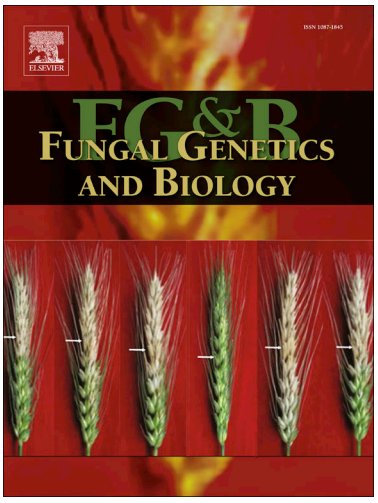
Journal Pre-proofs



Population structure of *Pyricularia oryzae* on rice in Vietnam reveals diversified populations with four pandemic and two endemic clusters

Lieu Thi Le, Henri Adreit, Loan Thi Ha, Joelle Milazzo, Michel Lebrun, Didier Tharreau, Xuan Hoi Pham, Hai Thanh Nguyen, Elisabeth Fournier, Giang Thi Hoang

PII: S1087-1845(23)00025-7
DOI: https://doi.org/10.1016/j.fgb.2023.103794
Reference: YFGBI 103794

To appear in: *Fungal Genetics and Biology*

Received Date: 11 August 2022
Revised Date: 25 February 2023
Accepted Date: 27 March 2023

Please cite this article as: Thi Le, L., Adreit, H., Thi Ha, L., Milazzo, J., Lebrun, M., Tharreau, D., Hoi Pham, X., Thanh Nguyen, H., Fournier, E., Thi Hoang, G., Population structure of *Pyricularia oryzae* on rice in Vietnam reveals diversified populations with four pandemic and two endemic clusters, *Fungal Genetics and Biology* (2023), doi: https://doi.org/10.1016/j.fgb.2023.103794

Population structure of *Pyricularia oryzae* on rice in Vietnam reveals diversified populations with four pandemic and two endemic clusters

Lieu Thi Le [a,b], Henri Adreit [c,d], Loan Thi Ha [a], Joelle Milazzo [c,d], Michel Lebrun [b], Didier Tharreau [c,d], Xuan Hoi Pham [a], Hai Thanh Nguyen[e], Elisabeth Fournier [c, *], Giang Thi Hoang [a,e *]

[a] *Agricultural Genetics Institute, VAAS, National Key Laboratory for Plant Cell Biotechnology, LMI RICE-2, Hanoi, Vietnam*

[b] *University of Science and Technology of Hanoi, Hanoi, Vietnam*

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

[d] *CIRAD, UMR PHIM, 34090 Montpellier, France*

[e] *Vietnam National University of Agriculture, Faculty of Biotechnology, Faculty of Agronomy, Hanoi, Vietnam*

*** Co-corresponding authors:**

Email address: nuocngamos@yahoo.com (G.T. Hoang); elisabeth.fournier@inrae.fr (E. Fournier)

Postal address: Agricultural Genetics Institute, Pham Van Dong street, Bac Tu Liem district, 100000 Hanoi, Vietnam (G.T. Hoang); UMR PHIM, TA A 120/K, Campus International de Baillarguet, F34398 Montpellier cedex 5, France (E. Fournier)

Running title: Population structure of *Pyricularia oryzae* in Vietnam

Keywords: *Pyricularia oryzae,* rice blast, SSR markers, genetic diversity, population structure, Vietnam

ABSTRACT

We characterized the genetic structure of 609 strains of *Pyricularia oryzae,* the fungal pathogen causing rice blast disease, in three main regions in Vietnam using microsatellites (SSR) markers. From the 447 distinct multilocus genotypes identified, six genetic clusters were defined, all of them showing elevated genetic and genotypic diversities. Four of these clusters were related to rice-attacking lineages already described at the worldwide scale, whereas the two remaining clusters were endemic to Vietnam. Strains were unevenly distributed into the six clusters depending on their groups of rice variety (*indica* / *japonica*) or type of varieties (traditional / modern) of origin, but none of the clusters was specifically related to these two factors. The highest diversity of blast population was found in Northern mountainous area, and the lowest in Red River Delta in both term of genetic diversity and gene diversity. Hierarchical AMOVAs confirmed that all three factors considered (rice variety group, type of variety origin and geography) significantly contributed to the population structure of *P. oryzae* in Vietnam, with highest contribution from rice variety group. Mating types were unevenly distributed among clusters. Combined with results of female fertility and linkage disequilibirum, we hypothesized that clonal reproduction probably occurred in all clusters, but that sexual reproduction likely took place at least in some restricted areas in the Northern mountainous area for strains belonging to the cluster related to the previously described recombinant lineage (worldwide lineage 1). Our study pictures the genetic diversity, population structure and reproductive mode of the blast fungus in central and north Vietnam, and shows that the observed population structure is explained by several factors, the most important one being the variability of rice variety. All these new information might help for elaborating appropriate strategies to controlling the blast disease.

1. Introduction

Rice (*Oryza sativa* L.) is one of the most important staple foods for Vietnamese people due to their cultural behavior. Vietnam is the fifth largest rice exporter in billion $ and the second or third (depending on years) in million tons of paddy. Rice also supplies approximately 50% of the food demands worldwide (Khush, 2005; Latif et al., 2011). This cereal thus plays a crucial role for economy and food security not only in Vietnam but also in the whole world. Vietnam is one of the Asia's richest hotspots of biological, ecological and cultural diversities (Higgins et al., 2019; Hue et al.,

2018). Notably, the exceptional richness of rice varieties in Vietnam, gathering both traditional and modern varieties, supplies elevated genetic diversity for several important genes involved in agronomically important traits, such as resistance to biotic and abiotic stresses (Higgins et al., 2021; Nguyen et al., 2012; Vu et al., 2016). Rice diversity represents an essential resource for plant breeders for improving breeding programs for high productivity, quality and resistance to diseases. The erosion of richness and genetic variance of traditional rice varieties has accelerated markedly in recent decades, being directly or indirectly linked to growing urbanization, industrialization and the development of tourist activities and services (Katayama et al., 2015; Nakagahra et al., 2000; Smith et al., 2019). Nevertheless, the global shrinking of rice cultivated biodiversity is a threatening bell for long-term sustainable development of agriculture in Vietnam (Lang et al., 2010).

Rice production is regularly jeopardized by various diseases, among which the airborne fungus *Pyricularia oryzae* Cavara (synonym *Magnaporthe oryzae*). This ascomycete causes blast disease on several *Poaceae* hosts (Couch et al., 2005; Gladieux et al., 2018a), responsible for dramatic damages on rice production, with yield losses in a range of 1-5% at the world scale (Savary, 2000) but sometimes locally reaching dramatic levels of 80-100% (Talbot, 2003). *P. oryzae* is a hemibiotrophic pathogen whose life cycle starts with a biotrophic phase followed by a necrotrophic phase (Wilson and Talbot, 2009). It is one type of rapidly evolving pathogen, and a premier model organism for plant pathology (Dean et al., 2012; Ebbole, 2007; Huang et al., 2014; Talbot, 2003; Valent, 1990). Rice immunity against blast, as against other fungal pathogens (Liu et al., 2014) includes quantitative mechanisms such as basal resistance (Vergne et al., 2010), and specific mechanisms based on the gene-for-gene relationship (Flor, 1971; Silué et al., 1992). However, specific genetic resistances against blast, triggered by so-called major resistance (R) genes, are only short-time effective when deployed in fields, due to the high capacity of *P. oryzae* to rapidly adapt to such resistances. Indeed, fungal effectors involved in interactions with the host comprise avirulence (AVR) factors acting in gene-for-gene interactions able to escape specific recognition R genes (Chiapello et al., 2015; Kim et al., 2019; Zhang and Xu, 2014). Different mechanisms including mutation and genetic recombination can contribute to the fast evolution of *P. oryzae* pathogen, especially at AVR loci in response to selection by R genes deployed in the field (Noguchi et al., 2006; Noguchi, 2011). To date, more than 100 R genes effective against blast have been described, among which 38 have been molecularly cloned and characterized (Manoj et al., 2019; Hu et al., 2022). Besides, at least 500 QTLs involved in partial resistance against blast have been defined, but only

few major QTLs have been identified by molecular markers (Ashkani et al., 2014; Ballini et al., 2008). On the pathogen's side, more than 40 *AVR* genes have been identified in *P. oryzae*, in which 12 have been cloned (Hu et al., 2022). Analyzing the variation of *AVR* genes can help evaluating the efficacy of cognate resistance genes (Selisana et al., 2017).

However, management of rice blast is still an issue. Chemical control is fast but not persistent and harmful for environment and people's health. Most prophylactic actions are ineffective, and genetic control is only short-time effective (Orasen et al., 2020; Silué et al., 1992). An accurate characterization of *P. oryzae* population biology allows a better understanding of the factors underlying population structure and spread of virulent genotypes, and provides useful knowledge for resistance screening and exploitation of resistance diversity.

Population structure of the rice blast fungus *P. oryzae* has been described at the worldwide scale by several studies during the last decade using various methods (Gladieux et al., 2018b; Saleh et al., 2012, 2014; Thierry et al., 2021; Zhong et al., 2018). By analyzing 55 populations from 15 countries with 12 simple-sequence repeat (SSR) markers, (Saleh et al., 2012, 2014) confirmed that South-East Asia was the center of origin of blast on rice and that most of the genetic diversity observed around the world was represented in this region. In this study, four main *P. oryzae* lineages pathogenic to rice were clarified, one of them being restricted to Asia and showing footprints of recombination, whereas the others are pandemic and strictly clonal. More recently, (Gladieux et al., 2018b) used the genome sequencing data of 50 worldwide *P. oryzae* rice-attacking strains and confirmed the existence of the four previously described lineages: only one lineage endemic to continental Southeast Asia (lineage 1) displayed signatures of sexual recombination, whereas the three other lineages were pandemic and clonal, with two of them associated with *japonica* (lineage 2) and *indica* rice (lineage 3) in rainfed and irrigated conditions. This global picture of *P. oryzae* populations divided in four main lineages on rice was also confirmed by recent studies using population genomics (Lattore et al., 2020, Thierry et al., 2021, Zhong et al., 2018), these studies showing that lineages 5 and 6 described by (Gladieux et al., 2018b) were in fact grouped with worldwide lineage 1. The time at which *P. oryzae* rice-attacking lineages separated from their last common ancestor is still a matter of debate (about 1,000 years ago in (Gladieux et al., 2018b), about 200 years ago in (Latorre et al., 2020) but likely occurred long after the initial domestication of rice. The reproduction of *P. oryzae in natura* is asexual in most parts of the world except in some restricted regions of South East Asia (Gladieux et al., 2018b; Saleh et al., 2012, 2014). However,

sexual reproduction is possible *in vitro*. *P. oryzae* is a heterothallic fungus, in which the two different mating types (called MAT1 and MAT2) are different idiomorphs present at the MAT locus in different thalli (Kanamori et al., 2007; Kang et al., 1994; Saleh et al., 2012). Sexual reproduction occurs when two compatible strains (i.e. of opposite mating types) meet, with at least one of them being able of producing perithecia (the organs in which meiosis occurs), i.e. female fertile (Saleh et al., 2012). The fertility of *P. oryzae* ranges from total sterility (unable to mate with any other strain), to female sterility (able to mate as a male parent), to full fertility (able to mate as either male or female parents) (Saleh et al., 2012; Valent et al., 1986). The sexual reproduction can increase the genotypic diversity of *P. oryzae* populations and the recombinant individuals may have new ability to infect different host cultivars.

Until now, knowledge of genetic population structure and reproductive mode of *P. oryzae* in Vietnam is still limited, which has obstructed the efforts towards selection and breeding of resistant cultivars adapted to local agro-climatic conditions. Using MAGGY probe, (Don et al., 1999) identified five endemic lineages from 78 strains from several provinces of the Red River Delta and the Mekong Delta. Four of them came from Red River Delta, and a single genetic group was described in the Mekong Delta, but no genetic groups were shared between both Deltas. They also observed that 10 strains formed perithecia when crossed *in vitro* with a reference strain, showing that fertile strains existed in their sample. (Thuan et al., 2006) found 12 genetic groups by using 160 AFLP markers to classify 123 strains collected in the Red River Delta. Strains sampled on *japonica* and *indica* hosts clustered separately.

Understanding the population biology, genetic diversity and population structure of pathogen is key for a better comprehension of the occurrence of new pathotypes and for the screening and breeding of cultivars with efficient resistances. The aim of the present study was to fill the information gap on the genetic diversity and population biology and structure of *P. oryzae* in Vietnam, using large population sampling and reliable molecular markers. Especially, we aimed at (i) collecting fungal populations on a large geographic range in order to have an overview of population diversity of rice blast throughout Vietnam (as compared to previous works conducted in the same country), (ii) using SSR genetic markers allowing comparison with former studies performed at the worldwide scale, (iii) collecting information (when available) on rice varieties on which fungal isolates were sampled and connecting this information to pathogen diversity, and (iv) exploring the reproductive mode of blast in Vietnam by analyzing mating types and fertility of fungal

isolates. For that purpose, we gathered a comprehensive collection of 609 strains covering three major rice growing regions of the country: Central region, Red River Delta and Northern mountainous area, obtained with the same sampling methodology and representing real populations (i.e. several strains per field). This collection was analyzed using SSR markers used for previous studies at the worldwide scale, and was characterized for mating type and fertility using molecular and biological tests. We then analyzed how population structure relates with several factors (geography, rice variety group, type of rice variety).

2. Materials and methods

2.1. Sampling

Extensive sampling was performed from May 2018 to May 2019 throughout Red River Delta, Northern mountainous area and Central region of Vietnam (Table S1). In each province of those regions, we targeted two location points. At each location point, two different rice varieties from two different fields were selected for harvesting diseased plants. The diseased plants harvested from one variety, at one location were considered as one sample. In order to get 10 fungal strains per sample, we harvested as far as possible at least 20 diseased plants (neck, leave, collar or panicle) per field. Samples were harvested from farmer's field with permission of individual farmers or their neighbor. For each sample, the name of rice variety, locality and global positioning systems (GPS) coordinates were recorded. Harvested samples were kept dry for next steps in laboratory.

2.2. Isolation and storage of P. oryzae strains

For each sample, pieces of diseased plant were placed in humidified Petri dishes at 25 °C for 1 to 2 days to allow fungal sporulation. After 24 hours, monospore isolation was performed to obtain genetically pure *P. oryzae* strains, as previously described by (Silué and Nottéghem, 1990). Strains were stored on dried filter paper at -20 °C after seven days growth (25 °C, fluorescent lighting, 12-h photoperiod) on a 7-cm-diameter paper disk placed on autoclaved rice flour medium, as described by (Valent et al., 1986).

2.3. DNA extraction and SSR amplification

Fungal DNA was extracted by enzymatic cell wall digestion procedure (Sweigard et al., 1990) as described by (Adreit et al., 2007), after four days of culture of each *P. oryzae* strain in liquid medium (per liter: 10 g glucose, 3 g KNO_3, 2 g KH_2PO_4, 2 g yeast extract, 500 mg Penicillin G) in the dark at 25 °C. Extracted DNA was quantified and qualified by Nano Drop measurement (Nanodrop 2000). To validate their fungal origin, 1 µl of the extracted DNAs (20 ng/µl) was amplified with primer pair ITS5/ITS4 (ITS5: 5'-GGAAGTAAAAGTCGTAACAAGG-3', ITS4: 5'-TCCTCCGCTTATTGATATGC-3') specific to Ascomycetes, with the following PCR program: predenaturation at 95 °C for 15 min, denaturation at 94 °C for 30 s, hybridization at 60 °C for 90 s, extension at 72 °C for 60 s, repeat from step 2 to step 4 for 30 cycles, and final extension at 60 °C for 10 min. Five µl of amplification products were checked on 1% agarose gel.

Fungal population was genotyped with the following 13 SSR markers specifically developed for *P. oryzae*: pyrms47-48, pyrms427-428, pyrms99B-100, pyrms409-410, pyrms657-658, pyrms77B-78, pyrms63-64, pyrms83B-84, pyrms607-608, pyrms37-38, pyrms233-234, pyrms319-320, pyrms43B-44 (Adreit et al., 2007; Saleh et al., 2012). For PCR amplifications, SSR primers were combined in four multiplexes as follows: Multiplex 1 (pyrms47-48, 427-428, 99B-100), Multiplex 2 (pyrms409-410, 657-658, 77B-78), Multiplex 3 (pyrms63-64, 43B-44, 83B-84, 607-608), Multiplex 4 (pyrms37-38, 233-234, 319-320). PCR amplifications were run using QIAGEN Multiplex PCR kit in a total volume of 5 µl, including 2.5 µl of master Mix, 0.5 µl of 10X mix primers, 0.5 µl of 5X Q solution, and 1.5 µl of genomic DNA (10 ng/µl), with the following program: predenaturation at 95 °C for 15 min, denaturation at 94 °C for 30 s, hybridization at 57 to 63 °C for 90 s, extension at 72 °C for 60 s, repeat steps from 2 to 4 for 40 cycles, and final extension at 72 °C for 30 min. For some DNAs, SSR loci that failed to be amplified in multiplexes were re-amplified separately using the same protocol, then added to the corresponding multiplex. Amplification products of multiplexes were pooled together (1.5 µl of amplified products at 1/70 dilution mixed with 15 µl Formamide HiDi and GeneScan-500LIZ size marker, Applied Biosystems), and were then separated and sized by fluorescence measurement on a 16-capillary ABI Prism 3130XL device (Applied Biosystems, Foster City, USA). Raw released data were analyzed and converted to allele size using GeneMapper v. 4.0 (Applied Biosystems).

2.4. Analysis of P. oryzae population genetic structure

Genotypic diversity was evaluated with the R package poppr (Kamvar et al., 2014) using several estimators: number of multilocus genotypes (MLGs), number of MLGs expected at the lowest sample size (eMLGs, estimated using rarefaction method), Simpson's diversity index (λ, that is the probability that two genotypes randomly selected from a sample are different), and evenness of MLGs (E.5, which accounts for the distribution of MLGs' frequencies and which is less affected by unbalanced sample sizes than λ). Poppr was also used to estimate gene diversity as Nei's expected heterozygosity (H_e, that is the probability that two alleles picked randomly in the population are different, averaged over the 13 SSR loci) and the diversity of Stoddart and Taylor's index, G. The occurrence of genetic recombination was tested by estimating the unbiased index of association $\bar{r}_D$ (Agapow and Burt, 2005), that tends to 0 in populations where recombination occurs freely and to 1 in clonal populations (significant deviation from 0 was tested using 1000 replicates of random reshuffling of alleles).

Genetic population structure was inferred using Discriminant Analysis of Principal Components (DAPC) implemented in the R package adegenet 2.1.3 (Jombart et al., 2010; Jombart and Ahmed, 2011). The find.clusters function was used to determine the optimal number of genetic clusters (*K*) based on the Bayesian information criterion (BIC) with *K* varying from 2 to 12. The optimal number of clusters were confirmed by calculating the cross-entropy criterion as a function of the number of clusters using the software sNMF (Frichot et al., 2014), and by conducting a hierarchical agglomerative clustering with Ward's linkage (Ward, 1963) using the software DARwin (Perrier and Jacquemoud-Collet, 2006). Individuals with a probability of ancestry ≥ 70% in a given cluster were assigned to this cluster, otherwise they were classified as admixed. In parallel to DAPC, we used the R package ape 5.3 (Paradis and Schliep, 2019) to infer genealogical relationships among individuals by building a genealogical tree based on the "relative dissimilarity" distance matrix (which reflects the number of alleles differing between two individuals), using the Neighbour Joining tree reconstruction algorithm. We visually compared genealogical relationships with DAPC clustering by coloring branches on the tree according to the DAPC clusters, to which individuals were assigned. To estimate genetic differentiation among genetic clusters and to estimate the effect of different factors (geography, rice variety group, type of variety origin) on population subdivision, we performed analyses of molecular variance (AMOVA) (Excoffier et al., 1992) as implemented in the R package ade4 (Dray and Dufour, 2007).

2.5. Mating type and fertility assay

All genotyped strains were also tested for mating type and fertility. Mating type was determined by both PCR reactions and biological test. PCR amplification were performed with the primers specific for MAT1 (Primers A1 and A5) and MAT2 (Primers B15 and B16) defined by (Xu and Hamer, 1995). The PCR reactions were prepared in a total volume of 10 µl containing 20 ng of DNA template, 1 mM of PCR buffer, 0.2 mM of dNTP, 0.5 unit of Tag polymerase and 1 mM of each primer. Thermal cycling conditions for MAT1 were: initial denaturation step at 95 °C for 5 min; 30 cycles of 95 °C for 30 s, annealing at 60 °C for 40 s and extension at 72 °C for 1 min; followed by 72 °C for 5 min; final at 15 °C. The conditions for MAT2 were: initial denaturation step at 95 °C for 5 min; 30 cycles of 95 °C for 40 s, annealing at 55 °C for 1 min and extension at 72 °C for 1 min; followed by 72 °C for 5 min; final at 15 °C. PCR products were separated by electrophoresis in a 1% agarose gel. The biological mating type test that allows determining both the mating type and the fertility of a tested strain, was performed as described in (Saleh et al., 2012). Briefly, each tested strain is confronted to a hermaphrodite (i.e. female and male fertile) reference strain of known mating type on rice flour agar medium in a 90 mm Petri dish, and then let to grow for 21 days (2 days at 25 °C, then at 20 °C with continuous light). The mating type and female fertility status of the tested strain was deduced from its ability to form perithecia in the contact with the reference strain: (i) if there is no line of perithecia formed neither on the tested strain nor on the reference strain side, then the tested strain is of the same mating type as the reference strain; (ii) if there is one line of perithecia formed on the reference strain side, then the tested strain is of opposite mating type to the reference strain, and is female-sterile; (iii) if there are two lines of perithecia formed both on the tested strain side and on the reference strain side, then the tested strain is of opposite mating type to the reference strain, and is female-fertile. Each tested strain was confronted to four reference strains (two replicates for each cross).

3. Results

3.1. Population diversity within the whole population of P. oryzae

Among the samples collected during the whole campaign (Fig. 1), 647 monospore strains were recovered and successfully genotyped with 13 SSR markers. Genotypes with more than 25%

missing data were removed from the dataset, which finally comprised 609 genotyped strains: 263 from Northern mountainous area, 136 from Red River Delta and 210 from Central region (Table 1). Genotypic diversity was elevated in the whole population as well as in the three different geographic regions. A total of 447 distinct MLGs were identified, representing 73.4% of 609 genotyped strains. The number of distinct MLGs represented 79.1% (208 MLGs out of 263 genotyped strains) in Northern mountainous area, 66.2% (90 MLGs out of 136 genotyped strains) in the Red River Delta, and 76.7% (161 MLGs out of 210 genotyped strains) in the Central region. This still held true after rarefaction analysis, the number of expected MLGs (eMLGs) representing 87.3%, 67.2%, and 83.6% of the smallest common sample size (134 genotypes) in the whole population for Northern mountainous area, Red River Delta and Central region, respectively. For the 60 (90%) sampling locations in which at least four strains were genotyped, the percentage of distinct MLGs recovered at the scale of a single paddy field varied between 12.5% and 100%, and was above 50% for 54 out of 60 (88%) sampling locations, showing that the elevated genotypic diversity was also effective at the scale of the paddy field. The elevated genotypic diversity was also attested by Simpson's index λ ranging between 0.980 and 0.993 (0.996 in the whole population), and by evenness of MLGs ranging between 0.707 and 0.809 (0.663 in the whole population). Gene diversity estimated by Nei's expected heterozygosity H_e was also elevated in the whole population (0.586) as well as in the three different geographic regions (0.675, 0.417 and 0.585 in Northern mountainous area, Red River Delta and Central region, respectively).

3.2. Genetic structure of P. oryzae

The genetic subdivision of *P. oryzae* population in Vietnam was inferred using DAPC analysis, performed on the clone corrected dataset comprising 447 distinct MLGs (Fig. 2A, Fig. 3A). Scatterplots of coordinates of individuals and of the groups showed that the clearest subdivision was captured with *K*=6 clusters (Fig. 2A). The definition of genetic clusters was consolidated by running sNMF, with a clear drop-down of cross-entropy observed for *K*=6 (Fig. 2B), indicating that further subdivision in more than six clusters only added noise to the analysis. Hierarchical agglomerative clustering with Ward's linkage, performed using DARwin, also confirmed the subdivision into six clusters, equivalent to the six clusters inferred with DAPC (Fig. 2C). Finally, genealogical relationships, built from the relative dissimilarity distance matrix among the 447 MLGs, was concordant with the genetic subdivision obtained with DAPC: only 10 out of 447 MLGs grouped

with individuals not assigned to the same DAPC cluster (Fig. 3B). For the following analyses, the assignment of strains to genetic clusters is that defined by DAPC. In total, cluster 1 (red) comprised 232 strains (135 MLGs), cluster 2 (yellow) comprised 120 strains (106 MLGs), cluster 3 (light green) comprised 80 strains (74 MLGs), cluster 4 (turquoise blue) comprised 28 strains (21 MLGs), cluster 5 (dark blue) comprised 103 strains (67 MLGs), and cluster 6 (pink) comprised 46 strains (44 MLGs) (Table 2, Fig. 3B). At the scale of the paddy field, in 52% of the sampling locations we recovered strains that belong to at least two different genetic clusters, showing that these genetic clusters frequently coexisted in sympatry.

To investigate the relationships between genetic clusters inferred in Vietnam and the four rice-attacking genetic lineages already described at the worldwide scale (Gladieux et al., 2018b; Lattore et al., 2020; Thierry et al., 2021), we combined the 447 Vietnamese MLGs in this study with 338 unique MLGs representative of (and unambiguously assigned to) the four worldwide lineages (WL) WL1 to WL4. These 785 genotypes had in common 10 out of the 13 SSR markers (for the set of 338 genotypes representative of the worldwide lineages, Pyrms43B marker was not used, and Pyrms99B and Pyrms409 markers generated large number of missing values). This DAPC (named DAPC-785 in order to distinguish with the previous DAPC with 447 Vietnamese MLGs named DAPC-447) showed that four out of the six clusters inferred in Vietnam were related to worldwide genetic lineages (Fig. S1): strains from Vietnamese clusters 3, 4, 2 and 6 grouped with reference strains from WL1, WL2, WL3 and WL4, respectively. The two other Vietnamese clusters, i.e. clusters 1 and 5, were specific to Vietnam. Both analyses (DAPC-785 and DAPC-447) with only 13 MLGs misassigned.

Estimates of diversity within clusters confirmed that both genotypic and genetic diversities were elevated whatever the cluster considered (Table 2). Cluster 3 (WL1) was the most diverse both in terms of genotypic and genetic diversities, with almost the highest proportion of MLGs (92.5%) and Nei's genetic diversity index (H_e=0.634). Conversely, cluster 1 was the largest cluster in size, but was the less diverse (58.2% of MLGs, H_e=0.258). The remaining clusters 2 (WL3), 4 (WL2), 5 and 6 (WL4) showed elevated genotypic diversities (proportions of MLGs: 88.3%, 75%, 65% and 95.6% respectively) and low to intermediate gene diversities (H_e = 0.429, 0.311, 0.408 and 0.374 respectively). Genetic differentiation among clusters was estimated using pairwise F_{ST} implemented in the AMOVA (Table 3). F_{ST} values were all significantly different from 0 and ranged from 0.30 (between clusters 2 and 3) to 0.68 (between clusters 1 and 6), with a mean overall value of 0.49. All

of the pairwise F_{ST} values were higher than 0.30, indicating that all pairs of clusters were significantly differentiated with restricted gene flow between them.

3.3. Factors explaining genetic subdivision of P. oryzae populations

We first examined the geographic distribution of the six genetic DAPC clusters (Table 4, Fig. 1, Fig. S2). Cluster 1, the biggest one was present in all three regions, but ancestry coefficients were highest in Red River Delta. Moreover, this cluster was the most popular in Red River Delta, other clusters contributed very few isolates in this region. Cluster 2 distributed mostly in Northern mountainous area and Central region. Cluster 3 was mostly found in the Northern mountainous area. Cluster 4, the smallest cluster among all six defined clusters, distributed relatively balanced in the three regions with highest probabilities of ancestry in restricted places of Northern mountainous area and South-West of Central region. Clusters 5 distributed mostly in the Central provinces, being scarce in the Red River Delta region. Finally, cluster 6 was present mostly in Northern Vietnam, and more seldom in Central region. In conclusion, although there was some relationship between geographic localization and genetic assignment of individuals for some of the six genetic clusters, geography was clearly not the single factor explaining population subdivision of *P. oryzae* in Vietnam. The six clusters were found in sympatry in all three regions.

To better understand what best explains the observed population subdivision, we addressed how genetic clusters were organized according to the following factors: geographic origin of strains, rice variety group (*japonica* or *indica*) and type of variety (traditional or modern) of host plant of strains (Table 4). χ^2 tests were all highly significant, indicating a non-random distribution of genotypes among genetic clusters according to each factor. Hierarchical AMOVAs were performed for each factor by grouping genotypes by clusters then by factor (Table 5). In all cases, the percentage of molecular variation between strains within populations accounted in average for 52.24% (ranging from 50.89% to 53.61%) of total variation, and the contribution of variation among genetic clusters to total variation averaged 38.29% (ranging from 35.3% to 39.8%), confirming strong genetic differentiation among clusters. For each factor considered, the percentage of variation among levels of factor was significant, representing 9.25%, 12.47% and 6.62% of total variation for geographical region, rice group and variety type, respectively. Altogether, these results showed that all three factors significantly contributed to population subdivision of *P. oryzae* in Vietnam.

3.4. Recombination and possible occurrence of sexual reproduction

Recombination within each genetic cluster was addressed by estimating linkage disequilibrium using the $\bar{r}_D$ index. Values estimated with the total number of genotypes within each cluster (no clone correction) varied from 0.007 for cluster 6 to 0.265 for cluster 4 (Table 6). Cluster 1 and 6 had $\bar{r}_D$ value close to zero, the value expected for population with random sexual mating. Moreover, the observed $\bar{r}_D$ value after permutations for cluster 6 fell in the area of no linkage. After clone correction, $\bar{r}_D$ value of these two clusters reduced closer to zero and the observed $\bar{r}_D$ value after permutations for both clusters 1 and 6 fell in the area of no linkage. This indicated sexual reproduction might occur in these two clusters. The $\bar{r}_D$ values of cluster 2, cluster 3, cluster 4 and cluster 5 were significantly different from 0, meaning that the null hypothesis of full random mating could be rejected in those clusters; however, the highest value was 0.265 for cluster 4, i.e. far from 1 (which is the expected value under clonality) indicating that these clusters might have mixed reproductive modes (mainly clonal with some occurrence of recombination). This remained true when clone correction was applied (Fig. S3). The calculation of $\bar{r}_D$ for each cluster*region combination (Table 6) showed that the null hypothesis of random mating could not be rejected for cluster 2 in Northern mountainous area, cluster 4 in Central region (for clone corrected dataset) and in Northern mountains area. However, $\bar{r}_D$ estimates of linkage disequilibrium are hard to interpret *per se* since this index has been shown to be sensitive to other factors than recombination such as migration or hidden genetic substructure (De Meeûs and Balloux, 2004). Therefore, we also search for the possible occurrence of sexual reproduction by scoring mating types and female fertility in all clusters. Mating types, determined using PCR assays and confirmed by *in vitro* test crosses for all genotyped strains, were unevenly distributed among clusters. Clusters 1, 2 and 5 mainly comprised MAT2 strains (MAT1:MAT2 ratio of 4:228, 2:118, and 2:101 respectively). Clusters 3, 4 and 6 comprised mostly MAT1 strains (MAT1:MAT2 ratio of 70:10, 27:1, and 42:4, respectively). The less unbalanced ratio is therefore found in cluster 3 (12,5% of MAT2 isolates). *In vitro* test crosses against fertile testers of known mating types also allowed to score female fertility for almost all genotypes (541/609): only 16 fertile strains were recovered and were distributed in clusters 1 (2 MAT2 strains/232), cluster 2 (1 MAT2 strain/120) and cluster 3 (8 MAT1 and 5 MAT2 strains/80). The three female fertile strains in cluster 1 and 2, from Central region and Red River Delta, respectively (Table 6), were sampled in fields in which only MAT2 strains were found, indicating that biological

conditions were not fulfilled for sexual reproduction to occur. Interestingly, the 13 female fertile strains in cluster 3 were found in sympatry in two localities in the Northern mountainous area in which strains from both mating types were present, confirming the likeliness of occurrence of sexual reproduction in this cluster. In the first locality (Dien Bien Dong district, Dien Bien province), 9 out of the 16 strains sampled (10 MAT1, 6 MAT2) were female fertile; in the second locality (one field in the Bac Ha district, Lao Cai province), 4 out of the 7 strains sampled (6 MAT1, 1 MAT2) were female fertile. To conclude, the possibility of sexual reproduction was supported by genetic data (low values of linkage disequilibrium) for cluster 1 and 6, and also by biological data (presence of both mating types and of female fertile strains) for cluster 3 at least in some restricted areas of the Northern mountainous area. Conversely, clusters 2, 4 and 5 might be mainly clonal.

4. Discussion

The present study was based on SSR genotyping of a comprehensive sample of 609 rice blast strains covering three main areas of Vietnam: Northern mountainous area, Red River Delta and Central region. The sampled area covered a variety of geographic and climatic conditions (Beck et al., 2018) (Fig. S4). Seasonality is marked in the three studied areas. Northern mountainous area and Red River Delta are characterized by the cultivation of highly diverse rice varieties, including both traditional and modern, *indica* and *japonica* rice varieties, and traditional farming systems are the most frequent in almost Northern mountainous area. Fungicide is applied strongly in Red River Delta but rarely in Northern mountainous area. In the Central region, both intensive and traditional farming systems are found, cultivated rice varieties are less diverse, mostly from *indica* variety type and of modern type, since farmers here mainly follow governmental instructions of the local agricultural development sections. However, many fungal strains were also sampled from unknown rice varieties in this region. Therefore, the maintenance of an elevated genetic diversity of host in the Northern mountainous area seems to be associated with the maintenance of an elevated genetic diversity of rice blast .

The global *P. oryzae* rice-infecting population in Vietnam analyzed in this work was genetically diversified, with both high genetic and genotypic diversities. This agreed with previous results obtained by (Saleh et al., 2012, 2014) showing elevated genetic diversity of blast population in South-East Asia using the same set of molecular markers. The highest genetic diversity was

encountered in the Northern mountainous area (where diversity of rice varieties is the highest with a predominance of traditional), followed by Central region and Red River Delta.

We inferred that the *P. oryzae* rice-infecting population in Vietnam was subdivided into six significantly differentiated genetic clusters. Although they were found in sympatry in the three main geographical regions explored and throughout Vietnam (Fig. 1), strains of cluster 1 dominated in Red River Delta, and strains from cluster 1, 2 and 5 dominated in Central region. The highest pairwise F_{ST} value (0.68) was observed between clusters 6 and 1 that did not geographically overlap. Our results were coherent with previous studies by (Saleh et al., 2012, 2014) who identified four clusters found in sympatry in most regions of Asia, three out of which migrated outside of Asia in specific regions of the world. The colocalization of differentiated genetic lineages in South Asia was confirmed by population genomics data (Gladieux et al., 2018b; Thierry et al., 2021). Conversely to these results, another study conducted in Vietnam identified five genetic groups, one in the South and four in the North, by fingerprinting 78 strains with the MAGGY probe (Don et al., 1999). The very small sample size and low resolution of genetic markers could explain the discrepancies between Don's study and ours. Our results also showed significant uneven distributions of strains within the six clusters according to geographic location, rice variety type and traditional or modern variety type (Tables 4 and 5), meaning that all factors conjugate to shape the genetic diversity and structure of blast population in Vietnam. The resulting picture is a patchy distribution of the six clusters (Fig. 1). The highly diversified rice blast population in Vietnam likely reflects the elevated capability of blast fungus to adapt to heterogeneous environmental conditions. This heterogeneous feature and the detected recombination signals will be key factors to choose methods for controlling this pathogen.

Strains from cluster 3 in this study were clearly related to worldwide lineage 1. This cluster was the most diverse one and seemed to exhibit signals of recombination with low linkage disequilibrium value, both mating types (although the ratio was biased towards MAT1) and female fertile strains found in sympatry at least in two localities. This agrees with former results by (Gladieux et al., 2018b; Thierry et al., 2021) identifying worldwide lineage 1 as a recombinant lineage. Cluster 3 dominated in the Northern mountainous area (Fig. 1), and was significantly associated with traditional *japonica* rice varieties. Altogether, these results agree with previous studies performed at the worldwide scale (Tharreau et al., 2009; Thierry et al., 2021) showing that female fertile strains and recombining populations (almost all included in worldwide lineage 1) were

recovered in hillside and upland areas where traditional varieties are grown under rainfed water regime. Cluster 6 in our study was related to worldwide lineage 4 and also dominated in Northern areas and on traditional *japonica* varieties. The other clusters inferred in our study were mainly found on *indica* rice subspecies (Table 4), and were probably mainly clonal. The fact that the strains coming from *japonica* and *indica* varieties mainly segregated into different clusters was in accordance with (Liao et al., 2016): using the same set of SSR markers, these authors showed that strains sampled on *japonica* and *indica* traditional rice varieties in the YuanYang Terraces (Yunnan province of China) clustered into two genetically differentiated groups. (Thuan et al., 2006) also found similar results by characterizing 12 genetic groups from 123 strains collected in Red River Delta with AFLP markers, strains sampled on *indica* hosts being separated from *japonica* hosts. Clusters 2 and 4 in this study were related to the previously described worldwide clonal lineages 3 and 2, respectively. As expected according to previous results (Gladieux et al., 2018b; Saleh et al., 2012, 2014; Thierry et al., 2021), linkage disequilibrium inferred in these clusters, paucity of fertile strains and predominance of a single mating type indicated that clonal reproduction likely dominated in these clusters. Clusters 1 and 5 in this study were endemic to Vietnam.

Altogether our findings showed that blast population structure in Vietnam was shaped by a combination of factors (geography, rice variety group and variety type) that resulted in a patchy distribution of the six inferred genetic clusters over the country. Both sexual and clonal reproduction occurs in *P. oryzae* population in Vietnam. The endemicity of this blast pathogen existed in two clusters, 1 and 5, in which many strains contributed from all three studied regions. Taking into account this heterogeneity will be key for screening new sources of resistance to blast in Vietnamese rice germplasm.

Declaration of Competing Interest

The authors declare no conflicts of interest

Author Contributions

G.T.H. and E.F. directed the research, designed experiments and acquired funding; L.T.L, L.T.H., H.A., M.L., X.H.P., E.F. and G.T.H. contributed to field sampling; L.T.L., L.T.H., G.T.H., H.A., H.T.N. and J.M. performed phenotyping and genotyping experiments; L.T.L. and E.F. performed the data analyses;

L.T.L wrote the manuscript with input from E.F., G.T.H., D.T., M.L. and X.H.P. All authors edited and agreed on the final manuscript.

Acknowledgements

This work was supported by grants from the SPE department of INRAE, the BIOS department of CIRAD, the French Embassy in Vietnam, and the RICE CRP. Genotyping data used in this study were produced through the GenSeq technical facilities of the "Institut des Sciences de l'Evolution de Montpellier" with the support of LabEx CeMEB, an ANR "Investissements d'avenir" program (ANR-10-LABX-04-01).

Research data

The SSR dataset supporting the results of this article has been deposited as a downloadable CSV file in DRYAD: https://datadryad.org/stash/dataset/doi:10.5061/dryad.dv41ns236.

References

Adreit, H., Santoso Andriantsimialona D., Utami, D. W., Notteghem, J. L., Lebrun, M. H. and Tharreau, D. 2007. Microsatellite markers for population studies of the rice blast fungus, *Magnaporthe grisea*. Mol. Ecol. Notes 7, 667–670. http://doi.org/10.1111/j.1471-8286.2006.01672.x.

Agapow, P-M. and Burt, A. 2001. Indices of multilocus linkage disequilibrium. Mol. Ecol. Notes 1, 101–102. http://doi.org/doi: 10.1046/j.1471-8278.2000.00014.x.

Ashkani, S., Rafii, M.Y., Shabanimofrad, M., Ghasemzadeh, A., Ravanfar, S.A. and Latif, M.A. 2014. Molecular progress on the mapping and cloning of functional genes for blast disease in rice (*Oryza sativa* L.): current status and future considerations. Crit. Rev. Biotechnol. 36(2), 353–367. http://doi.org/doi: 10.3109/07388551.2014.961403.

Ballini, E., Morel, J.B., Droc, G., Price, A., Courtois, B., Notteghem, J.L., and Tharreau, D. 2008. A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. Mol. Plant Microbe Interact. 21(7), 859–68. http://doi.org/doi: 10.1094/MPMI-21-7-0859.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A. and Wood, E.F. 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Nature Scientific Data 5, 180214. http://doi.org/doi: 10.1038/sdata.2018.214.

Chiapello, H., Mallet, L., Guérin, C., Aguileta, G., Amselem, J., Kroj, T., Ortega-Abboud, E., Lebrun, M-H., Henrissat, B., Gendrault, A., Rodolphe, F., Tharreau, D. and Fournier, E. 2015. Deciphering genome content and evolutionary realtionships of strains from the fungus *Magnaporthe oryzae* attacking different hosts. Genome Biol. Evol. 7, 2896–2912. http://doi.org/doi: 10.1093/gbe/evv187.

Couch, B.C., Fudal, I., Lebrun, M.H., Tharreau, D., Valent, B., Kim, P.V., Nottéghem, J.L. and Kohn, L.M. 2005. Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. Genetics 170(2), 613–630. doi:http://doi.org/10.1534/genetics.105.041780.

De Meeûs, T. and Balloux, F. 2005. Clonal reproduction and linkage disequilibrium in diploids: a simulation study. Infect. Genet. Evol. Infection Genetics and Evolution 4(4, 345–351. http://doi.org/10.1016/j.meegid.2004.05.002.

Dean, R., Van Kan, J., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Span, P. D., Rudd, J. J., Dickman, M., Kahmann, R., Ellis, J., and Foster, G. D. 2012. The Top 10 fungal pathogens in molecular plant pathology. Mol. Plant Pathol. 13(7), 804. http://doi.org/10.1111/j.1364-3703.2012.00822.x.

Don, L.D., Tosa, Y., Nakayashiki, H. and Mayama, S. 1999. Population structure of the rice blast fungus in Vietnam. Ann. Phytopathol. Soc. Jpn. 65, 475–479. http://doi.org/10.3186/jjphytopath.65.15.

Dray, S. and Dufour, A. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. J. Stat. Software 22(4), 1–20. http://doi.org/10.18637/jss.v022.i04.

Ebbole, D.J. 2007. *Magnaporthe* as a model for understanding host-pathogen interactions. Annu. Rev. Phytopathol. 45, 437–56. http://doi.org/10.1146/annurev.phyto.45.062806.094346.

Excoffier, L., Smouse, P.E. and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131, 479–491. http://doi.org/10.1093/genetics/131.2.479.

Flor, H.H. 1971. Current status of the gene-for-gene concept. Ann. Rev. Phytopathology 9, 275–296. http://doi.org/10.1146/annurev.py.09.090171.001423.

Frichot E., Mathieu F., Trouillon T., Bouchard G., François O. (2014) Fast and efficient estimation of individual ancestry coefficients. Genetics 196, 973–983. https://doi.org/10.1534/genetics.113.160572.

Garnier-Géré, P. and Chikhi, L. 2013. Population subdivision, Hardy–Weinberg equilibrium and the Wahlund Effecteffect. eLS. http://doi.org/10.1002/9780470015902.a0005446.pub3.

Gladieux, P., Condon, B., Ravel, S., Soanes, D., Maciel, J.L.N., Nhani, A. Jr, Chen, L., Terauchi, R., Lebrun, M-H., Tharreau, D., Mitchell, T., Pedley, K.F., Valent, B., Talbot, N.J., Farman, M. and Fournier, E. 2018a. Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. mBio 9(1), e01219-17. http://doi.org/10.1128/mBio.01219-17.

Gladieux, P., Ravel, S., Rieux, A., Cros-Arteil, S., Adreit, H., Milazzo, J., Thierry, M., Fournier, E., Terauchi, R. and and Tharreau, D. 2018b. Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. mBio 9(2), e01806-17. http://doi.org/10.1128/mBio.01806-17.

Higgins, J., Santos, B., Khanh, T.D. et al., 2021. Resequencing of 672 Native Rice Accessions to Explore Genetic Diversity and Trait Associations in Vietnam. Rice 14, 52. http://doi.org/10.1186/s12284-021-00481-0.

Huang, J., Si, W., Deng, Q., Li P., and Yang, S. 2014. Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. BMC Genetics 15, 45. http://doi.org/0.1186/1471-2156-15-45.

Hu Z-J, Huang Y-Y, Lin X-Y, Feng H, Zhou S-X, Xie Y, Liu X-X, Liu C, Zhao R-M, Zhao W-S, Feng C-H, Pu M, Ji Y-P, Hu X-H, Li G-B, Zhao J-H, Zhao Z-X, Wang H, Zhang J-W, Fan J, Li Y, Peng Y-L, He M, Li D-Q, Huang F, Peng Y-L and Wang W-M.2022. Loss and Natural Variations of Blast Fungal Avirulence Genes Breakdown Rice Resistance Genes in the Sichuan Basin of China. *Front. Plant Sci.* 13:788876. doi: 10.3389/fpls.2022.788876

Hue, H.T., Nghia, L.T., Minh, H.T., Anh, L.H., Trang, L.T.T. and Khanh, T.D. 2018. Evaluation of genetic diversity of local-colored rice landraces using SSR Markers. Nat. Sci. 67, 24–34. http://doi.org/10.18052/www.scipress.com/ILNS.67.24.

Jombart, T. and Ahmed, I. 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27(21), 3070-1. http://doi.org/10.1093/bioinformatics/btr521.

Jombart, T., Devillard, S. and Balloux, F. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics 11(1), 94. http://doi.org/10.1186/1471-2156-11-94.

557 Kamvar, Z.N., Tabima, J.F. and Grünwald, N.J. 2014. Poppr: An R package for genetic analysis of
558 populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2, e281.
559 http://doi.org/doi:10.7717/peerj.281.
560 Kanamori, M., Kato, H., Yasuda, N., Koizumi, S.L., Peever, T.L., Kamakura, T., Teraoka, T. and Arie, T.
561 2007. Novel mating type-dependent transcripts at the mating type locus in *Magnaporthe oryzae*.
562 Gene 403 (1-2), 6–17. http://doi.org/10.1016/j.gene.2007.06.015.
563 Kang, S., Chumley, F.G. and Valent, B. 1994. Isolation of the mating-type genes of the
564 phytopathogenic fungus *Magnaporthe grisea* using genomic subtraction. Genetics 138, 289–96.
565 http://doi.org/10.1093/genetics/138.2.289.
566 Katayama, N., Baba, Y.G., Kusumoto, Y. and Tanaka, K. 2015. A review of post-war changes in rice
567 farming and biodiversity in Japan. Agricultural Systems 132, 73–84.
568 http://doi.org/10.1016/j.agsy.2014.09.001.
569 Khush, G.S. 2005. What it will take to feed 5.0 billion rice consumers in 2030. Plant Mol. Biol. 59,
570 1–6. http://doi.org/10.1007/s11103-005-2159-5.
571 Kim, K-T., Ko, J., Song, H., Choi, G., Kim, H., Jeon, J., Cheong, K., Kang, S. and Lee, Y-H. 2019. Evolution
572 of the genes encoding effector candidates within multiple pathotypes of *Magnaporthe oryzae*.
573 Front. Microbiol. 10, 2575. http://doi.org/10.3389/fmicb.2019.02575.
574 Lang, N.T., Luy, T.T., Ha, P.T.T., and Buu, B.C. 2010. Genotype analysis to blast disease in rice (*Oryza*
575 *sativa L*.) in Mekong delta. OmonRice 17, 87–98.
576 Latif, M.A., Rahman, M.M., Kabir, M.S., Ali, M.A., Islam, M.T. and Rafii, M.Y. 2011. Genetic diversity
577 analyzed by quantitative traits among rice (*Oryza sativa L*.) genotypes resistant to blast disease.
578 Afr. J. Microbiol. Res. 5(25), 4383–4391. http://doi.org/doi:10.5897/AJMR11.492.
579 Latorre, S., Reyes-Avila, C.S., Malmgren, A., Win, J., Kamoun, S. and Burbano, H.A. 2020. Differential
580 loss of effector genes in three recently expanded pandemic clonal lineages of the rice blast
581 fungus. BMC Biology 18, 88. http://doi.org/10.1186/s12915-020-00818-z.
582 Le, M.T., Arie, T. and Teraoka, T. 2010. Population dynamics and pathogenic races of rice blast
583 fungus, *Magnaporthe oryzae* in the Mekong Delta in Vietnam. J. Gen. Plant Pathol. 76, 177–182.
584 http://doi.org/10.1007/s10327-010-0231-8.
585 Liao, J., Huang, H., Meusnier, I., Adreit, H., Ducasse, A., Bonnot, F., Pan, L., He, X., Kroj, T., Fournier,
586 E., Tharreau, D., Gladieux, P. and Morel, J.B. 2016. Pathogen effectors and plant immunity

587 determine specialization of the blast fungus to rice subspecies. eLife 5, e19377.
588 http://doi.org/10.7554/eLife.19377.
589 Liu, W., Liu, J., Triplett, L., Leach, J.E. and Wang, G-L. 2014. Novel insights into rice innate immunity
590 against bacterial and fungal pathogens. Annu. Rev. Phytopathol. 52, 213–241.
591 http://doi.org/10.1146/annurev-phyto-102313-045926.
592 Manoj, K.Y., Aravindan, S., Umakanta, N., Raghu, S., Prabhukarthikeyan, S.R., Keerthana, U., Marndi,
593 B.C., Totan, A., Susmita, M., Rupesh, D., Pramesh, I.D., Sanghamitra, S. and Rath, P.C. 2019. Blast
594 resistance in Indian rice landraces: Genetic dissection by gene specific markers. PloS One 14(3),
595 e0213566. http://doi.org/10.1371/journal.pone.0213566.
596 Nakagahra, M., Miyazaki, S. and Vaughan, D.A. 2000. Challenge to the crisis of the earth's biosphere
597 in the 21st century. Genebank Manag. of Crop Genet. Res., pp. 130–150.
598 Nguyen, T.T.T., Nguyen, T.M.N., Hoang, H.L., Naruto, F. and Kenichi, T. 2012. Genetic diversity in
599 Vietnamese upland rice germplasm revealed by SSR markers. J. Fac. Agr. Kyushu Univ. 57(2),
600 :383–391. http://doi.org/10.5109/25195.
601
602
603 Noguchi, M.T., Yasuda, N., and Fujita, Y. 2006. Evidence of genetic exchange by parasexual
604 recombination and genetic analysis of pathogenicity and mating type of parasexual recombinants
605 in rice blast fungus, *Magnaporthe oryzae*. *Phytopathology* R 96, 746–750. doi: 10.1094/PHYTO-
606 96-0746.
607 Noguchi, M.T. 2011. Parasexual recombination in *Magnaporthe oryzae*. *Japan Agricult. Res. Q.: JARQ*
608 45, 39–45. doi: 10.6090/jarq.45.39.
609 Orasen, G., Greco, R., Puja, E., Pozzi, C. and Stile, M.R. 2020. Blast resistance R genes pyramiding in
610 temperate japonica rice. Euphytica 216, 40. http://doi.org/doi:10.1007/s10681-020-2575-2.
611 Paradis, E. and Schliep, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary
612 analyses in R. Bioinformatics 35, 526–528. http://doi.org/10.1093/bioinformatics/bty633.
613 Perrier X., Jacquemoud-Collet J.P. (2006) DARwin Software. http://darwin.cirad.fr/darwin.
614 Saleh, D., Milazzo, J., Adreit, H., Fournier, E. and Tharreau, D. 2014. South-East Asia is the center of
615 origin, diversity and dispersion of the rice blast fungus, *Magnaporthe oryzae*. New. Phytol. 201,
616 1440–1456. http://doi.org/10.1111/nph.12627.

617 Saleh, D., Xu, P., Shen, Y., Li, C., Adreit, H., Milazzo, J., Ravigné, V., Bazin, E., Nottéghem, J.L., Fournier,
618 E. and Tharreau D. 2012. Sex at the origin: An Asian population of the rice blast fungus
619 *Magnaporthe oryzae* reproduces sexually. Mol. Ecol. 21, 1330–1344.
620 http://doi.org/10.1111/j.1365-294X.2012.05469.x.
621 Savary, S., Willocquet, L., Elazegui, F.A., Castilla, N.P., and Teng, P.S. 2000. Rice pest constraints in
622 tropical Asia: quantification of yield losses due to rice pests in a range of production situations.
623 Plant Dis. 84, 357–369. http://doi.org/doi: 10.1094/PDIS.2000.84.3.357.
624 Selisana, S.M., Yanoria, M.J., Quime, B., Chaipanya, C., Lu, G., Opulencia, R., 2017. Avirulence
625 (*AVR*) gene-based diagnosis complements existing pathogen surveillance tools for effective
626 deployment of resistance (*r*) genes against rice blast disease. *Phytopathology* 107, 711–720. doi:
627 10.1094/PHYTO-12-16-0451-R.
628 Silué, D. and Nottéghem, J.L. 1990. Production of perithecia of *Magnaporthe oryzae* on rice plants.
629 Mycol. Res. 94(4), 1151–1152. http://doi.org/10.1016/S0953-7562(09)81351-8.
630 Silué, D., Nottéghem, J.L. and Tharreau, D. 1992. Evidence of gene-for-gene relationship in the *Oryza*
631 *sativa* - *Magnaporthe grisea* pathosystem. Phytopathology 82, 577–580.
632 http://doi.org/10.1094/Phyto-82-577.
633 Smith, J.C., Ghosh, A. and Hijmans, R.J. 2019. Agricultural intensification was associated with crop
634 diversification in India (1947-2014). PLoS ONE 14(12), e0225555.
635 http://doi.org/10.1371/journal.pone.0225555.
636 Sweigard, J.A., Orbach, M.J., Valent, B. and Chumley, F.G. 1990. A miniprep procedure for isolating
637 genomic DNA from *Magnaporthe grisea*. Fung. Genet. Newsletter 37, 28.
638 http://doi.org/10.4148/1941-4765.1493.
639 Talbot, N.J. 2003. On the trail of a cereal killer: Exploring the Biology of *Magnaporthe grisea*. Annual.
640 Rev. Microbiol. 57, 177–202. http://doi.org/10.1146/annurev.micro.57.030502.090957.
641 Tharreau, D., Fudal, I., Santoso Andriantsimialona, D., Utami D., Fournier E., Lebrun M-H., and
642 Notteghem J-L. 2009. World population structure and migration of the rice blast fungus,
643 *Magnaporthe oryzae*. In: Wang GL., Valent B. (edsEds), Advances in genetics, genomics and
644 control of rice blast disease. Springer, Dordrecht, pp. 209–215. http://doi.org/10.1007/978-1-
645 4020-9500-9_21.
646 Thierry, M., Milazzo, J., Adreit, H., Ravel, S., Borron, S., Sella, V., Ioos, R., Fournier, E., Tharreau, D.
647 and Gladieux, P. 2021. Ecological differentiation and incipient speciation in the fungal pathogen

causing rice blast. bioRxiv 2020.06.02.129296. PLOS Pathogens, in press. https://doi.org/10.1101/2020.06.02.129296.

Thuan, N.T.N., Bigirimana, J., Roumen, E., Van, D.D. and Hofte, M. 2006. Molecular and pathotype analysis of the rice blast fungus in North Vietnam. Europ. J. Plant. Pathol. 114, 381–396. http://doi.org/10.1007/s10658-006-0002-8.

Valent, B., 1990. Rice blast as a model system for plant pathology. Phytopathology 80, 33–36. http://doi.org/doi:10.1094/Phyto-80-33.

Valent, B., Crawford, M.S., Weaver, C.G. and Chumley, F.G. 1986. Genetic studies of fertility and pathogenicity in *Magnaporthe grisea*. Iowa State J. Res. 60, 569–594.

Vergne, E., Grand, X., Ballini, E., Chalvon, V., Saindrenan, P., Tharreau, D., Nottéghem, J.L. and Morel, J.B. 2010. Preformed expression of defense is a hallmark of partial resistance to rice blast fungal pathogen *Magnaporthe oryzae*. BMC Plant Biol. 10, 206. doi: http://doi.org/10.1186/1471-2229-10-206.

Vu, T.T.H., Nguyen, T.T.H., Tran, D.K., Khuat, H.T. and Nakamura, C. 2016. Genetic diversity of Vietnamese lowland rice germplasms as revealed by SSR markers in relation to seedling vigour under submergence. Biotechnology & Biotechnological Equipment 30(1), 17–25. http://doi.org/10.1080/13102818.2015.1085330.

Ward, J.H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58(301), 236–244. http://doi.org/10.1080/01621459.1963.10500845.

Wilson, R.A. and Talbot, N.J. 2009. Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. Nat. Rev. Microbiol. 7(3), 185–195. http://doi.org/10.1038/nrmicro2032.

Xu, J-R. and Hamer, J.E. 1995. Assessment of *Magnaporthe grisea* mating type by spore PCR. Fung. Genet. Newsletter 42, 80–81. doi: http://doi.org/10.4148/1941-4765.1358.

Zhang, S. and Xu J-R. 2014. Effectors and effector delivery in *Magnaporthe oryzae*. PLoS Pathogens 10(1), e1003826. http://doi.org/10.1371/journal.ppat.1003826.

Zhong, Z., Chen, M., Lin, L., Han, Y., Bao, J., Tang, W., Lin, L., Lin, Y., Somai, R., Lu, L., Zhang, W., Chen, J., Hong, Y., Chen, X., Wang, B., Shen, W., Lu, G., Norvienyeku, J., Ebbole, J.D. and Wang, Z. 2018. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. The ISME Journal 12, 1867–1878. http://doi.org/10.1038/s41396-018-0100-6.

678 **Tables**

679 **Table 1. Diversity indices of *Pyricularia oryzae* in the whole population and in the three different**
680 **geographical regions (Northern mountainous area, Red River Delta, Central region).**

681

	N	MLG	eMLG (± SE)	H	G	λ	E.5	H_e
Northern mountainous area	263	208	117 (± 3.14)	5.21	148.1	0.993	0.808	0.675
Red River Delta	136	90	90 (± 0)	4.25	49.7	0.980	0.707	0.417
Central region	210	161	112 (± 2.96)	4.95	114.2	0.991	0.809	0.585
Whole population	**609**	**447**	**134 (± 3.74)**	**5.88**	**238.5**	**0.996**	**0.663**	**0.586**

682

683 N: number of strains genotyped. MLG: number of different multilocus genotypes. eMLG: number of
684 MLG expected at the lowest common sample size, estimated using rarefaction analysis (SE: standard
685 error for the rarefaction analysis). H: Shannon-Wiener index. G: Stoddart and Taylor's index. λ:
686 Simpson's index of genotypic diversity. E.5: Evenness of MLG. H_e: Nei's gene diversity (expected
687 heterozygosity).

688 **Table 2. Genetic diversity within each of the six genetic clusters.**

	N	MLG	eMLG (± SE)	H	G	λ	E.5	*He*	MLG/N
Cluster 1	232	135	24.1 (± 1.80)	4.53	55.6	0.982	0.597	0.258	58.2%
Cluster 2 (WL3)	120	106	27.1 (± 0.90)	4.61	92.3	0.989	0.918	0.429	88.3%
Cluster 3 (WL1)	80	74	27.2 (± 0.82)	4.27	68.1	0.985	0.950	0.634	92.5%
Cluster 4 (WL2)	28	21	21.0 (± 0)	2.95	17.0	0.941	0.888	0.311	75%
Cluster 5	103	67	23.9 (± 1.67)	4.01	44.0	0.977	0.791	0.408	65%
Cluster 6 (WL4)	46	44	27.3 (± 0.67)	3.77	42.3	0.976	0.977	0.374	95.6%

689 N: number of strains genotyped. MLG: number of different multilocus genotypes. eMLG: number of
690 MLG expected at the lowest common sample size, estimated using rarefaction analysis (SE: standard
691 error for the rarefaction analysis). H: Shannon-Wiener index. G: Stoddart and Taylor's index. λ:
692 Simpson's index of genotypic diversity. E.5: Evenness of MLG. *He*: Nei's gene diversity (expected
693 heterozygosity).

694 **Table 3. Pairwise F_{ST} values between the six genetic clusters estimated by AMOVA.** Associated P-
695 values are all below 0.01, indicating that F_{ST} values are significantly different from 0.

Cluster	2	3	4	5	6
1	0.43	0.53	0.65	0.54	0.68
2		0.30	0.49	0.31	0.54
3			0.47	0.35	0.33
4				0.53	0.65
5					0.57

696

Table 4. Distribution of *P. oryzae* genotypes into the six genetic clusters, according to geographic region, rice variety group, and variety type of original host plant. Results of χ^2 tests, performed to address random distribution of genotypes into clusters, are indicated for three factors.

		DAPC clusters					
		1	**2**	**3**	**4**	**5**	**6**
Region	Central region	66	51	4	9	76	4
	Red River Delta	102	6	2	7	4	15
	Northern mountainous area	64	63	74	12	23	27
χ^2=250.2, P<10^{-4}, df=10							
Rice group and variety type	Indica - Modern	167	82	9	15	92	2
	Indica - Traditional	25	9	24	6	11	0
	Japonica - Modern	25	4	0	7	0	13
	Japonica - Traditional	15	25	47	0	0	31
χ^2=325.54, P<10^{-4}, df=15							
Total		232	120	80	28	103	46

Table 5. Hierarchical analysis of molecular variance (AMOVA) performed for each factor (geographic region, rice variety group and variety type of original host plant). The 447 genotypes were grouped by DAPC genetic clusters then by factor. All associated probability below 10^{-3} (statistical significance tested after 999 permutations).

	Percentage of variation		
	among clusters	**among level of factor within clusters**	**within individuals**
Region	39.8%	9.25%	50.89%
Rice subspecies	35.3%	12.47%	52.23%
Variety type	39.77%	6.62%	53.61%

707 **Table 6. Index association $\bar{r}_D$ and numbers of female fertile, Mat1and Mat2 strains in each region and each genetic cluster.** The probability
708 that $\bar{r}_D$ is significantly different from 0, indicated between brackets, was inferred after 999 permutations.

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
$\bar{r}_D$ of all strains		0.037 (P=0.001)	0.079 (P=0.001)	0.106 (P=0.001)	0.265 (P=0.001)	0.124 (P=0.001)	0.007 (P=0.082)
$\bar{r}_D$ of clone-corrected		-0.001 (P=0.282)	0.076 (P=0.001)	0.097 (P=0.001)	0.224 (P=0.001)	0.091 (P=0.001)	0.004 (P=0.123)
Central region	Female fertile/Mat1	0/1	-	0/3	0/8	-	0/4
	Female fertile /Mat2	1/65	1/51	0/1	0/1	0/76	-
	$\bar{r}_D$ (all strains)	0.022 (P=0.066)	0.122 (P=0.001)	0.381 (P=0.001)	0.583 (P=0.001)	0.190 (P=0.001)	-
	$\bar{r}_D$ (clone-corrected)	-0.001 (P=0.455)	0.115 (P=0.001)	0.551 (P=1.000)	0.022 (P=0.116)	0.147 (P=0.001)	-
Red River Delta	Female fertile/Mat1	0/2	-	0/1	0/7	0/1	0/13
	Female fertile /Mat2	1/100	0/6	0/1	-	0/3	0/2
	$\bar{r}_D$ (all strains)	0.049 (P = 0.001)	0.204 (P=0.057)	-	0.351 (P=0.001)	-	0.148 (P=0.001)
	$\bar{r}_D$ (clone-corrected)	0.012 (P=0.126)	0.204 (P=0.041)	-	0.335 (P=0.015)	-	0.137 (P=0.002)
Northern mountainous area	Female fertile/Mat1	0/1	0/2	8/66	0/12	0/1	0/25
	Female fertile /Mat2	0/63	0/61	5/8	-	0/22	0/2
	$\bar{r}_D$ (of all strains)	0.116 (P=0.001)	0.089 (P=0.001)	0.118 (P=0.001)	- 0.016 (P=0.625)	0.277 (P=0.001)	0.017 (P=0.122)
	$\bar{r}_D$ (clone-corrected)	0.065 (P=0.001)	0.083 (P=0.001)	0.107 (P=0.001)	- 0.075 (P=0.709)	0.145 (P=0.003)	0.007 (P=0.259)

709

710 **Supplementary tables**

711 **Table S1. Names of provinces and districts where samples were collected in three geographical**
712 **regions.**

Region	Province	District
Central region	Binh Dinh	An Nhon, Hoai Nhon, Phu My
	Gia Lai	Chu Pah, Pleiku
	Ha Tinh	Hong Linh
	Hue	Huong Tra, Phu Loc
	Khanh Hoa	Phu Yen
	Kon Tum	Vinh Quang
	Phu Yen	Tuy An
	Quang Binh	Bo Trach, Quang Ninh
	Quang Nam	Dai Loc, Duy Xuyen, Que Son
	Quang Ngai	Binh Son, Son Tinh
	Thanh Hoa	Tinh Gia, Tp. Thanh Hoa
Northern mountainous area	Bac Giang	Viet Yen, Xuong Giang
	Bac Kan	Bach Thong, Cho Don, Ngan Son
	Cao Bang	Hoa An, Thach An
	Dien Bien	Dien Bien Dong
	Ha Giang	Bac Quang, Vi Xuyen
	Lai Chau	Tam Duong
	Lao Cai	Bac Ha, Bat Xat, Sa Pa
	Phu Tho	Doan Hung
	Son La	Quynh Nhai
	Thai Nguyen	Dai Tu, Phu Luong
	Tuyen Quang	Son Duong, Thai Hoa, Tuyen Quang
	Yen Bai	Van Chan
Red River Delta	Bac Ninh	Tu Son
	Ha Nam	Binh Luc, Duy Tien

	Hai Duong	Kim Thanh, Nam Sach
	Hai Phong	An Lao, Vinh Bao
	Hung Yen	Kim Dong
	Nam Dinh	Giao Thuy, Vu Ban
	Ninh Binh	Kim Khanh, Kim Son
	Quang Ninh	Tien Yen, Uong Bi
	Thai Binh	Dong Hung, Quynh Phu
	Vinh Phuc	Phu Yen

713

Captions of figures

Main figures

Fig. 1. Map of Vietnam showing sampling locations and the wheel shape circle showing frequency of occurrence of the six genetic clusters inferred with DAPC in three regions in Vietnam. Each point in map represents one sampling location. The black points were in Northern mountainous area; the violetred points were in Red River Delta and steelblue points were in Central region. Each pie represents one cluster in one region, with cluster 1: red, cluster 2: yellow, cluster 3: green, cluster 4: turquoise blue, cluster 5: darkblue and cluster 6: pink.

Fig. 2. A. Scatterplot of coordinates of genotypes assigned to 6 genetic clusters. Each dot represents one genotype and the inertia ellipses represent the groups. **B.** Cross-entropy as a function of the number of ancestral populations (*K*) calculated using sNMF. **C.** Hierarchical agglomerative clustering with Ward's linkage of 447 genotypes using DARwin.

Fig. 3. Population subdividsion in *P. oryzae* inferred with DAPC and genealogical relationships. A. Barplots showing ancestry proportions of the 447 unique genotypes in *K* clusters, with *K* varying from 3 to 8. Each genotype is represented by a vertical bar divided into *K* segments, indicating membership in the *K* clusters. Numbering and coloring of clusters at *K*=6 remains as is throughout this study clusters (red: cluster 1, yellow: cluster 2, green: cluster 3, turquoise blue: cluster 4, dark blue: cluster 5, pink: cluster 6). **B.** Neighbor Joining tree showing genealogical relationships among individuals based on the "relative dissimilarity" distance matrix. Branches are colored according to the assignation of genotypes to the six genetic clusters defined by DAPC.

Supplementary figures

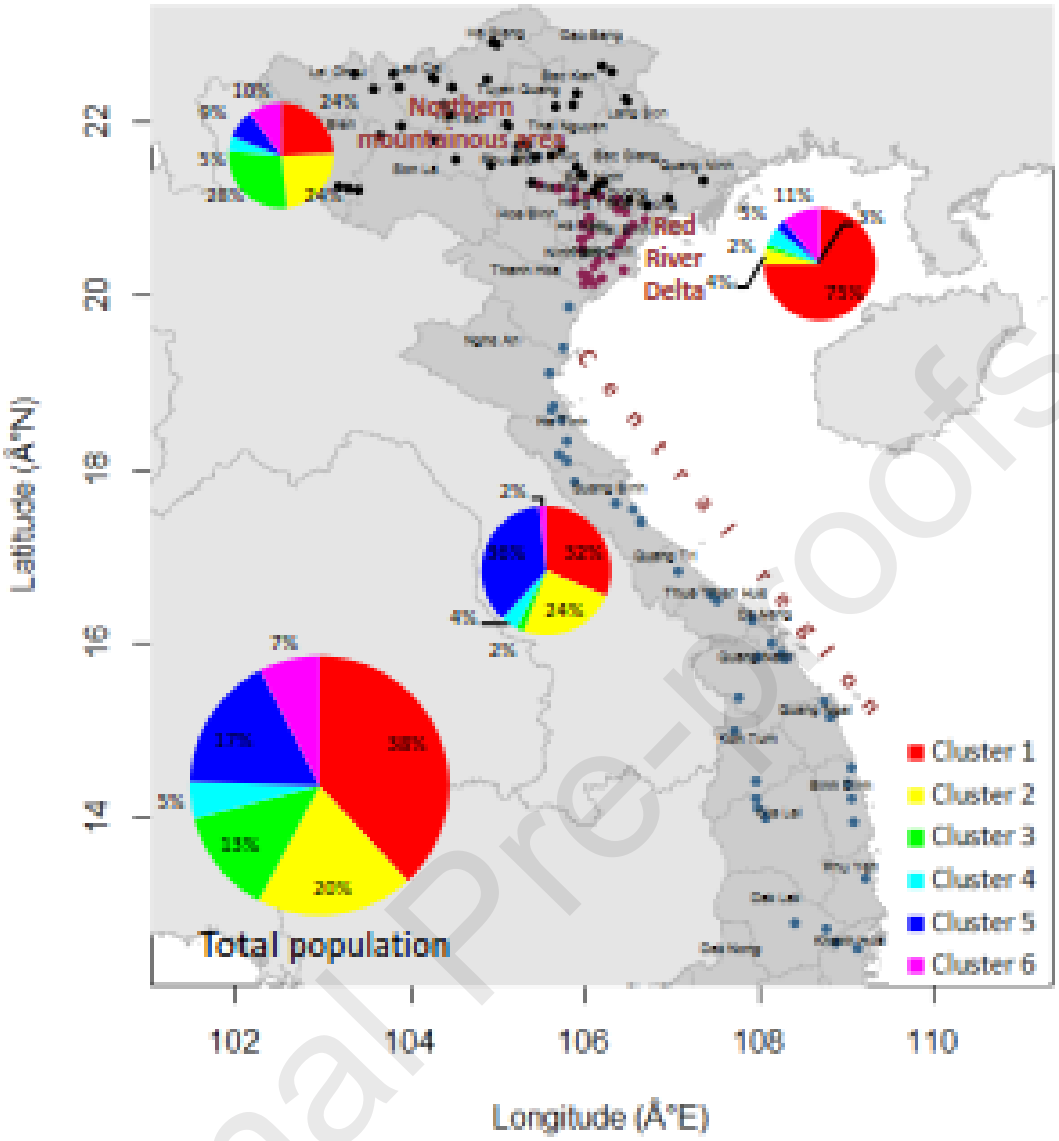
Fig. S1. Relationships between the six genetic clusters inferred in Vietnam (Fig. 3, this study) and the four rice-attacking genetic lineages already described at the worldwide scale (WL lineages). A. Barplot showing the probability of ancestry inferred from DAPC-785 (i.e. DAPC performed with the 447 Vietnamese MLGs + 338 worldwide reference MLGs) into *K*=6 ancestral clusters. The worldwide reference MLGs are framed in black with the corresponding worldwide lineage indicated above (WL1 to WL4). The colors chosen for the six clusters inferred from DAPC-785 are those of clusters

742 inferred from DAPC-447 to whom the majority of MLGs are also unambiguously assigned. Grey
743 arrows below the barplot indicate the 13 MLGs for which the two analyses are not congruent. **B.**
744 Distribution of the 447 Vietnamese MLGs assigned to the six clusters inferred by DAPC-785 (lines)
745 into the six clusters inferred by DAPC-447 (columns). Light green boxes indicated correct matches,
746 grey boxes indicate incorrect matches.

747 **Fig. S2. Interpolation of maximum ancestry coefficients in each cluster at each geographic position**
748 **throughout Vietnam.** Red: cluster 1, yellow: cluster 2, green: cluster 3, turquoise blue: cluster 4,
749 dark blue: cluster 5, pink: cluster 6. The darkest color corresponds to an ancestry coefficient of 1,
750 whereas white corresponds to an ancestry coefficient of 0.

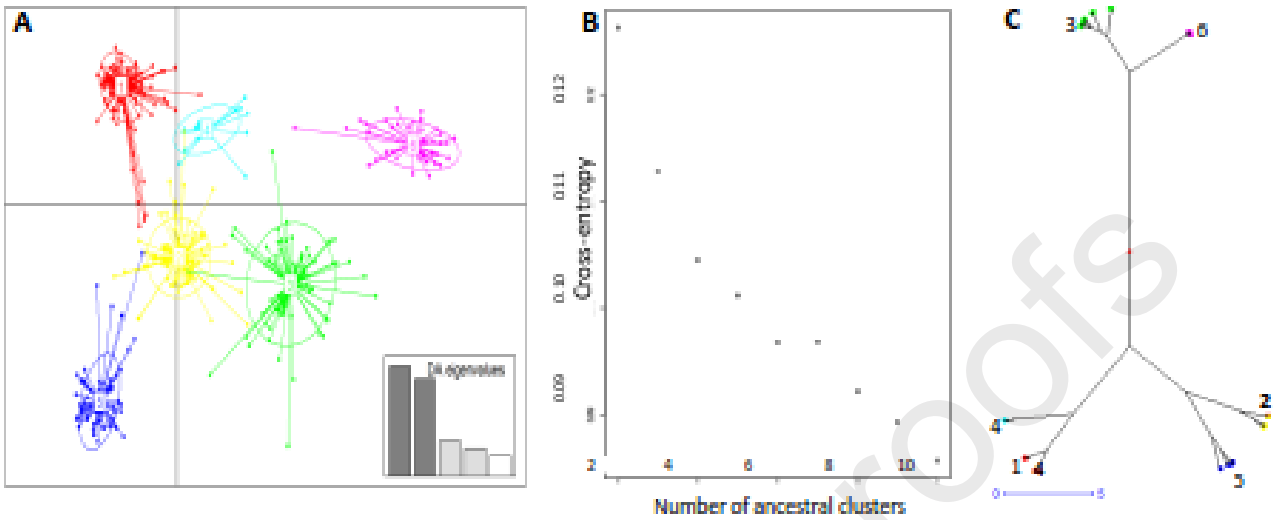
751 **Fig. S3. Distribution of $\bar{r}_D$ values in each genetic cluster after 999 randomizations of alleles among**
752 **individuals, either for the full dataset (609 genotypes) or after clone correction (447 genotypes).**
753 Dotted blue lines indicate the observed value of $\bar{r}_D$.

754 **Fig. S4. Köppen-Geiger climate classification map for Vietnam (Beck et al., 2018).** Limits of the
755 geographic regions are highlighted in grey, with the three geographic regions in which samples
756 were taken indicated in bold font.



757

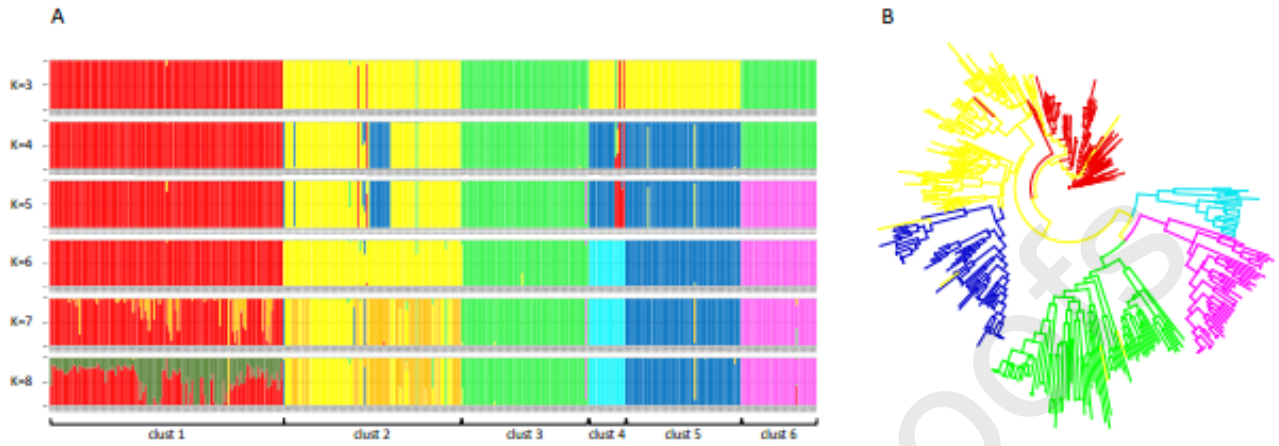
758



759

760

Lieu et al. Figure 3



761

762

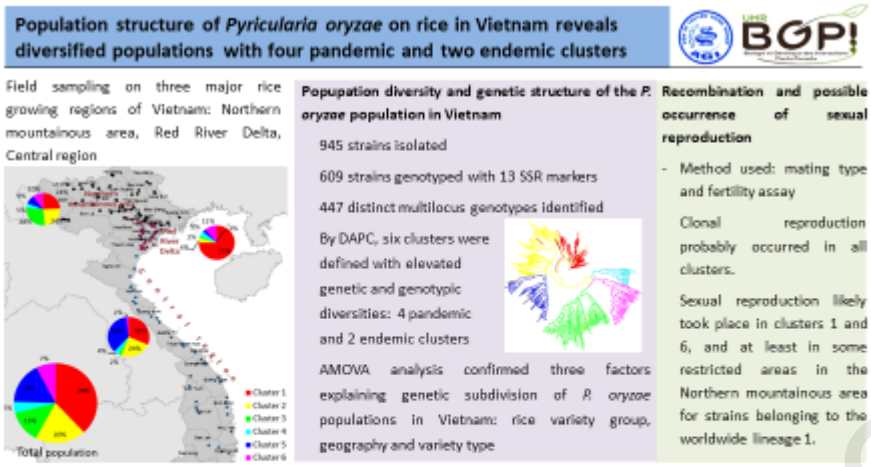
763 **Highlights**

764 • The genetic diversity of *Pyricularia oryzae* populations was elevated in Vietnam
765 • Six genetic clusters were defined: 4 pandemic and 2 endemic
766 • Blast population structure in Vietnam was shaped by a combination of three factors
767 • Mating types were unevenly distributed in the whole population and among clusters
768 • Recombination signals were detected in some restricted areas of Northern mountains

769

770

771 Graphical abstract



772

773

774 **Author statement**

775 **Hoang G.T. and Fournier E.** directed the research, designed experiments and acquired funding; **Le**
776 **L.T., Ha L.T., Adreit H., Lebrun M., X.H.P., Fournier E. and Hoang G.T.** contributed to field sampling;
777 **Le L.T., Ha L.T., Hoang G.T., Adreit H., Nguyen H.T. and Milazzo J.** performed phenotyping and
778 genotyping experiments; **Le L.T. and Elisabeth Fournier** performed the data analyses; **Le L.T.** wrote
779 the manuscript with input from **Elisabeth Fournier, Hoang G.T., Tharreau D., Lebrun M. and Pham**
780 **X.H.** All authors edited and agreed on the final manuscript.

781

782