent of the phenotypic differences. Here, we tested the hypothesis that lineages thrive in geographically and/or environmentally different conditions, which would provide indirect evidence for specialization

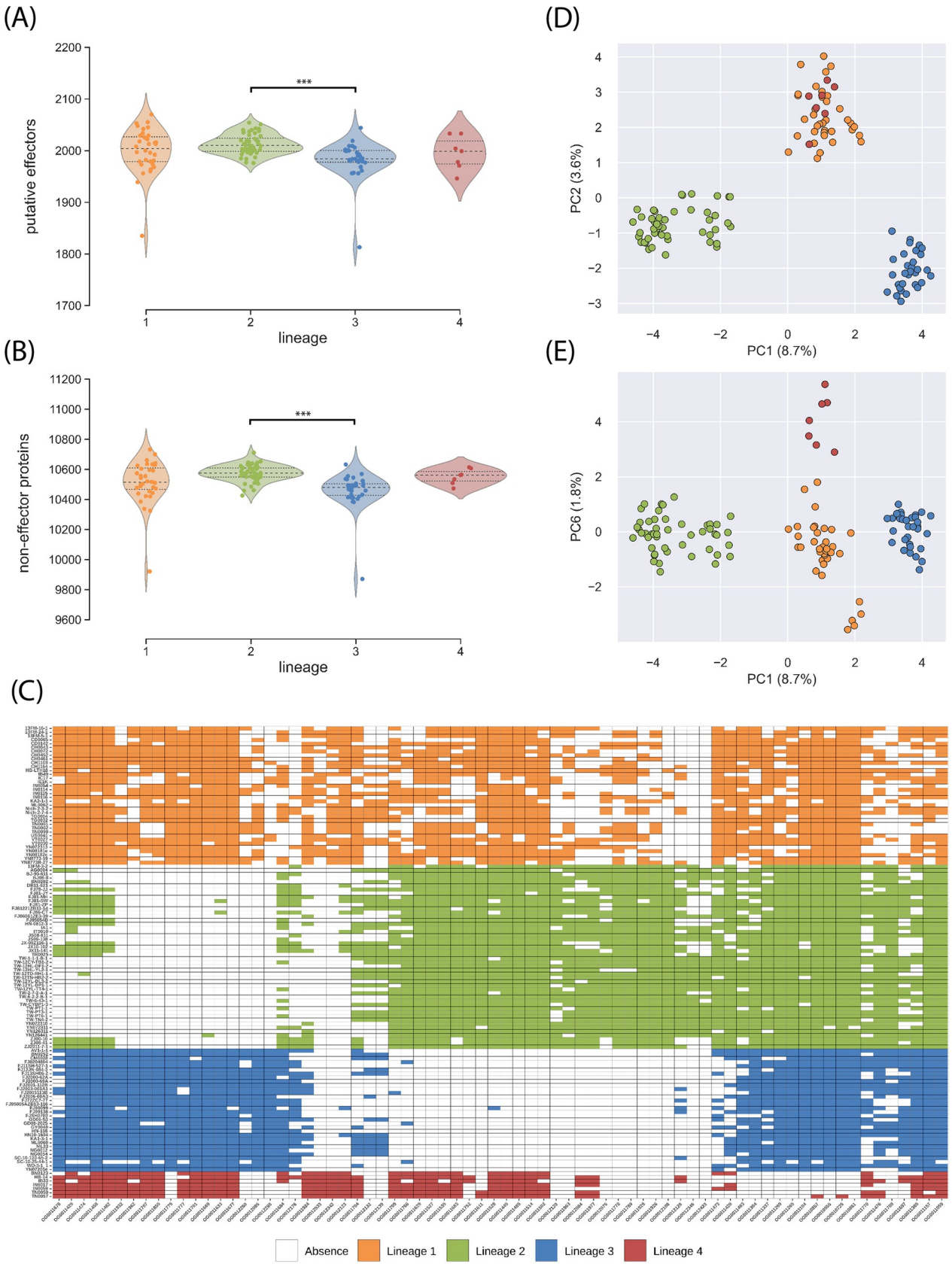
(A)
putative effectors
2200
2100
2000
1900
1800
1700
***
1
2
3
4
lineage
(B)
non-effector proteins
11200
11000
10800
10600
10400
10200
10000
9800
9600
***
1
2
3
4
lineage
(C)
Absence
Lineage 1
Lineage 2
Lineage 3
Lineage 4
(D)
PC2 (3.6%)
4
3
2
1
0
−1
−2
−3
−4
−2
0
2
4
PC1 (8.7%)
(E)
PC6 (1.8%)
4
2
0
−2
−4
−2
0
2
4
PC1 (8.7%)

Fig 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-value<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 component. (E) Presence and absence in lineage 1–4 of the 72 putative effectors with the highest contribution to principal components 1, 2 and 6, in a principal component analysis of presence/absence data.

https://doi.org/10.1371/journal.ppat.1010687.g005

to different habitats. We tested this hypothesis by collecting geographic and climatic data for all isolates, using the GPS position of the nearest city or the geographic center of their region of origin when the exact GPS position of isolates was not available (S1 Table). At a larger scale, clonal lineages 2 and 3 were both widespread, with lineage 2 found on all continents, and lineage 3 on all continents except Europe. Lineage 2 was the only lineage sampled in Europe (with the exception of a single isolate from lineage 1), whereas lineage 3 was more widespread in intertropical regions. Lineage 4 was mostly found in South Asia (India, Bangladesh, Nepal), and in the USA and Africa (Benin, Tanzania). The recombining lineage, lineage 1, was found on all continents, but mostly in Asia (79% of all isolates and 94% of all genotypes). At a smaller scale, two, or even three different lineages were sampled in the same year in the same village or city in 11 different countries from all continents. Such coexistence was observed for 24 of the 65 municipalities x year combinations for which multiple isolates were collected. A Kruskal-Wallis test showed that there was a statistically significant difference in altitudinal distribution between lineages ($H = 163.9$; p-value<0.0001; Fig 6). On average, lineage 1 was distributed at higher altitudes (mean: 908m), lineage 2 at lower altitudes (mean: 209m) and lineages 3 and 4 were intermediate (means: 521 and 669m, respectively). Mann-Whitney tests for multiple comparisons showed that the altitudinal distribution of lineages 1 and 2, lineages 1 and 3, and lineages 2 and 3 differed significantly at p-value < 0.05. Altitudinal distributions were relatively broad, with all lineages distributed at high altitudes (>1500m), and all lineages except lineage 4 distributed at low altitudes (<250m). Together, these analyses reveal that lineages had different, but partially overlapping latitudinal, longitudinal and altitudinal distributions, including some overlap with the distribution of the sexually reproducing lineage (lineage 1).

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 [62] 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 (Fig 7A and 7B and S18 Data and S4 Text). 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 [63] for all sampling locations, we obtained statistically significant results, in a permutation test on OMI values, for lineages 2, 3 and 4 (permutation test: lineage 1: OMI = 2.14, p-value = 0.620; lineage 2: OMI = 6.54, p-value<0.001; lineage 3: OMI = 2.08, p-value<0.001; lineage 4: OMI = 13.73, p-value = 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 (Figs 7C, 7D and S6 and S18 Data and S5 Text).

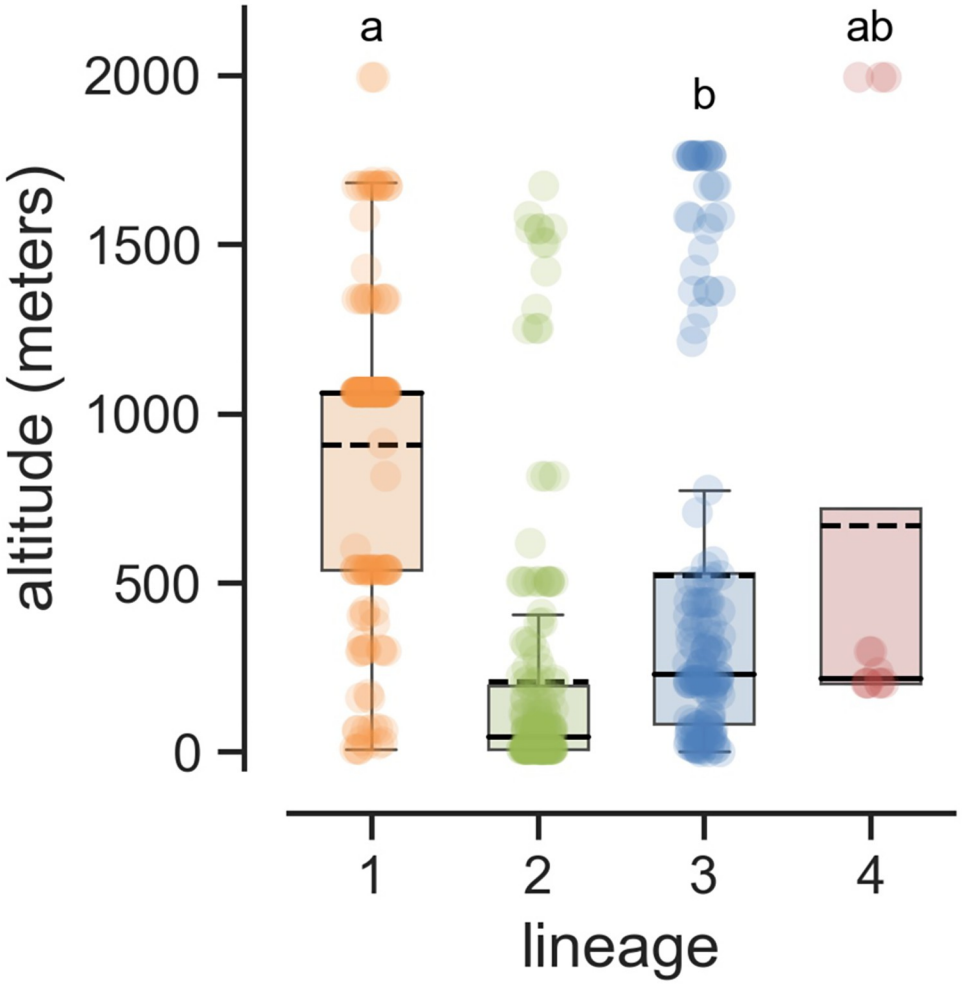

Fig 6. Altitudinal distributions of four lineages of *P. oryzae*. Dashed and solid black lines represent the mean and median, respectively. Shared superscripts indicate non-significant differences (Mann-Whitney post-hoc tests, p-value > 0.05, Holm-Bonferroni correction).

https://doi.org/10.1371/journal.ppat.1010687.g006

Discussion

This study reports on the worldwide distribution of genetic and phenotypic diversity in *P. oryzae*, and of the eco-evolutionary factors underlying the 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 geographically more restricted samplings (43, 44, 47–49). Our study substantially extends these previous 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 one international and three Asian clusters. This deep sampling of the natural genetic variation also reveals that clonal lineages of rice blast are more frequent in areas differing in terms of the prevailing environmental conditions. We also provide additional experimental evidence that lineages are host-specialized, and new evidence that breeding system isolation and early post-mating genetic incompatibilities act as strong barriers to gene flow between lineages.

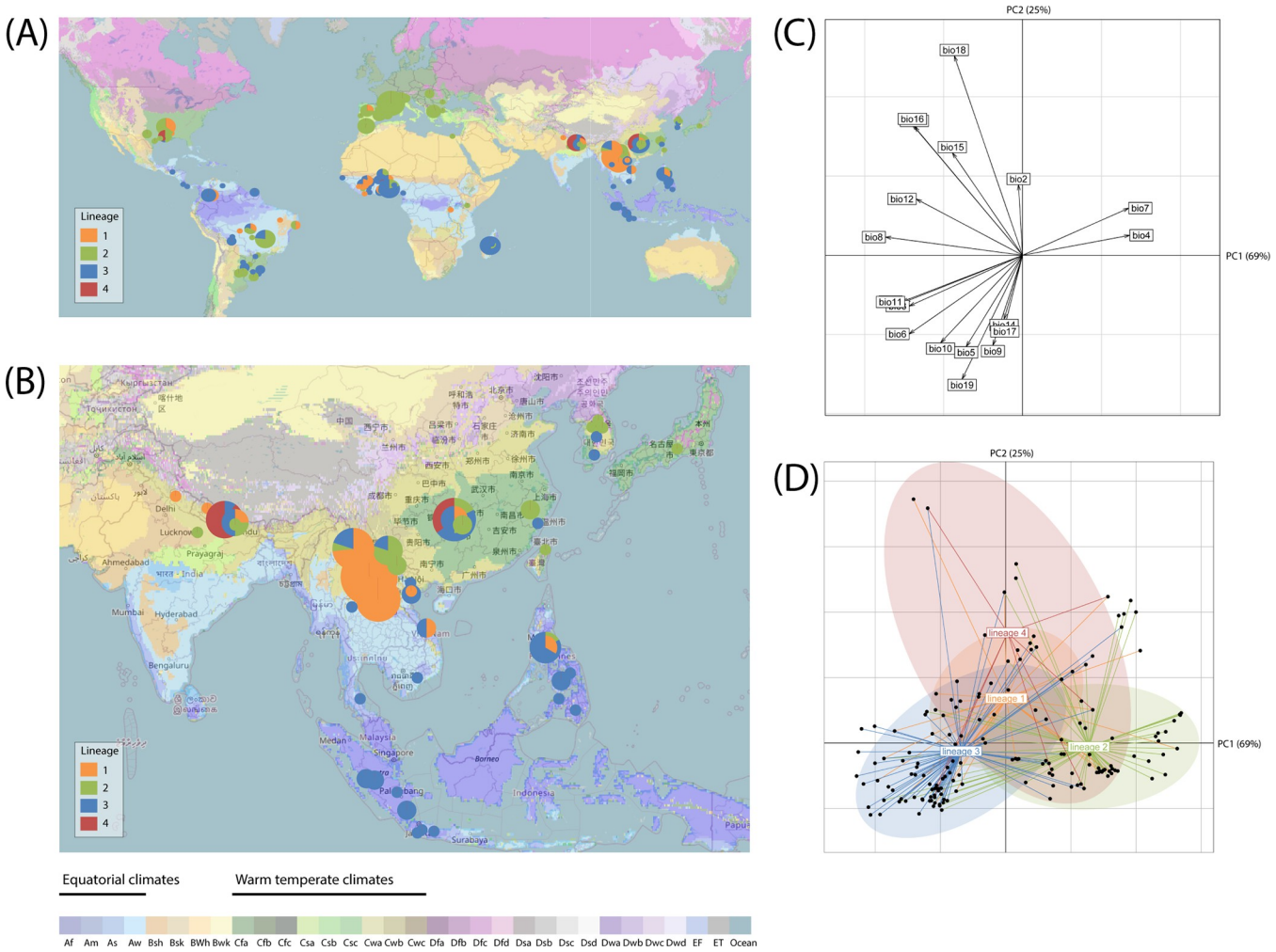

Fig 7. Geographic distribution of four lineages of *P. oryzae* and corresponding climatic data. (A, B) Pie charts representing the distribution of the four lineages in the world (A) and in South, East and Southeast Asia (B), keeping only isolates for which the sampling position was precisely known (i.e., 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 (62). (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 Materials and Methods section. PC1 and PC2 in panels C and D represent the first and second principal components, respectively, with percentage of variation between parentheses.

https://doi.org/10.1371/journal.ppat.1010687.g007

Niche separation and barriers to gene flow between clonal lineages

Many fungal plant pathogens consist of multiple divergent lineages coexisting on the same crop species, over large areas [64]. This raises questions about the maintenance of different lineages in the face of potential gene flow. 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. These barriers to gene flow may also 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.

To test for host specialization, we conducted pathogenicity experiments using a larger set of isolates [49,54], but also a larger set of rice varieties representing all five rice subgroups, in

order to assess whether certain features of adaptation to host may have been missed by previous studies based on more limited datasets. Our results were consistent with earlier findings in that the host range varied across lineages, but all host ranges overlapped, indicating that host specialization was not strict, or was not fully captured by our experiments. We therefore set out to examine the geographic and climatic distribution of lineages as additional barriers to gene flow, by relying on our dataset of unprecedented geographic density (152 sites in 51 countries, vs. 23 countries in ref. [47–49] or 55 sites in 15 countries in ref. [44]). We 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 did not predominate in one particular type of climate but was mostly sampled in Southeast Asia, lineage 4 predominated in regions with high levels of seasonal precipitation (typically the Indian subcontinent), and lineages 2 and 3 had circum-global distributions but predominated in temperate and tropical climates, respectively. The types of rice grown in these geo-climatic zones tend to be different, with in first approximation indica rice growing in the plains throughout the tropics and sub-tropics, tropical japonica in upland high-rainfall environments, temperate japonica in mid-latitude temperate zones, aromatic rice in the northwest of the Indian subcontinent, and aus in ecologically variable areas of South and Southeast Asia [65–70]. We cannot, therefore, rule out that host specialization contributes to the observed geoclimatic differentiation, but it remains unlikely that specialization is the main cause of this differentiation, because when observing at a finer scale the distribution ranges of rice subgroups largely overlap [65–70]. Disentangling host specialization from climate adaptation would require the combination of dense sampling and characterization of the genetic make-up of the varieties of origin of isolates.

Despite the finding of lineage separation in different geo-climatic regions, our experiments revealed no strong differences between lineages in terms of sporulation and mycelial growth on synthetic media at different temperatures. This suggests that, if adaptation to temperature occurs in this pathogen, it was not measured by our experiments, either because it involves traits other than sporulation and hyphal growth, or because the *in vitro* conditions were not suitable to demonstrate differences. These small differences in temperature optima may nonetheless, together with adaptation to the host, contribute to niche separation and thereby reduce gene flow between lineages, via pre-mating barriers, such as immigrant non-viability (i.e., lower rates of contact between potential mates due to the mortality of immigrants; [14,71]), and post-mating barriers, such as ecological hybrid non-viability (i.e., lower survival of ill-adapted hybrid offspring).

Link between disease spread and pathogen lineage divergence

There is a direct connection between disease emergence and speciation, because the emergence of new pathogens can be facilitated, caused by or associated with strong divergent selection in organisms mating within or on their hosts, and by major changes in demographics and geographic isolation [15,72–74]. There is also a connection between disease emergence and mode of reproduction, because disease emergence should be facilitated by multiple asexual cycles, corresponding to multiple cycles of selection for local adaptation without migration introducing locally deleterious immigrant alleles and recombination breaking down locally advantageous allelic combinations [15]. However, theoretically, in the long term, asexual organisms should accumulate a greater load of deleterious mutations and be less able to fix advantageous mutations, resulting in higher rates of extinction and lower rates of speciation [75,76]. The pattern of population structure revealed in *P. oryzae* seems to be intermediate between observations for introduced fungi with fully clonal population structures—such as the

pathogens causing wheat brown rust [28], boxwood blight [77] or verticillium wilt [78,79]—and introduced fungi with barriers maintaining lineages that are permeable and allow hybridization and introgression to occur, such as the pathogens causing *Dothistroma* needle blight [80], Dutch elm disease [11], and soybean anthracnose *Colletotrichum truncatum* [81]. The global structure of *P. oryzae* is more reminiscent of that of the pathogens causing chestnut blight [82,83], wheat yellow rust [84,85], and hop powdery mildew [86], for which clonal lineages coexist with recombining lineages, although the proximal causes for the loss of sexual reproduction may differ between these models and rice blast. Another marked difference from most of the abovementioned pathogens is that the clonal lineages of rice blast are relatively older, having diverged about a thousand years ago [47–49], which corresponds to several thousand asexual generations. Recombining lineage 1 may, at some point, become a source of adaptive introgression, reinvigorating the genomes of clonal lineages to reduce the mutational load, introducing beneficial variation, and leading to marked changes in population structure. Thus, although our work shows that the spread of a pathogen over heterogeneous habitats and over divergent populations of a crop species can drive the emergence of reproductive barriers, we cannot predict whether these lineages will continue to exist in their current form for the foreseeable future.

Loss of female sterility as the proximal cause of shifts to clonality

The loss of sexual reproduction, caused by demographic events or selection against sex, is a frequent change in the life history traits of introduced fungal pathogens [8,24,25,87]. However, the underlying changes and driving factors remain largely unexplored. In *P. oryzae*, the release of constraints maintaining sex in the short term, such as the need for sexual structures for winter survival or long-distance dispersal, is probably not responsible for the loss of sex, because the likelihood of overwintering on seeds and the rate of human-mediated transportation would not be expected to differ between clonal lineages 2, 3 and 4 and recombining lineage 1. The protection of locally advantageous allelic combinations is a more plausible explanation for the loss of sex in lineages 2, 3 and 4, because, as shown here, these lineages are found in, and adapted to, different hosts and environmental conditions. This leaves open the question of the maintenance of sex in lineage 1 open, but it is possible that the diversity of cultivated and wild rice diversity of rice is higher in the geographical range of this lineage [45]. It is difficult to draw firm conclusions about the evolutionary causes for selection against sexual reproduction, but the finding of an internationally distributed cluster in lineage 1 sheds light on the possible sequence of proximal events underlying the emergence of clonal lineages in this pathogen. The *International* cluster of lineage 1 displayed genetic signatures of recombination and sexual reproduction (in the form of an excess of homoplastic variants or relatively balanced mating types), but the signal of recombination detected here may be purely historical, because we also found that the clonal fraction and the frequency of sterile females were very high in this cluster. The loss of female fertility may, therefore, be an early and major cause of the shift towards asexual reproduction. Some genotypes assigned to the international cluster of lineage 1 may, in the future, come to resemble other clonal lineages in analyses of genealogical relationships, and form a long branch stemming from the central lineage 1, particularly for genotypes already displaying clonal propagation, such as genotypes 2, 18, 58, 98 and 254. This model could also apply to other fungal pathogens with similar life history strategies.

Effector repertoires differ between lineages

Effector repertoires play a key role in pathogen specialization, in fungal pathogens in general [21], and *P. oryzae* in particular [23, 58, 88–90]. Previous work suggested that clonal lineages

of *P. oryzae* possess smaller effector repertoires [48] and that the clonal lineage associated with japonica rice (lineage 2) possesses more effectors, and, in particular, Avr-effectors, than other clonal lineages [48,58]. Our analyzes on a larger set of putative effectors (ca. 2000 in our study, vs. 13 Avr-effectors in [58] and 178 known and candidate effectors in [48]) do not completely confirm this trend as we do not find a significantly larger effector complement in lineage 1 compared to other lineages. Lineage 2 associated with temperate japonica has significantly more putative effectors than lineage 3 associated with indica. However, the number of non-effector genes is also greater in lineage 2 than in lineage 3, and thus it remains possible that the larger effector complement of lineage 2 is a mere consequence of a larger number of genes. We also found that patterns of presence / absence variation mirrored patterns of population subdivision based on SNP allelic variation, and, thus, that the differential sorting of both nucleotide polymorphism and gene content across lineages reflects similar genealogical processes. Our multivariate analyses identified 72 effectors making the greatest contribution to the differentiation of the four lineages in terms of presence / absence. It is possible that the frequency differences for these 72 effectors were due to chance events during bottlenecks at the onset of lineage formation, or, alternatively, that their differential loss was important for the initial adaptation of lineages to new rice populations or subgroups. Interestingly, two of these 72 effectors, AvrRmg8 (= OG0011611) and PWL3 (= OG0011928), are host range determinants in wheat-infecting *P. oryzae* because they trigger immunity in wheat varieties carrying the resistance genes *Rmg8* [91] or *PWT3* [38, 92]. Nucleotide diversity at synonymous and non-synonymous sites and measures of relative nucleotide divergence (F_{ST}) also identified effectors with signatures of diversifying or divergent selection, possibly mediated by coevolutionary interactions with host molecules. In the future, it will be interesting to determine the molecular targets of these effectors and to decipher the relationship between their polymorphism and their mode of action.

Lineage 1 is a threat to global rice production

Our results also indicate that lineage 1 may pose a major threat to rice production. Most of the effectors with signatures of diversifying selection were identified in lineage 1, highlighting the greater ability of this lineage to fix advantageous mutations rapidly. Eleven of the 12 most multivirulent isolates (lesion type > 2 on more than 40 varieties tested) belonged to lineage 1, and the *Yule* cluster, in particular, was highly pathogenic on indica varieties. The propagation of such highly pathogenic genotypes, belonging to a lineage that has both mating types and fertile females, should be monitored closely. This monitoring of lineage 1 is all the more critical because this lineage has an intermediate geoclimatic distribution that overlaps largely with the distributions of the other three lineages, and some of the attempted crosses between clonal lineage 2, 3 and 4 isolates and recombining lineage 1 isolates produced viable progeny. This finding confirms the possibility of gene flow into this lineage, as previously demonstrated on the basis of admixture mapping (48, 49). It also raises the question of the threat posed by gene flow from the highly pathogenic lineage 1 to the other lineages.

Concluding remarks

Our study of genetic and phenotypic variation within and between clonal lineages, and within the recombinant lineage of *P. oryzae* suggests a scenario for the emergence of widespread clonal lineages. The loss of female fertility may be a potent driver of the emergence of asexually reproducing groups of clones. The reproductive isolation generated by the loss of sex and the accumulation of mutations due to the absence of sexual purging would facilitate the specialization of some of these clonal groups, leading to the extinction of the least efficient clonal groups,

and, finally, to the propagation of clonal lineages fixed for a single mating type. These results highlight that disease spread over heterogeneous habitats and over divergent populations of a crop species can lead to niche separation and barriers to gene flow between different lineages of a widely distributed fungal plant pathogen.

Materials and methods

Biological material

We chose 886 *P. oryzae* isolates collected on Asian rice between 1954 and 2014 as representative of the global genetic diversity of the fungus. Isolates were selected on the basis of microsatellite data, to maximize the number of multilocus genotypes represented [44], or based on geographic origin in the absence of genotypic data, to maximize the number of countries represented in the dataset.

Seventy-two and seventy isolates were selected for experimental measurements of reproductive success and pathogenicity, respectively (S1 Table). This subset of isolates included 10 isolates from each of the three clonal lineages (lineages 2, 3 and 4), and 40 isolates (42 for mating experiments) from the various clusters within lineage 1 [*Baoshan* (9 isolates), *International* [10], *Laos* [11], and *Yule* (12 for mating experiments, 10 for pathogenicity experiments)]. A subset of 41 isolates was selected for growth and sporulation experiments at different temperatures (S1 Table) including 10 isolates from each of the three clonal lineages (lineages 2, 3 and 4), and 11 isolates from the *Yule* cluster of lineage 1. Isolate CH0999 was included in the pathogenicity tests, although not genotyped with the Infinium array. We assigned this isolate to the *Yule* cluster using a combination of microsatellite data [44] and Infinium SNPs (not shown).

Forty-five varieties were chosen as representative of the five main genetic subgroups of Asian rice [93]: indica (CO39, Chau, Chiem chanh, DA11, De abril, IR8, JC120, Pappaku), aus (Arc 10177, Baran boro, Black gora, DA8, Dholi boro, Dular, FR13 A, JC148, Jhona 26, Jhona 149, Kalamkati, T1, Tchampa, Tepi boro), temperate japonica (Aichi asahi, Kaw luyoeng, Leung pratew, Maratelli, Nep hoa vang, Nipponbare, Sariceltik, Som Cau 70A), tropical japonica (Azucena, Binulawan, Canella de ferro, Dholi boro, Gogo lempuk, Gotak gatik, Trembese) and aromatic (Arc 10497, Basmati lamo, Dom zard, Firooz, JC1, Kaukkyisaw, N12).

Genotyping

Pyricularia oryzae isolates were genotyped at 5,657 genomic positions with an Illumina Infinium beadchip microarray carrying SNPs identified in whole genome sequencing data for 29 rice- and barley-infecting *P. oryzae* isolates [49]. The set of 29 whole genome sequences corresponded to four rice-infecting isolates characterized in ref. [94] and 25 isolates sequenced using paired-end Illumina reads in reference [49]. Read mapping and SNP calling were carried out as in reference [49]. The SNPs were selected to be polymorphic and biallelic, and not to distinguish between the different lineages identified previously. Of the ca. 20,000 SNPs initially identified in 29 genomes, we retained 7,000 SNPs that were biallelic and whose flanking sequences (50 bp) were unique and monomorphic. We then randomly subsampled 5,692 SNPs, and 5,657 could be successfully characterized in our set of 886 isolates. The final dataset included 3,686 biallelic SNPs without missing data.

Whole genome sequencing

We used a set of 123 whole-genome sequences to investigate differences in effector content between clusters and lineages, including 94 publicly available genomes and 29 new genomes that we selected to increase sample size for the International and Laos clusters of lineage 1, and

for lineage 4 (S3 Table). The 29 isolates were sequenced using Illumina HiSeq 3000. Isolates were grown on rice flour-agar medium for mycelium regeneration, then in liquid rice flour medium [95]. Genomic DNA extraction was carried out with >100 mg of fresh mycelium from liquid culture. Fresh mycelium dried on Miracloth paper was crushed in liquid nitrogen. Nucleic acids were subsequently extracted using an extraction buffer (2% CTAB– 1.4 M NaCl–0.1 M Tris-HCl pH 8–20 mM EDTA pH 8 added before use with a final concentration of 1% Na_2SO_3), then purified with a chloroform:isoamyl alcohol (24:1) treatment, precipitated overnight in isopropanol, and rinsed with 70% ethanol. The extracted nucleic acids were further treated with RNase A (0.2 mg/mL final concentration) and purified with another chloroform: isoamyl alcohol (24:1) treatment followed by overnight precipitation in ethanol. The concentration of extracted genomic DNA was assessed on Qubit using the dsDNA HS Assay Kit. The purity of the extracted DNA was checked by verifying that the 260/280 and 260/230 absorbance ratios measured with a NanoDrop spectrophotometer were between 1.8 and 2.0. Preparation of TruSeq nano library preparation and HiSeq3000 sequencing (150 nucleotide reads, and 500 bp insert size) were performed at GeT-PlaGe (INRAE, Toulouse, France).

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

For the 123 sequenced isolates included in the dataset, low-quality reads were removed with Cutadapt software [96]. Reads were assembled with ABySS 2.2.3 [97,98], using different K-mer sizes. For each isolate, we chose the assembled sequence with the highest N50 for further analyses. Genes were predicted with Braker 2.1.5 [99,100] using RNAseq data [60] as extrinsic evidence for model refinement. Genes were also predicted with Augustus 3.4.0 [101] (training set = *Magnaporthe grisea*) and gene models that did not overlap with the gene models identified with Braker were added to the GFF file generated with the latter. Repeated regions were masked with RepeatMasker 4.1.0 (http://www.repeatmasker.org/). The completeness of genome assemblies and gene predictions was assessed with BUSCO [102], and completeness in BUSCO genes was greater than 93.7% (S3 Table). Putative effector genes were identified as genes encoding proteins predicted to be secreted by at least two of three methods [SignalP 4.1 [103], TargetP [104] and Phobius [105]], with no predicted transmembrane domain based on TMHMM analysis [106], no predicted motif of retention in the endoplasmic reticulum based on PS-scan [107], and no CAZy annotation based on DBscan v7 [108]. Differences in the numbers of putative effectors and non-effectors between lineages (ANOVA and Tukey's HSD test) were assessed with the scipy 1.6.0 package in python 3.6. Homology relationships between predicted genes were established with OrthoFinder v2.4.0 [109]. Principal component analysis of presence/absence variation for putative effectors and non-effector proteins was performed with the R package prcomp. We used the get_pca_var() in prcomp to extract the results for variables (i.e. effectors) and to identify the effectors making the largest contributions to the principal components, defined as effectors with a contribution to loadings of PC1, PC2 and PC6 greater than 1%. For estimates of the numbers of core and accessory genes, we used a rarefaction approach to account for differences in sample size across lineages. For each pseudo-sample size, we estimated the size of the core and accessory genome in each lineage using a maximum of 2000 pseudosamples. Core and accessory genome sizes were compared using a Kruskal-Wallis test and posthoc Mann-Whitney U-tests as implemented in packages scikit_posthocs 0.6.6 and scipy 1.8.0 in Python 3.6. Sequences for each orthogroup were aligned and cleaned with TranslatorX [110], using default parameters. The ratio of non-synonymous to synonymous diversity π_N/π_S and Weir and Cockerham's F_{ST} [111] were calculated for each orthogroup with Egglib 3 (https://www.egglib.org), with minimal sample size of at least four

per lineage, excluding sites with more than 30% missing data, and excluding alignments shorter than 100 bp after removal of sites with excess missing data. Distributions of F_{ST} in effector and non-effector genes were compared using Mann-Whitney U-tests as implemented in package SCIPY 1.8.0 in PYTHON 3.6.

All raw sequencing data are deposited under accession codes PRJEB42377 and PRJEB46618. Published data sets with accession codes PRJNA417903, PRJEB41186 and PRJNA354675 were used. Single-nucleotide polymorphisms, genome assemblies, genome annotations, and gene predictions were deposited in Zenodo (doi: 10.5281/zenodo.4561581).

Population subdivision and recombination

The dataset used for population genetic analysis had no missing data. Clones were therefore identified as isolates with identical genotypes at 3,686 sites. Among the 886 *P. oryzae* isolates genotyped, we identified 264 different multilocus genotypes (i.e., "clones"), which were used for the analysis of population subdivision. We used the sNMF program to infer individual ancestry coefficients in *K* ancestral populations. This program is optimized for the analysis of large datasets and does not assume Hardy-Weinberg equilibrium. It is, therefore, more appropriate to deal with inbred or clonal lineages [50]. We used SPLITSTREE version 4 [112] to visualize relationships between genotypes in a phylogenetic network, with reticulations to represent the conflicting phylogenetic signals caused by homoplasy. Weir and Cockerham's F_{ST} [111] was calculated using the same method as whole genome sequencing data.

Analyses of recombination were conducted on a clone-corrected dataset (i.e. keeping only one representative of each the genotypes repeated multiple times). We tested the null hypothesis of clonality using the pairwise homoplasy index (PHI; [51]), maximum χ^2 (Max χ^2, [52]) and neighbor similarity score (NSS; [53]), as implemented in the PHIPACK program (https://www.maths.otago.ac.nz/~dbryant/software.html). Homoplastic sites have sequence identities that are not inherited from a common ancestor, but instead derived from independent events in different branches, such as recurrent mutations, sequencing errors or recombination. Homoplastic sites were identified by concatenating genotypes at all sites, inferring a genealogy with the GTR+GAMMA model in RAXML v.8.2.12 [113], and mapping mutations onto the genealogy with the 'Trace All Characters' function of MESQUITE under maximum parsimony [114]. In the resulting matrix of ancestral states for all nodes, the number of mutations occurring at each site was counted and sites displaying multiple mutations across the genealogy were considered to be homoplastic. We used POPLDDECAY [115] to measure linkage disequilibrium (r^2) as a function of distance between pairs of SNPs. POPLDDECAY was configured with a maximum distance between SNPs of 400 kbp, and a minor allele frequency of 0.005.

Experimental measurement of reproductive success and female fertility

Pyricularia oryzae is a heterothallic fungus with two mating types (*Mat1.1* and *Mat1.2*). Sexual reproduction between strains of opposite mating type can be observed in laboratory conditions, and results in the production of ascospores within female sexual structures called perithecia [116]. Crosses were carried out on a rice flour agar medium (20 g rice flour, 2.5 g yeast extract, 15 g agar and 1 L water, supplemented with 500 000 IU penicillin G after autoclaving for 20 min at 120˚C), as described by [116]. We assessed reproductive success by determining the production of perithecia produced after three weeks of culture at 20˚C under continuous light. Each cross was performed twice and we determined the mean number of perithecia across repeats. We further assessed the presence of asci and germinating ascospores in perithecia for a subset of crosses, by excising perithecia with a scalpel and counting the number of germinated filaments for each individual ascus after overnight incubation on water agar. The

subset of crosses included 10 Mat1.1 isolates (lineage 1: CH0999, CH1065, CH1076; lineage 2: CH0092, MC0016, SP0006; lineage 4: IN0017, IN0092, NP0070, CH0718) and 6 Mat1.2 isolates (lineage 1: CH0997; CH1083, CH1120; lineage 3: BR0019, CH0549, MD0929).

We measured female fertility for 210 isolates (listed in S1 Table), by monitoring perithecium production in crosses involving the tester strains CH0997 (Mat1.2) and CH0999 (Mat1.1). On synthetic media, perithecia are formed at the zone of contact between parental mycelia. Isolates forming perithecia are described as female-fertile. Perithecia can be formed by the two interacting partners or by one partner only. If the two partners are female-fertile, two parallel lines of perithecia are observed at the contact zone between mycelia. If only one partner is female-fertile, only one line of perithecia is observed, at the edge of the mycelium of the fertile isolate.

Pathogenicity tests

Compatibility between *P. oryzae* isolates and rice plants from the five main genetic subgroups of rice (indica, temperate japonica, tropical japonica, aus and aromatic) was measured in controlled conditions. We inoculated on 45 varieties (see section Biological Material), with 70 isolates. Inoculations were performed as described by [54]. Conidial suspensions (25 000 conidia. mL^{-1}) in 0.5% gelatin were sprayed onto three-week-old rice seedlings (> 6 plants/variety). The inoculated plants were incubated for 16 hours in the dark at 27˚C and 100% humidity and then for seven days with a day/night alternation (13 hours at 27˚C/11 hours at 21˚C), before the scoring of symptoms. Lesion type was rated from 1 to 6 and the symptom type was assessed visually on leaves. Each interaction was assessed in three independent experiments, which represented 9,450 scored interactions in total. Symptoms scores are ordinal-scale variables (i.e. variables that can be ranked but are not evenly spaced). We therefore analyzed the data with a proportional-odds model with the clm() function implemented in ORDINAL version 2018.8–25 in R, including rice type and pathogen lineage as main effects, and the rice type–pathogen lineage interaction. A comprehensive mixed model including all experimental replicates, with isolate as a random effect, was also tested, but did not display convergence. We therefore used the median of the three scores in calculations, as this is what most accurately represented the biological and experimental variation in the ability of an isolate to exploit a given host genotype. We used an analysis of deviance to evaluate whether effect and interaction terms were statistically significant. Pairwise comparisons of significant factors were performed after computing least-squares means with LSMEANS version 2.30–0 in R, with Tukey adjustment. The R commands and their outputs are provided in S8 Text.

Mycelial growth and sporulation at different temperatures

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

second and third experiments. At 30˚C, the number of spores was measured after 7 days only, versus 10 days at other temperatures, because by this stage, the mycelium had already reached the edges of the plates.

Mycelium growth was analyzed independently for each temperature with linear mixed-effects models. The variable used in statistical analysis was the square root of the mean mycelial diameter measured along two perpendicular axes on a Petri plate. Post-hoc pairwise comparisons were performed after computing least-squares means with LSMEANS version 2.30–0 in R, with Tukey adjustment. The R commands and their outputs are provided in S6 Text. Data are reported in S11 Data.

Sporulation data were analyzed with a Kruskal-Wallis test with RSTATIX version 0.7.0 in R. The variable used in statistical analysis was the median number of spores weighted by mycelium size, calculated across replicates. Post-hoc pairwise comparisons were performed with Dunn's non-parametric multiple comparison test. The R commands and their outputs are provided in S7 Text. Data are reported in S12 Data.

Statistical analyses of climatic data

The outlying mean index (OMI), or marginality, is used to study niche separation and niche breadth. The OMI approach evaluates whether a particular lineage is mostly associated with some climatic conditions or can be sampled with the same probability throughout the range of the species. The approach is based on a multivariate analysis (here a PCA) of climatic data associated with the locations included in the dataset, regardless of the identity of the lineages sampled or the number of isolates sampled. An index of marginality or specialization (the outlying mean index) is then calculated for each lineage represented in the dataset. This index measures the distance between the average habitat of a given lineage and the average habitat conditions of the area studied (corresponding to the distribution of a hypothetical species uniformly distributed under all conditions). Lineages are then placed in environmental conditions that maximize their OMIs, and permutation tests are used to test whether the distribution of a lineage differs from random. OMI gives the same weight to all sampling locations, whether rich or poor in individuals or lineages, and is particularly suitable in cases in which sampling is not homogeneous. Environmental values, consisting of 19 biome values (WorldClim bioclimatic variables [63]), were retrieved for all sampling locations. The 11 temperature variables included in the analysis were (bio1) annual mean temperature, (bio2) mean diurnal range, (bio3) isothermality [100*(bio2/bio7)], (bio4) temperature seasonality [standard deviation*100], (bio5) maximum temperature of the warmest month, (bio6) minimum temperature of the coldest month, (bio7) annual temperature range [bio5-bio6], (bio8) mean temperature of the wettest quarter, (bio9) mean temperature of the driest quarter, (bio10) mean temperature of the warmest quarter, (bio11) mean temperature of the coldest quarter. The eight precipitation variables included in the analysis were (bio12) annual precipitation, (bio13) precipitation of the wettest month, (bio14) precipitation of the driest month, (bio15) precipitation seasonality (coefficient of variation), (bio16) precipitation of the wettest quarter, (bio17) precipitation of the driest quarter, (bio18) precipitation of the warmest quarter, (bio19) precipitation of the coldest quarter. The 11 climatic variables were normalized before analysis. A contingency table was generated to associate the number of isolates from each lineage with each sampling location. Only isolates from our dataset with a precisely known geographic origin (region, city or GPS position) were included in the analysis. A random permutation test with 10,000 permutations was used to assess the statistical significance of marginality for each lineage. Altitudinal data were analyzed using a Kruskal-Wallis test and posthoc Mann-Whitney U-tests as implemented in packages SCIKIT_POSTHOCS 0.6.6 and SCIPY 1.8.0 in PYTHON 3.6.

Supporting information

S1 Data. Multilocus genotypes for neighbor-net phylogenetic inference, principal component analysis and recombination analyses. Used to generate Fig 1.
(TXT)

S2 Data. Ancestry proportions in K = 4 sNMF clusters. Used to generate Fig 1.
(TXT)

S3 Data. Multilocus genotypes for neighbor-net phylogenetic inference. Used to generate Fig 2.
(TXT)

S4 Data. Ancestry proportions in K = 4 sNMF clusters. Used to generate Fig 2.
(TXT)

S5 Data. Geographic distribution of clones (i.e. multilocus genotypes repeated multiple times). Used to generate Fig 2.
(TXT)

S6 Data. RAxML genealogy for homoplasy analysis, estimated from S1 Data. Used to generate Table 1.
(TXT)

S7 Data. Pairs of nodes included in homoplasy analysis, as determined from S6 Data. Used to generate Table 1.
(TXT)

S8 Data. Result of 'Trace all characters' analysis conducted with Mesquite. Used to generate Table 1.
(TXT)

S9 Data. Production of perithecia. Used to generate Fig 3.
(TXT)

S10 Data. Production of germinating ascospores. Used to generate Fig 3.
(TXT)

S11 Data. Mycelial growth rates. Used in S6 Text.
(TXT)

S12 Data. Sporulation data. Used in S7 Text.
(TXT)

S13 Data. Symptom scores for *P. oryzae* isolates inoculated onto rice varieties. Used to generate Fig 4 and in S8 Text.
(TXT)

S14 Data. Size of core and accessory genomes for putative effectors and non-effector proteins. Used to generate Fig 5.
(TXT)

S15 Data. Presence and absence of putative effectors in *P. oryzae* genomes. Used to generate Fig 5.
(TXT)

S16 Data. Presence and absence of non-effector genes in *P. oryzae* genomes. Used to generate Fig 5.
(TXT)

S17 Data. Orthology table, with one orthogroup per line and one isolate per column, and correspondence with gene names (i.e. MGG) in reference genome 70–15 (version MG8 – Ensembl fungi). Used to generate Fig 5.
(TXT)

S18 Data. GPS position of isolates. Used to generate Fig 7.
(TXT)

S1 Text. Analysis of inter-lineage and inter-cluster differentiation.
(DOCX)

S2 Text. Analysis of shared ancestry between isolates collected on rice and lineages infecting non-rice hosts.
(DOCX)

S3 Text. Assignment of sequenced isolates.
(DOCX)

S4 Text. R code used to prepare maps in Fig 7.
(TXT)

S5 Text. R code used for analyses based on the Outlying Mean Index, including Principal Component Analysis presented in Fig 7.
(TXT)

S6 Text. Analysis of mycelial growth rates.
(DOCX)

S7 Text. Analysis of sporulation data.
(DOCX)

S8 Text. Analysis of pathogenicity data.
(DOCX)

S1 Table. Isolates of *Pyricularia oryzae* and analysis of genotyping technical replicates.
(XLSX)

S2 Table. Matrix of compatibility between four lineages of *P. oryzae* isolates and five subgroups 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.
(XLSX)

S3 Table. Sequenced isolates.
(XLSX)

S4 Table. 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.
(TXT)

S5 Table. Genes in the top 10% of inter-lineage F_{ST}. Only orthogroups with sample size > = 4 were included in calculations.
(TXT)

S1 Fig. Cross-entropy (CE) as a function of the number of clusters K modeled in sNMF analyses of population subdivision.
(DOCX)

S2 Fig. Assignment of genotypes to clusters identified in this study, in Gladieux et al. 2018b (A) (49) and in Saleh et al. 2014 (44) (B and C).
(XLSX)

S3 Fig. Linkage disequilibrium as a function of physical distance.
(DOCX)

S4 Fig. 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.
(DOCX)

S5 Fig. 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.
(DOCX)

S6 Fig. 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.
(DOCX)

Acknowledgments

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

Author Contributions

Conceptualization: Maud Thierry, Thomas Kroj, Renaud Ioos, Elisabeth Fournier, Didier Tharreau, Pierre Gladieux.

Data curation: Maud Thierry, Florian Charriat, Joëlle Milazzo, Henri Adreit, Didier Tharreau, Pierre Gladieux.

Formal analysis: Maud Thierry, Florian Charriat, Sébastien Ravel, Pierre Gladieux.

Funding acquisition: Thomas Kroj, Renaud Ioos, Didier Tharreau, Pierre Gladieux.

Investigation: Maud Thierry, Florian Charriat, Joëlle Milazzo, Henri Adreit, Sébastien Ravel, Sandrine Cros-Arteil, Sonia borron, Violaine Sella, Pierre Gladieux.

Methodology: Maud Thierry, Thomas Kroj, Elisabeth Fournier, Didier Tharreau, Pierre Gladieux.

Project administration: Thomas Kroj, Renaud Ioos, Elisabeth Fournier, Didier Tharreau, Pierre Gladieux.

Resources: Didier Tharreau, Pierre Gladieux.

Software: Maud Thierry, Florian Charriat, Sébastien Ravel, Pierre Gladieux.

Supervision: Thomas Kroj, Renaud Ioos, Elisabeth Fournier, Didier Tharreau, Pierre Gladieux.

Validation: Maud Thierry, Joëlle Milazzo, Didier Tharreau.

Visualization: Maud Thierry, Florian Charriat, Pierre Gladieux.

Writing – original draft: Maud Thierry, Pierre Gladieux.

Writing – review & editing: Maud Thierry, Florian Charriat, Thomas Kroj, Elisabeth Fournier, Didier Tharreau, Pierre Gladieux.

References

1. Desprez-Loustau ML, Robin C, Buée M, Courtecuisse R, Garbaye J, Suffert F, et al. The fungal dimension of biological invasions. Trends in Ecology & Evolution. 2007; 22(9):472–80. https://doi.org/10.1016/j.tree.2007.04.005 PMID: 17509727
2. Rizzo DM. Exotic species and fungi: Interactions with fungal, plant and animal communities. The Fungal Community: Its Organization and Role in the Ecosystem. Boca Raton: CRC Press; 2005. p. 857–80.
3. Strange RN, Scott PR. Plant Disease: A Threat to Global Food Security. Annual Review of Phytopathology. 2005; 43(1):83–116. https://doi.org/10.1146/annurev.phyto.43.113004.133839 PMID: 16078878
4. Stukenbrock EH, McDonald BA. The Origins of Plant Pathogens in Agro-ecosystems. Annual Review of Phytopathology. 2008; 46(1):75–100. https://doi.org/10.1146/annurev.phyto.010708.154114 PMID: 18680424
5. Parker IM, Gilbert GS. The Evolutionary Ecology of Novel Plant-Pathogen Interactions. Annual Review of Ecology, Evolution and Systematics. 2004; 35(1):675–700.
6. Yarwood CE. Man-Made Plant Diseases. Science. 1970; 168:218–20. https://doi.org/10.1126/science.168.3928.218 PMID: 17747092
7. Palm ME. Systematics and the Impact of Invasive Fungi on Agriculture in the United States. BioScience. 2001; 51(2):141–7.
8. Gladieux P, Feurtey A, Hood ME, Snirc A, Clavel J, Dutech C, et al. The population biology of fungal invasions. Molecular Ecology. 2015:1969–86. https://doi.org/10.1111/mec.13028 PMID: 25469955
9. Thrall PH, Oakeshott JG, Fitt G, Southerton S, Burdon JJ, Sheppard A, et al. Evolution in agriculture: the application of evolutionary approaches to the management of biotic interactions in agro-ecosystems. Evol Appl. 2011; 4(2):200–15. https://doi.org/10.1111/j.1752-4571.2010.00179.x PMID: 25567968
10. Williams PD. Darwinian interventions: taming pathogens through evolutionary ecology. Trends in Parasitology. 2009; 26(2):83–92. https://doi.org/10.1016/j.pt.2009.11.009 PMID: 20036799
11. Hessenauer P, Fijarczyk A, Martin H, Prunier J, Charron G, Chapuis J, et al. Hybridization and introgression drive genome evolution of Dutch elm disease pathogens. Nature ecology & evolution. 2020; 4(4):626–38. https://doi.org/10.1038/s41559-020-1133-6 PMID: 32123324
12. Barrett LG, Thrall PH, Burdon JJ, Linde CC. Life history determines genetic structure and evolutionary potential of host-parasite interactions. Trends in Ecology & Evolution. 2008; 23(12):678–85. https://doi.org/10.1016/j.tree.2008.06.017 PMID: 18947899
13. Giraud T, Gladieux P, Gavrilets S. Linking the emergence of fungal plant diseases with ecological speciation. Trends in ecology & evolution. 2010; 25(7):387–95. https://doi.org/10.1016/j.tree.2010.03.006 PMID: 20434790

14. Gladieux P, Guerin F, Giraud T, Caffier V, Lemaire C, Parisi L, et al. Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. Molecular Ecology. 2011; 20(21):4521–32. https://doi.org/10.1111/j.1365-294X.2011.05288.x PMID: 21967446

15. Giraud T, Villareal LMMA, Austerlitz F, Le Gac M, Lavigne C. Importance of the Life Cycle in Sympatric Host Race Formation and Speciation of Pathogens. Phytopathology. 2006; 96(3):280–7. https://doi.org/10.1094/PHYTO-96-0280 PMID: 18944443

16. Servedio MR, Doorn GSV, Kopp M, Frame AM, Nosil P. Magic traits in speciation: 'magic' but not rare? Trends in Ecology & Evolution. 2011; 26(8):389–97.

17. Peever TL. Role of host specificity in the speciation of Ascochyta pathogens of cool season food legumes. European Journal of Plant Pathology. 2007; 119(1):119–26.

18. Stukenbrock EH, Bataillon T, Dutheil JY, Hansen TT, Li R, Zala M, et al. The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen Mycosphaerella graminicola and its wild sister species. Genome research. 2011; 21(12):2157–66. https://doi.org/10.1101/gr.118851.110 PMID: 21994252

19. Mercier A, Simon A, Lapalu N, Giraud T, Bardin M, Walker A-S, et al. Population genomics reveals molecular determinants of specialization to tomato in the polyphagous fungal pathogen Botrytis cinerea in France. Phytopathology. 2021(ja). https://doi.org/10.1094/PHYTO-07-20-0302-FI PMID: 33829853

20. Lemaire C, De Gracia M, Leroy T, Michalecka M, Lindhard-Pedersen H, Guerin F, et al. Emergence of new virulent populations of apple scab from nonagricultural disease reservoirs. New Phytologist. 2016; 209(3):1220–9. https://doi.org/10.1111/nph.13658 PMID: 26428268

21. Schulze-Lefert P, Panstruga R. A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends in Plant Science. 2011; 16(3):117–25. https://doi.org/10.1016/j.tplants.2011.01.001 PMID: 21317020

22. Uhse S, Djamei A. Effectors of plant-colonizing fungi and beyond. PLoS Pathogens. 2018; 14(6): e1006992. https://doi.org/10.1371/journal.ppat.1006992 PMID: 29879221

23. Couch BC, Fudal I, Lebrun M-H, Tharreau D, Valent B, van Kim P, et al. Origins of Host-Specific Populations of the Blast Pathogen *Magnaporthe oryzae* in Crop Domestication With Subsequent Expansion of Pandemic Clones on Rice and Weeds of Rice. Genetics. 2005; 170(2):613–30. https://doi.org/10.1534/genetics.105.041780 PMID: 15802503

24. Taylor JW, Jacobson D, Fisher M. The Evolution of Asexual Fungi: Reproduction, Speciation and Classification. Annual Review of Phytopathology. 1999; 37:197–246. https://doi.org/10.1146/annurev.phyto.37.1.197 PMID: 11701822

25. Taylor JW, Hann-Soden C, Branco S, Sylvain I, Ellison CE. Clonal reproduction in fungi. Proceedings of the National Academy of Sciences. 2015; 112(29):8901–8. https://doi.org/10.1073/pnas.1503159112 PMID: 26195774

26. Lin X, Heitman J. The biology of the Cryptococcus neoformans species complex. Annu Rev Microbiol. 2006; 60:69–105. https://doi.org/10.1146/annurev.micro.60.080805.142102 PMID: 16704346

27. Li F, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, Mago R, et al. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. Nature Communications. 2019; 10(1):5068. https://doi.org/10.1038/s41467-019-12927-7 PMID: 31699975

28. Fellers JP, Sakthikumar S, He F, McRell K, Bakkeren G, Cuomo CA, et al. Whole-genome sequencing of multiple isolates of Puccinia triticina reveals asexual lineages evolving by recurrent mutations. G3 Genes|Genomes|Genetics. 2021. https://doi.org/10.1093/g3journal/jkab219 PMID: 34544127

29. de Vienne DM, Giraud T, Gouyon P-H. Lineage selection and the maintenance of sex. PLoS One. 2013; 8(6):e66906. https://doi.org/10.1371/journal.pone.0066906 PMID: 23825582

30. Salcedo A, Rutter W, Wang S, Akhunova A, Bolus S, Chao S, et al. Variation in the AvrSr35 gene determines Sr35 resistance against wheat stem rust race Ug99. Science. 2017; 358(6370):1604–6. https://doi.org/10.1126/science.aao7294 PMID: 29269474

31. Mboup M, Bahri B, Leconte M, De Vallavieille-Pope C, Kaltz O, Enjalbert J. Genetic structure and local adaptation of European wheat yellow rust populations: the role of temperature-specific adaptation. Evol Appl. 2012; 5(4):341–52. https://doi.org/10.1111/j.1752-4571.2011.00228.x PMID: 25568055

32. Frenkel O, Peever TL, Chilvers MI, Ozkilinc H, Can C, Abbo S, et al. Ecological Genetic Divergence of the Fungal Pathogen Didymella rabiei on Sympatric Wild and Domesticated Cicer spp. (Chickpea). Applied and Environmental Microbiology. 2010; 76(1):30–9. https://doi.org/10.1128/AEM.01181-09 PMID: 19897759

33. Farman M, Peterson G, Chen L, Starnes J, Valent B, Bachi P, et al. The Lolium pathotype of Magnaporthe oryzae recovered from a single blasted wheat plant in the United States. Plant Disease. 2017; 101(5):684–92. https://doi.org/10.1094/PDIS-05-16-0700-RE PMID: 30678560

34. Milazzo J, Pordel A, Ravel S, Tharreau D. First scientific report of Pyricularia oryzae causing gray leaf spot disease on perennial ryegrass (Lolium perenne) in France. Plant Disease. 2019; 103(5):1024–.

35. Pordel A, Ravel S, Charriat F, Gladieux P, Cros-Arteil S, Milazzo J, et al. Tracing the origin and evolutionary history of Pyricularia oryzae infecting maize and barnyard grass. Phytopathology. 2021; 111 (1):128–36. https://doi.org/10.1094/PHYTO-09-20-0423-R PMID: 33100147

36. Tanaka M, Nakayashiki H, Tosa Y. Population structure of Eleusine isolates of Pyricularia oryzae and its evolutionary implications. Journal of General Plant Pathology. 2009; 75(3):173–80.

37. Islam MT, Croll D, Gladieux P, Soanes DM, Persoons A, Bhattacharjee P, et al. Emergence of wheat blast in Bangladesh was caused by a South American lineage of Magnaporthe oryzae. BMC biology. 2016; 14(1):84. https://doi.org/10.1186/s12915-016-0309-7 PMID: 27716181

38. Inoue Y, Vy TTP, Yoshida K, Asano H, Mitsuoka C, Asuke S, et al. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. Science. 2017; 357(6346):80–3. https://doi.org/10.1126/science.aam9654 PMID: 28684523

39. Gladieux P, Condon B, Ravel S, Soanes D, Maciel JLN, Nhani A, et al. Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus Magnaporthe oryzae. mBio. 2018; 9(1). https://doi.org/10.1128/mBio.01219-17 PMID: 29487238

40. Valent B, Farman M, Tosa Y, Begerow D, Fournier E, Gladieux P, et al. Pyricularia graminis-tritici is not the correct species name for the wheat blast fungus: response to Ceresini et al.(MPP 20: 2). Molecular plant pathology. 2019; 20(2):173. https://doi.org/10.1111/mpp.12778 PMID: 30697917

41. Langner T, Białas A, Kamoun S. The blast fungus decoded: Genomes in flux. MBio. 2018; 9(2): e00571–18. https://doi.org/10.1128/mBio.00571-18 PMID: 29666287

42. Zeigler RS. Recombination in *Magnaporthe Grisea*. Annual Review of Phytopathology. 1998; 36 (1):249–75. https://doi.org/10.1146/annurev.phyto.36.1.249 PMID: 15012500

43. Tharreau D, Fudal I, Andriantsimialona D, Utami D, Fournier E, Lebrun M-H, et al. World population structure and migration of the rice blast fungus, Magnaporthe oryzae. Advances in genetics, genomics and control of rice blast disease: Springer; 2009. p. 209–15.

44. Saleh D, Milazzo J, Adreit H, Fournier E, Tharreau D. South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, Magnaporthe oryzae. New Phytologist. 2014; 201(4):1440–56. https://doi.org/10.1111/nph.12627 PMID: 24320224

45. Saleh D, Xu P, Shen Y, Li C, Adreit H, Milazzo J, et al. Sex at the origin: an Asian population of the rice blast fungus Magnaporthe oryzae reproduces sexually. Molecular Ecology. 2012; 21(6):1330–44. https://doi.org/10.1111/j.1365-294X.2012.05469.x PMID: 22313491

46. Kumar J, Nelson RJ, Zeigler RS. Population structure and dynamics of Magnaporthe grisea in the Indian Himalayas. Genetics. 1999; 152(3):971–84. https://doi.org/10.1093/genetics/152.3.971 PMID: 10388817

47. Zhong Z, Chen M, Lin L, Han Y, Bao J, Tang W, et al. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. The ISME journal. 2018; 12(8):1867–78. https://doi.org/10.1038/s41396-018-0100-6 PMID: 29568114

48. Latorre SM, Reyes-Avila CS, Malmgren A, Win J, Kamoun S, Burbano HA. Differential loss of effector genes in three recently expanded pandemic clonal lineages of the rice blast fungus. BMC biology. 2020; 18(1):1–15.

49. Gladieux P, Ravel S, Rieux A, Cros-Arteil S, Adreit H, Milazzo J, et al. Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. mBio. 2018; 9(2). https://doi.org/10.1128/mBio.01806-17 PMID: 29615506

50. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of individual ancestry coefficients. Genetics. 2014; 196(4):973–83. https://doi.org/10.1534/genetics.113.160572 PMID: 24496008

51. Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. Genetics. 2006; 172(4):2665–81. https://doi.org/10.1534/genetics.105.048975 PMID: 16489234

52. Maynard Smith J. Analyzing the mosaic structure of genes. Journal of molecular evolution. 1992; 34 (2):126–9. https://doi.org/10.1007/BF00182389 PMID: 1556748

53. Jakobsen IB, Easteal S. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. Bioinformatics. 1996; 12(4):291–5. https://doi.org/10.1093/bioinformatics/12.4.291 PMID: 8902355

54. Gallet R, Fontaine C, Bonnot F, Milazzo J, Tertois C, Adreit H, et al. Evolution of Compatibility Range in the Rice-Magnaporthe oryzae System: An Uneven Distribution of R Genes Between Rice Subspecies. Phytopathology. 2016; 106(4):348–54. https://doi.org/10.1094/PHYTO-07-15-0169-R PMID: 26667186

55. Silué D, Notteghem JL, Tharreau D. Evidence of a gene-for-gene relationship in the Oryza sativa-Magnaporthe grisea pathosystem. Phytopathology. 1992; 82(5):577–80.

56. Khan MAI, Ali MA, Monsur MA, Kawasaki-Tanaka A, Hayashi N, Yanagihara S, et al. Diversity and distribution of rice blast (Pyricularia oryzae Cavara) races in Bangladesh. Plant Disease. 2016; 100 (10):2025–33. https://doi.org/10.1094/PDIS-12-15-1486-RE PMID: 30683013

57. Valent B, Farrall L, Chumley FG. Magnaporthe grisea genes for pathogenicity and virulence identified through a series of backcrosses. Genetics. 1991; 127(1):87–101. https://doi.org/10.1093/genetics/127.1.87 PMID: 2016048

58. Liao J, Huang H, Meusnier I, Adreit H, Ducasse A, Bonnot F, et al. Pathogen effectors and plant immunity determine specialization of the blast fungus to rice subspecies. eLife. 2016; 5:e19377. https://doi.org/10.7554/eLife.19377 PMID: 28008850

59. Yoshida K, Saunders DGO, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, et al. Host specialization of the blast fungus Magnaporthe oryzae is associated with dynamic gain and loss of genes linked to transposable elements. BMC genomics. 2016; 17(1):1. https://doi.org/10.1186/s12864-016-2690-6 PMID: 27194050

60. Pordel A, Ravel S, Charriat F, Gladieux P, Cros-Arteil S, Milazzo J, et al. Tracing the origin and evolutionary history of Pyricularia oryzae infecting maize and barnyard grass. Phytopathology. 2020: PHYTO-09. https://doi.org/10.1094/PHYTO-09-20-0423-R PMID: 33100147

61. Seong K, Krasileva KV. Computational Structural Genomics Unravels Common Folds and Novel Families in the Secretome of Fungal Phytopathogen Magnaporthe oryzae. 2021(0894–0282 (Print)).

62. Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. World map of the Köppen-Geiger climate classification updated. 2006.

63. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International journal of climatology. 2017; 37(12):4302–15.

64. Fitt BD, Huang Yj Fau—van den Bosch F, van den Bosch F Fau—West JS, West JS. Coexistence of related pathogen species on arable crops in space and time. Annual Review of Phytopathology. 2006; 44(0066–4286 (Print)):163–82. https://doi.org/10.1146/annurev.phyto.44.070505.143417 PMID: 16602949

65. Anonymous. Revised Consensus Document on the Biology of Rice (Oryza sativa L.). Organisation for Economic Co-operation and Development; 2021.

66. Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD, et al. Global dissemination of a single mutation conferring white pericarp in rice. PLoS genetics. 2007; 3(8):e133. https://doi.org/10.1371/journal.pgen.0030133 PMID: 17696613

67. Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nature communications. 2011; 2(1):1–10. https://doi.org/10.1038/ncomms1467 PMID: 21915109

68. Khush GS. Origin, dispersal, cultivation and variation of rice. Plant molecular biology. 1997; 35(1):25–34.

69. Gupta PC, O'toole JC. Upland rice: a global perspective: Int. Rice Res. Inst.; 1986. 186 p.

70. Civáň P, Ali S, Batista-Navarro R, Drosou K, Ihejieto C, Chakraborty D, et al. Origin of the aromatic group of cultivated rice (Oryza sativa L.) traced to the Indian subcontinent. Genome biology and evolution. 2019; 11(3):832–43. https://doi.org/10.1093/gbe/evz039 PMID: 30793171

71. Nosil P, Vines TH, Funk DJ. Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution. 2005; 59(4):705–19. PMID: 15926683

72. Hansen EM. Speciation in plant pathogenic fungi—the influence of agricultural practice. Canadian Journal of Plant Pathology. 1987; 9(4):403–10.

73. Stukenbrock EH. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. New Phytologist. 2013; 199(4):895–907. https://doi.org/10.1111/nph.12374 PMID: 23782262

74. Restrepo S, Tabima JF, Mideros MF, Grünwald NJ, Matute DR. Speciation in fungal and oomycete plant pathogens. Annual review of phytopathology. 2014; 52:289–316. https://doi.org/10.1146/annurev-phyto-102313-050056 PMID: 24906125

75. Otto SP, Lenormand T. Resolving the paradox of sex and recombination. Nature Reviews Genetics. 2002; 3(4):252–61. https://doi.org/10.1038/nrg761 PMID: 11967550

76. Orr HA. The rate of adaptation in asexuals. Genetics. 2000; 155(2):961–8. https://doi.org/10.1093/genetics/155.2.961 PMID: 10835413

77. LeBlanc N, Cubeta MA, Crouch JA. Population Genomics Trace Clonal Diversification and Intercontinental Migration of an Emerging Fungal Pathogen of Boxwood. Phytopathology. 2020; 111(1):184–93. https://doi.org/10.1094/PHYTO-06-20-0219-FI PMID: 33048629

78. Short DPG, Gurung S, Gladieux P, Inderbitzin P, Atallah ZK, Nigro F, et al. Globally invading populations of the fungal plant pathogen V erticillium dahliae are dominated by multiple divergent lineages. Environmental Microbiology. 2015; 17(8):2824–40. https://doi.org/10.1111/1462-2920.12789 PMID: 25630463

79. Short DPG, Gurung S, Hu X, Inderbitzin P, Subbarao KV. Maintenance of sex-related genes and the co-occurrence of both mating types in Verticillium dahliae. PLoS One. 2014; 9(11):e112145. https://doi.org/10.1371/journal.pone.0112145 PMID: 25383550

80. Capron A, Feau N, Heinzelmann R, Barnes I, Benowicz A, Bradshaw RE, et al. Signatures of post-glacial genetic isolation and human-driven migration in the Dothistroma needle blight pathogen in western Canada. Phytopathology. 2021; 111(1):116–27. https://doi.org/10.1094/PHYTO-08-20-0350-FI PMID: 33112215

81. Rogério F, Van Oosterhout C, Ciampi-Guillardi M, Correr FH, Hosaka GK, Cros-Arteil S, et al. Means, motive, and opportunity for biological invasions: genetic introgression in a fungal pathogen. bioRxiv. 2021.

82. Stauber L, Badet T, Feurtey A, Prospero S, Croll D. Emergence and diversification of a highly invasive chestnut pathogen lineage across southeastern Europe. Elife. 2021; 10:e56279. https://doi.org/10.7554/eLife.56279 PMID: 33666552

83. Dutech C, Barres B, Bridier J, Robin C, Milgroom MG, Ravigné V. The chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. Molecular Ecology. 2012; 21(16):3931–46. https://doi.org/10.1111/j.1365-294X.2012.05575.x PMID: 22548317

84. Ali S, Gladieux P, Leconte M, Gautier A, Justesen AF, Hovmoller MS, et al. Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Puccinia striiformis f. sp tritici. Plos Pathogens. 2014; 10(1). https://doi.org/10.1371/journal.ppat.1003903 PMID: 24465211

85. Bueno-Sancho V, Persoons A, Hubbard A, Cabrera-Quio LE, Lewis CM, Corredor-Moreno P, et al. Pathogenomic Analysis of Wheat Yellow Rust Lineages Detects Seasonal Variation and Host Specificity. Genome Biology and Evolution. 2017; 9(12):3282–96. https://doi.org/10.1093/gbe/evx241 PMID: 29177504

86. Weldon WA, Knaus BJ, Grünwald NJ, Havill JS, Block MH, Gent DH, et al. Transcriptome-derived amplicon sequencing markers elucidate the US Podosphaera macularis population structure across feral and commercial plantings of Humulus lupulus. Phytopathology. 2021; 111(1):194–203. https://doi.org/10.1094/PHYTO-07-20-0299-FI PMID: 33044132

87. Hessenauer P, Feau N, Gill U, Schwessinger B, Brar GS, Hamelin RC. Evolution and adaptation of forest and crop pathogens in the Anthropocene. Phytopathology. 2021; 111(1):49–67. https://doi.org/10.1094/PHYTO-08-20-0358-FI PMID: 33200962

88. Takabayashi N, Tosa Y, Oh HS, Mayama S. A Gene-for-Gene Relationship Underlying the Species-Specific Parasitism of Avena/Triticum Isolates of Magnaporthe grisea on Wheat Cultivars. Phytopathology. 2002; 92(11):1182–8. https://doi.org/10.1094/PHYTO.2002.92.11.1182 PMID: 18944243

89. Sweigard JA, Carroll AM, Kang S, Farrall L, Chumley FG, Valent B. Identification, Cloning, and Characterization of *PWL2*, a Gene for Host Species Specificity in the Rice Blast Fungus. The Plant cell. 1995; 7(8):1221–33. https://doi.org/10.1105/tpc.7.8.1221 PMID: 7549480

90. Asuke S, Tanaka M, Hyon G-S, Inoue Y, Vy TTP, Niwamoto D, et al. Evolution of an Eleusine-specific subgroup of Pyricularia oryzae through a gain of an avirulence gene. Molecular Plant-Microbe Interactions. 2020; 33(2):153–65. https://doi.org/10.1094/MPMI-03-19-0083-R PMID: 31804154

91. Anh VL, Inoue Y, Asuke S, Vy TTP, Anh NT, Wang S, et al. Rmg8 and Rmg7, wheat genes for resistance to the wheat blast fungus, recognize the same avirulence gene AVR-Rmg8. Molecular plant pathology. 2018; 19(5):1252–6. https://doi.org/10.1111/mpp.12609 PMID: 28846191

92. Kang S, Sweigard Ja Fau—Valent B, Valent B. The PWL host specificity gene family in the blast fungus Magnaporthe grisea. Molecular Plant Microbe Interactions. 1995(0894–0282 (Print)). https://doi.org/10.1094/mpmi-8-0939 PMID: 8664503

93. Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. Genetic structure and diversity in Oryza sativa L. Genetics. 2005; 169(3):1631–8. https://doi.org/10.1534/genetics.104.035642 PMID: 15654106

94. Chiapello H, Mallet L, Guerin C, Aguileta G, Amselem J, Kroj T, et al. Deciphering Genome Content and Evolutionary Relationships of Isolates from the Fungus Magnaporthe oryzae Attacking Different Host Plants. Genome Biol Evol. 2015; 7(10):2896–912. https://doi.org/10.1093/gbe/evv187 PMID: 26454013

95. Adreit H, Santoso, Andriantsimialona D, Utami DW, Notteghem JL, Lebrun MH, et al. Microsatellite markers for population studies of the rice blast fungus, Magnaporthe grisea. Molecular Ecology Notes. 2007; 7(4):667–70.

96. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011; 17(1):10–2.

97. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: a parallel assembler for short read sequence data. Genome research. 2009; 19(6):1117–23. https://doi.org/10.1101/gr.089532.108 PMID: 19251739

98. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome research. 2017; 27(5):768–77. https://doi.org/10.1101/gr.214346.116 PMID: 28232478

99. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 2015; 32(5):767–9. https://doi.org/10.1093/bioinformatics/btv661 PMID: 26559507

100. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. Gene Prediction: Springer; 2019. p. 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5 PMID: 31020555

101. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic acids research. 2005; 33(suppl_2):W465–W7. https://doi.org/10.1093/nar/gki458 PMID: 15980513

102. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; 31(19):3210–2. https://doi.org/10.1093/bioinformatics/btv351 PMID: 26059717

103. Nielsen H. Predicting secretory proteins with SignalP. Protein function prediction: Springer; 2017. p. 59–73. https://doi.org/10.1007/978-1-4939-7015-5_6 PMID: 28451972

104. Emanuelsson O, Nielsen H, Brunak S, Von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of molecular biology. 2000; 300(4):1005–16. https://doi.org/10.1006/jmbi.2000.3903 PMID: 10891285

105. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. Journal of molecular biology. 2004; 338(5):1027–36. https://doi.org/10.1016/j.jmb.2004.03.016 PMID: 15111065

106. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology. 2001; 305(3):567–80. https://doi.org/10.1006/jmbi.2000.4315 PMID: 11152613

107. Bhagwat M, Aravind L. Psi-blast tutorial. Comparative genomics: Springer; 2007. p. 177–86. https://doi.org/10.1007/978-1-59745-514-5_10 PMID: 17993673

108. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic acids research. 2012; 40(W1):W445–W51. https://doi.org/10.1093/nar/gks479 PMID: 22645317

109. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome biology. 2015; 16(1):157. https://doi.org/10.1186/s13059-015-0721-2 PMID: 26243257

110. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic acids research. 2010; 38(suppl_2):W7–W13. https://doi.org/10.1093/nar/gkq291 PMID: 20435676

111. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 38:1358–70. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x PMID: 28563791

112. Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology and Evolution. 2006; 23(2):254–67. https://doi.org/10.1093/molbev/msj030 PMID: 16221896

113. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9):1312–3. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

114. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version 3.61 2019.

115. Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019; 35(10):1786–8. https://doi.org/10.1093/bioinformatics/bty875 PMID: 30321304

116. Saleh D, Milazzo J, Adreit H, Tharreau D, Fournier E. Asexual reproduction induces a rapid and permanent loss of sexual reproduction capacity in the rice fungal pathogen Magnaporthe oryzae: results of in vitro experimental evolution assays. BMC evolutionary biology. 2012; 12(1):1–16. https://doi.org/10.1186/1471-2148-12-42 PMID: 22458778