Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*

Pierre GLADIEUX [1*], Bradford CONDON [2*], Sebastien RAVEL[3], Darren SOANES[4], Joao Leodato Nunes MACIEL[5], Antonio NHANI Jr[5], Li CHEN [2], Ryohei TERAUCHI[7], Marc-Henri LEBRUN[8], Didier THARREAU[3], Thomas MITCHELL[9], Kerry F. PEDLEY[10], Barbara VALENT[11], Nicholas J. TALBOT[4], Mark FARMAN[2], Elisabeth FOURNIER[1]

[1] INRA, UMR BGPI, Campus de Baillarguet, Montpellier F34398; [2] Department of Plant Pathology, University of Kentucky, Lexington, KY 40546; [3] CIRAD, UMR BGPI, Campus de Baillarguet, Montpellier F34398; [4]College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, UK; [5] Embrapa Wheat, Passo Fundo, Brazil, [6] Embrapa Agricultural Informatics, Campinas, Brazil ; [7] Iwate Biotechnology Research Center, Kitakami, Iwate, Japan; [8]INRA-AgroParisTech UMR BIOGER, Campus AgroparisTech, Thiverval-Grignon, F78850; [9]Department of Plant Pathology, Ohio State University, Columbus, OH 43210; [10]USDA, Agricultural Research Service, FDWSRU, Ft. Detrick, Maryland, 21702,[11]Department of Plant Pathology, Kansas State University, Manhattan, KS 66506

* These authors contributed equally to the work

Corresponding author: Elisabeth Fournier

e-mail: elisabeth.fournier@inra.fr

phone: +33 4 99 62 48 63

Abstract

Delineating species and epidemic lineages in fungal plant pathogens is critical to our understanding of disease emergence and the structure of fungal biodiversity, and also informs international regulatory decisions. *Pyricularia oryzae (syn. Magnaporthe oryzae)* is a multi-host pathogen that infects multiple grasses and cereals, is responsible for the most damaging rice disease (rice blast), and of growing concern due to the recent introduction of wheat blast to Bangladesh from South America. However, the genetic structure and evolutionary history of *M. oryzae*, including the possible existence of cryptic phylogenetic species, remain poorly defined. Here, we use whole-genome sequence information for 76 *M. oryzae* isolates sampled from 12 grass and cereal genera to infer the population structure of *M. oryzae*, and to reassess the species status of wheat-infecting populations of the fungus. Species recognition based on genealogical concordance, using published data or extracting previously-used loci from genome assemblies, failed to confirm a prior assignment of wheat blast isolates to a new species (*Pyricularia graminis tritici*). Inference of population subdivisions revealed multiple divergent lineages within *M. oryzae*, each preferentially associated with one host genus, suggesting incipient speciation following host shift or host range expansion. Analyses of gene flow, taking into account the possibility of incomplete lineage sorting, revealed that genetic exchanges have contributed to the makeup of multiple lineages within *M. oryzae*. These findings provide greater understanding of the eco-evolutionary factors that underlie the diversification of *M. oryzae* and highlight the practicality of genomic data for epidemiological surveillance in this important multi-host pathogen.

Importance

Infection of novel hosts is a major route for disease emergence by pathogenic micro-organisms. Understanding the evolutionary history of multi-host pathogens is therefore important to better predict the likely spread and emergence of new diseases. *Magnaporthe oryzae* is a multi-host fungus that causes serious cereal diseases, including the devastating rice blast disease, and wheat blast, a cause of growing concern due to its recent spread from South America to Asia. Using whole genome analysis of 76 fungal strains from different hosts, we have documented the divergence of *M. oryzae* into numerous lineages, each infecting a limited number of host species. Our analyses provide evidence that inter-lineage gene flow has contributed to the genetic makeup of multiple *M. oryzae* lineages within the same species. Plant health surveillance is therefore warranted to safeguard against disease emergence in regions where multiple lineages of the fungus are in contact with one another.

Introduction

Investigating population genetic structure in relation to life history traits such as reproductive mode, host range or drug resistance is particularly relevant in pathogens [1, 2]. Knowledge of species, lineages, populations, levels of genetic variability and reproductive mode is essential to answer questions common to all infectious diseases, such as the tempo, origin, proximate (i.e. molecular) and ultimate (eco-evolutionary) causes of disease emergence and spread [3]. Multilocus molecular typing schemes have shown that cryptic species and lineages within species are often more numerous than estimated from phenotypic data alone. Genomic approaches are emerging as a new gold standard for detecting cryptic structure or speciation with increased resolution, allowing fine-grained epidemiological surveillance and science-based regulatory decisions. The added benefit of whole genomes approaches includes identifying the genetic basis of life history traits, and better understanding of both the genomic properties that influence the process of speciation and the signatures of (potentially incomplete) speciation that are observable in patterns of genomic variability [4, 5].

Many plant pathogenic ascomycete fungi are host-specific, and some of their life history traits have been shown to be conducive to the emergence of novel pathogen species adapted to new hosts [6, 7]. Investigating population structure within multi-host ascomycetes thus offers a unique opportunity to identify the genomic features associated with recent host-range expansions or host-shifts. In this study, our model is *Magnaporthe oryzae* (synonym of *Pyricularia oryzae*) [6-8], a fungal ascomycete causing blast disease on a variety of grass hosts. *Magnaporthe oryzae* is well studied as the causal agent of the most

important disease of rice (*Oryza sativa*), but it also causes blast disease on more than 50 cultivated and wild monocot plant species [9]. This includes other cereal crops such as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), finger millet (*Eleusine coracana*), and foxtail millet (*Setaria italica*, *S. viridis*), as well as wild and cultivated grass hosts including goosegrass (*Eleusine indica*), annual ryegrass (*Lolium multiflorum*), perennial ryegrass (*L. perenne*), tall fescue (*Festuca arundinacea*) and St. Augustine grass (*Stenotaphrum secundatum*) [10]. Previous studies based on multilocus sequence typing showed that *M. oryzae* is subdivided into multiple clades, each found on only a limited number of host species, with pathogenicity testing revealing host-specificity as a plausible driver of genetic divergence [11, 12]. More recently, comparative genomics of eight isolates infecting wheat, goosegrass, rice, foxtail millet, finger millet and barley revealed deep subdivision of *M. oryzae* into three groups infecting finger millet or wheat, foxtail millet, and rice or barley [13, 14]. Subsequent analysis of genomic data from nine wheat-infecting isolates, two ryegrass infecting isolates, and one weeping lovegrass-infecting isolate subdivided lineages infecting wheat only on the one hand and wheat or ryegrass on the other hand, and revealed an additional lineage associated with the weeping lovegrass strain [15]. Together, these studies suggest a history of host-range expansion or host-shifts and limited gene flow between lineages within *M. oryzae*.

Magnaporthe oryzae isolates causing wheat blast represent a growing concern in terms of food security. This seed-borne pathogen can spread around the world through movement of seed or grain. Therefore, understanding the evolutionary origin and structure of populations causing wheat blast is a top priority for researchers studying disease emergence and for regulatory agencies. Wheat blast was first discovered in southern Brazil

in 1985 [16] and the disease subsequently spread to the neighboring countries of Argentina, Bolivia and Paraguay [17-19] where it represents a considerable impediment to wheat production [20, 21]. Until recently, wheat blast had not been reported outside South America. In 2011, a single instance of infected wheat was discovered in the U.S., but analysis of the isolate responsible revealed that it was genetically similar to a local isolate from annual ryegrass and, therefore, unlikely to be an exotic introduction from South America [22]. More recently, in 2016, wheat blast was detected in Bangladesh [23]. Unlike the U.S. isolate, strains from this outbreak resembled South American wheat blast isolates rather than ryegrass-derived strains [15, 23], thereby confirming the spread of wheat blast from South America to Bangladesh.

It has recently been proposed that a subgroup of the wheat-infecting isolates, together with some strains pathogenic on *Eleusine* spp. and other *Poaceae* hosts, belongs to a new phylogenetic species, *Pyricularia graminis-tritici* (*Pgt*) that is well separated from other wheat- and ryegrass-infecting isolates, as well as pathogens of other grasses [24]. However, this proposed split was based on bootstrap support in a genealogy inferred from multilocus sequence concatenation, and Genealogical Concordance for Phylogenetic Species Recognition was not applied (GCPSR; [25, 26]). The observed lineage divergence appeared to be mostly driven by genetic divergence at one of 10 sequenced loci, raising questions on the phylogenetic support of this species.

The present study was designed to re-assess the hypothesis that *Pgt* constitutes a cryptic species within *M. oryzae* and, more generally, to infer population structure in relation to host of origin in this important plant pathogen. Using whole-genome sequences for 81

Magnaporthe isolates (76 *M. oryzae* from 12 host plant genera, four *M. grisea* from crabgrass [*Digitaria* spp.], and one *M. pennisetigena* from *Pennisetum* sp.) we addressed the following questions: do *M. oryzae* isolates form distinct host-specific lineages; and is there evidence for relatively long-term reproductive isolation between lineages (i.e. cryptic species) within *M. oryzae*? Our analyses of population subdivision and species identification revealed multiple divergent lineages within *M. oryzae*, each preferentially associated with one host plant genus, but refuted the existence of a novel cryptic phylogenetic species named *P. graminis-tritici*. In addition, analyses of gene flow revealed that genetic exchanges have contributed to the makeup of the multiple lineages within *M. oryzae*.

Results

Re-assessing the validity of the proposed *P. graminis-tritici* species by analyzing the original published data according to Phylogenetic Species Recognition by Genealogical Concordance (GCPSR)

To test the previous delineation of a subgroup of wheat-infecting isolates as a new phylogenetic species, we re-analyzed the Castroagudin et al. dataset [24], which mostly included sequences from Brazilian isolates. However, instead of using bootstrap support in a total evidence genealogy inferred from concatenated sequences for species delineation, we applied the GCPSR test [25, 26]. This test identifies a group as an independent evolutionary lineage (i.e. phylogenetic species) if it satisfies two conditions: (1) Genealogical concordance: the group is present in the majority of the single-locus genealogies, (2) Genealogical nondiscordance: the group is well-supported in at least one

147 single-locus genealogy and is not contradicted in any other genealogy at the same level of
148 support [25]. Visual inspection of the topologies and supports in each single-locus tree
149 revealed that GCPSR condition (1) was not satisfied since isolates previously identified as
150 belonging to the phylogenetic species *Pgt* grouped together in only one maximum
151 likelihood gene genealogy – the one produced using the *MPG1* locus (Figure S1A). The *Pgt*
152 separation was not supported by any of the nine other single-locus genealogies (Figures
153 S1B-S1J).

154 Next, we used the multilocus data as input to the program ASTRAL with the goal of
155 inferring a species tree that takes into account possible discrepancies among individual
156 gene genealogies [27-29]. The ASTRAL tree failed to provide strong support for the branch
157 holding the isolates previously identified as *Pgt* (Figure S2). Thus, analysis of the
158 Castroagudin et al. data according to GCPSR standards failed to support the existence of the
159 newly described *Pgt* species.

160 **Inferring population subdivision within *M. oryzae* using whole genome data**

161 We sought to test whether a phylogenomic study could provide better insight into the
162 possibility of speciation within *M. oryzae*. To this end, whole genome sequence data were
163 acquired for a comprehensive collection of 76 *M. oryzae* isolates from 12 host genera, four
164 *M. grisea* isolates from *Digitaria* spp. and one *M. pennisetigena* isolate from *Pennisetum*
165 (Table 1). The analysis included sequence data for strains collected on rice (*Oryza sativa*),
166 finger millet and goosegrass (*Eleusine spp.)*, wheat (*Triticum spp.*), tall fescue (*Festuca*
167 *arundinaceum)*, annual and perennial ryegrasses (*Lolium multiflorum and L. perenne,*
168 respectively), and barley (*Hordeum vulgare*). Representatives of previously unstudied host-

169 specialized populations from foxtails (*Setaria* sp.), St. Augustine grass (*Stenotaphrum*
170 *secundatum*), weeping lovegrass (*Eragrostis curvula*), signalgrass (*Brachiaria* sp.),
171 cheatgrass (*Bromus tectorum*) and oat (*Avena sativa*) were also included. SNPs identified in
172 aligned sequences of 2,682 orthologous single copy genes identified in all *M. oryzae*
173 genomes (in total ~6.6 Mb of sequence data), and from whole-genome SNPs identified from
174 pairwise blast alignments of repeat-masked genomes (average ~36 Mb aligned sequence).

175 First we employed the multivariate approach implemented in Discriminant Analysis of
176 Principal Components (DAPC; [30]) to examine population subdivision within *M. oryzae*.
177 Using the haplotypes identified from orthologous loci, the Bayesian Information Criterion
178 plateaued at K=10 in models varying K from 2 to 20 clusters, indicating that K=10 captures
179 the most salient features of population subdivision (Figure S3). Clusters identified at K=10
180 were as follows: (1) isolates from rice and two isolates from barley (dark green; referred to
181 as the *Oryza* lineage); (2) isolates from *Setaria* sp. (light green; referred to as the *Setaria*
182 lineage); (3) isolate Bm88324 from *Brachiaria mutica* (olive; referred to as the *Brachiaria1*
183 *lineage*); (4) isolate Bd8401 from *Brachiaria distachya* (brown; referred to as the
184 *Brachiaria2* lineage); (5) isolates from *Stenotaphrum* (red; referred to as the *Stenotaphrum*
185 lineage); (6) 17 of the 22 isolates from wheat and an isolate from *Bromus* (blue; referred to
186 as the *Triticum* lineage); (7) the remaining 3/22 isolates from wheat together with isolates
187 from *Lolium*, *Festuca*, oat and a second isolate from *Bromus* (purple; referred to as the
188 *Lolium* lineage); (8 & 9) isolates from *Eleusine* that formed two distinct clusters (light
189 orange and orange; referred to as the *Eleusine1* and *Eleusine2* lineages, respectively); and
190 (10) an isolate from *Eragrostis* (yellow; referred to as the *Eragrostis* lineage) (Figure 1).
191 Increasing K mostly resulted in further subdivision among the isolates from wheat, rice and

Lolium sp. The discovery of three wheat blast isolates that grouped with the *Festuca-Lolium* pathogens was important because it supports the idea that wheat-infecting isolates belong to at least two distinct populations.

Next, we inferred gene genealogies using maximum-likelihood and distance-based methods. Both approaches produced trees that corresponded well with the subdivisions identified in DAPC. The tree generated using maximum likelihood (ML) analysis of orthologous genes displayed a topology with ten lineages (Figure 2) showing one-to-one correspondence with the K clusters from DAPC (Figure 1; Figure S3). Nine of these lineages had >90% bootstrap support. The lineage that corresponded to the "blue" DAPC cluster (including the 17 isolates from wheat and isolate P29 from *Bromus*) had poor bootstrap support (50%).

The neighbor-joining tree built using "total genome" pairwise distances resolved very similar groupings to the DAPC (Figure 1; Figure S3) and ML ortholog tree (Figure 3). The only major discrepancy between ML and NJ trees was the confident placement of 87-120 – an isolate from rice – outside of the rice clade in the NJ tree (Figure 3).

Levels of polymorphism within and divergence between lineages/species

We compared levels of polymorphism within lineages to levels of divergence between lineages or species to apprehend the relative evolutionary depth of the lineages within *M. oryzae*. Genetic variability based on 2,682 orthologs was relatively low and one order of magnitude higher in the rice and wheat lineages (0.1% differences per site) than in the *Lolium* and *Setaria* lineages (other lineages not included in the calculations due to small sample sizes – only lineages with n>6 included; Table 2). The null hypothesis of no

recombination could be rejected in the *Lolium*, wheat, rice and *Setaria* lineages using the Pairwise Homoplasy Test implemented in the SPLITSTREE 4.13 program ([31]; p-value:0.0; Table 2).

Genome-wide nucleotide divergence was one order of magnitude higher between *M. oryzae* and its closest relatives, *M. grisea* and *M. pennisetigena*, than it was among isolates within *M. oryzae.* The maximum pairwise distance (number of differences per kilobase) between any two *M. oryzae* isolates was less than 1%, genome-wide (Figure S4; Table S1), compared with *M. oryzae* vs *M. grisea*, *M. oryzae* vs. *M. pennisetigena,* or *M. grisea* vs *M. pennisetigena*, all of which were consistently greater than 10%. The low level of genetic divergence among *M. oryzae* isolates, compared with that observed when comparing *M. oryzae* isolates to other established related species, provides good evidence against the existence of relatively ancient cryptic species within *M. oryzae*. (Table S1).

Re-assessment of Pgt as a novel species using whole genome data

While the 10 loci utilized in the Castroagudin et al. [24] study do not support the *Pgt* split based on GCPSR criteria, our DAPC and whole genome ML and NJ analyses supported the partitioning of wheat blast isolates into two, genetically-distinct lineages: one consisting almost exclusively of wheat-infecting isolates, the other comprising largely *Festuca*- and *Lolium*-infecting isolates as well as few wheat-infecting isolates (Figure 2; Figure 3). However, the Castroagudin et al. study did not include *Festuca*- and *Lolium*-infecting isolates and genome sequences from this study are not available. Therefore, to test for possible correspondence between the proposed *Pgt* species and the *Lolium* lineage (or indeed the *Triticum* lineage), we extended the 10 loci analysis to the *M. oryzae* genome

235 sequences used in the present study. For reference, we included the multilocus data for 16
236 isolates from the Castroagudin et al., representing all the major clades from that study. Nine
237 of the 10 loci were successfully recovered from 68 of our *M. oryzae* genome sequences. The
238 remaining locus, CH7-BAC9, was absent from too many genome sequences and, as a result,
239 was excluded from the analysis.

240 The nine concatenated loci produced a total-evidence RAxML tree in which very few
241 branches had bootstrap support greater than 50% (Figure 4). All of the *Pgt* isolates from
242 the Castroagudin et al. study were contained in a clade with 80% support. Inspection of the
243 *MPG1* marker that was reported to be diagnostic for *Pgt* (Castroagudin et al. 2016)
244 revealed that all of the isolates in this clade contained the *Pgt*-type allele (green dots) and
245 should therefore be classified as *Pgt* (Figure 4). Critically, however, a few isolates outside
246 this clade also harbored the *Pgt*-type allele. Moreover, the clade also included isolates from
247 the present study which came from wheat, annual ryegrass, perennial ryegrass, tall fescue,
248 finger millet, and goosegrass – isolates that did not group together in the DAPC analysis
249 (Figure 1), or in the ML and NJ trees built using the orthologous genes or whole genome
250 SNP data (Figure 2; Figure 3). Isolates carrying *Pgt*-type allele were in fact distributed
251 among three genetically distinct and well-supported clades (Figure 2; Figure 3).
252 Furthermore, visual inspection of the topologies and bootstrap supports for each single-
253 locus tree revealed that GCPSR criteria were not satisfied for the clade including all of the
254 *Pgt* isolates from the Castroagudin et al. Thus, isolates characterized by Castroagudin et al.
255 [24] as *Pgt* fail to constitute a phylogenetically cohesive group based on total genome
256 evidence and, thus, the existence of the *Pgt* species is not supported by our new genome-
257 wide data and analyses.

The basis for the previous designation of *Pgt* as a novel species was clearly revealed when *MPG1* alleles were mapped onto the ML and NJ trees. The distribution of *MPG1* alleles among different *M. oryzae* lineages was discontinuous (Figure S5). As an example, isolates from the *Triticum* lineage carried three different *MPG1* alleles. Two of these (including the *Pgt*-type) were also present in the *Lolium* lineage, while the third *MPG1* (ACT17T-C-6CAA140, Figure S5) was shared by distantly-related isolates from the *Stenotaphrum* lineage (Figure S5). Isolates from the *Eleusine* lineage also carried *Pgt*-type *MPG1* allele and two other variants, while isolates from the *Setaria* and *Oryza* lineages carried an *MPG1* allele distinct from all the others (Figure S5). Overall, the distribution of *MPG1* alleles points to the occurrence of incomplete lineage sorting and gene flow during *M. oryzae* diversification. Importantly, seven markers studied by Castroagudin et al. – including *MPG1* - showed discontinuities in their distributions among lineages defined using genome-wide data and analyses (Figure S5). The two other markers (*ACT1* and *CHS1*) used by Castroagudin et al. showed no sequence variations among the 68 *M. oryzae* isolates analyzed in the present study (data not shown) and are not useful for phylogenetic classification.

Species tree inference and phylogenetic species recognition from genome-wide data

The total evidence genealogies generated using sequence data from 76 *M. oryzae* genomes using either distance-based (whole genomes) or maximum likelihood (2,682 single-copy orthologs) phylogenetic methods were highly concordant in terms of lineage composition and branching order (Figure 2; Figure 3). However, concatenation methods can be positively misleading, as they assume that all gene trees are identical in topology and branch lengths and they do not explicitly model the relationship between the species tree

and gene trees [32]. To estimate the "species tree" and to re-assess previous findings of cryptic species within *M. oryzae*, we used a combination of species inference using the multispecies coalescent method implemented in ASTRAL [27-29] and a new implementation of the GCPSR that can handle genomic data.

The ASTRAL "species tree" with the local q1 support values on key branches is shown in Figure 5. The four *M. grisea* isolates from crabgrass (*Digitaria* sp.) and the *M. pennisetigena* isolate from fountaingrass (*Pennisetum* sp.) were included as outgroups, bringing the total number of isolates to 81 and reducing the dataset to 2,241 single-copy orthologous genes. The branches holding the clades containing the wheat blast isolates had q1 support values of 0.49, 0.39 and 0.37 which means that, in each case, fewer than 50% of the whole set of quartet gene trees recovered from the individual gene genealogies agreed with the local topology around these branches in the species tree. The branches that separated *M. grisea* and *M. pennisetigena* from *M. oryzae* had respective q1 values of 1, providing strong support for relatively ancient speciation. In contrast, the highest q1 value on any of the branches leading to the host-specialized clades was 0.8 for the *Setaria* pathogens, indicating that approximately 20% of the quartets recovered from individual gene trees were in conflict with the species tree around this branch. Together, these results indicate high levels of incomplete lineage sorting within, and/or gene flow involving these groups, and are thus inconsistent with the presence of genetically isolated lineages (i.e. species).

As a formal test for the presence of cryptic species within *M. oryzae*, we applied the phylogenetic species recognition criteria to the set of 2,241 single-copy orthologous genes using an implementation of the GCPSR scalable to any number of loci. Applying the GCPSR

following the non-discordance criterion of Dettman et al. (a clade has to be well supported by at least one single-locus genealogy and not contradicted by any other genealogy at the same level of support; [25]) resulted in the recognition of no species within *M. oryzae*.

Historical gene flow between lineages

The existence of gene flow and/or incomplete lineage sorting was also supported by phylogenetic network analysis. We used the network approach neighbor-net implemented in SPLITSTREE 4.13 [25] to visualize evolutionary relationships, while taking into account the possibility of recombination within or between lineages. The network inferred from haplotypes identified using the 2,682 single-copy orthologs in the 76 *M. oryzae* strains, showed extensive reticulation connecting all lineages, consistent with recombination or incomplete lineage sorting (Figure 6).

To disentangle the role of gene flow versus incomplete lineage sorting in observed network reticulations, but also to gain insight into the timing and extent of genetic exchanges, we used ABBA/BABA tests, which compare numbers of two classes of shared derived alleles (the ABBA and BABA classes). For three lineages P1, P2 and P3 and an outgroup with genealogical relationships (((P1,P2),P3),O), and under conditions of no gene flow, shared derived alleles between P2 and P3 (ABBA alleles) and shared derived alleles between P1 and P3 (BABA alleles) can only be produced by incomplete lineage sorting, and should be equally infrequent [33]. Differences in numbers of ABBA and BABA alleles are interpreted as gene flow assuming no recurrent mutation and no deep ancestral population structure within lineages. We computed *D*, which measures the imbalance between numbers of ABBA and BABA sites and is used to test for admixture in ((P1,P2),P3) triplets, with D>0 implying

gene flow between P2 and P3, and D<0 implying gene flow between P1 and P3 [34, 35]. We also made use of the heterogeneity in divergence time between members of ((P1,P2),P3) triplets to examine gene flow across three time periods [33], following these principles: (i) triplets including the most recently diverged lineages as P1 and P2 (i.e. the *Triticum* and *Lolium* lineages, the two *Eleusine* lineages, or the *Oryza* and *Setaria* lineages) carried information about gene flow across relatively recent times, (ii) triplets including, as P1 and P2, two lineages from the same main group of lineages (i.e. *Eragrostis*/*Eleusine*1/*Eleusine*2/*Triticum*/*Lolium* or *Brachiaria2*/*Setaria*/*Oryza*, excluding (P1,P2) pairs already used in (i)) carried information about gene flow across intermediate times, and (iii) triplets including, as P1 and P2, two lineages from different main groups of lineages (i.e. *Eragrostis* / *Eleusine*1 / *Eleusine*2 / *Triticum* / *Lolium* and *Brachiaria2* / *Setaria* / *Oryza*) and *Stenotaphrum* or *Brachiaria1* as P3 carried information about gene flow across a relatively long time period (Figure S6).

The D statistic measuring differences in counts of ABBA and BABA alleles was significantly different from zero (Z-score>3) in 104 of 120 lineage triplets, consistent with a history of gene flow between lineages within *M. oryzae* (Table S2). Given that a (P1,P2) pair can be represented as multiple ((P1,P2),P3) triplets, and that the sign of D indicates what is the pair involved in gene flow within each triplet, the 104 triplets with significant D values in fact represented 35 pairs connected by gene flow, spanning the three time scales defined by the phylogenetic affiliation of lineages (Figure S6). Lineages were equally represented in triplets deviating from null expectations assuming no gene flow, no ancient structure and no recurrent mutations. Consistent with historical gene flow, searches for private allele found no gene, among the 2241 gene surveyed, carrying mutations exclusive to a single

lineage. Together, these results indicate that gene flow was widespread, both across historical times and lineages, but it cannot be excluded that much of the signal was caused by events that happened prior to lineage splitting.

Recent admixture and gene flow between lineages

We then used the program STRUCTURE [36-38] to detect possible recent admixture between lineages (Figure S3). STRUCTURE uses Markov chain Monte Carlo simulations to infer the assignment of genotypes into K distinct clusters, minimizing deviations from Hardy–Weinberg and linkage disequilibria within each cluster. The patterns of clustering inferred with STRUCTURE were largely similar to those inferred with DAPC. STRUCTURE analysis provided evidence for admixture at all K values (Figure S3), suggesting that recent admixture events have recently shaped patterns of population subdivision within *M. oryzae*. 'Chromosome painting', a probabilistic method for reconstructing the chromosomes of each individual sample as a combination of all other homologous sequences [39], also supported the lack of strict genetic isolation between lineages (Text S1).

Discussion

Population subdivision but no cryptic phylogenetic species

Using population- and phylogenomic analyses of single-copy orthologous genes and whole-genome SNPs identified in *M. oryzae* genomes from multiple cereal and grass hosts, we provide evidence that *M. oryzae* is subdivided in multiple lineages preferentially associated with one host plant genus. Neither the re-analysis of previous data, nor the analysis of new data using previous phylogenetic species recognition markers, supports the existence of a wheat blast-associated species called *P. graminis-tritici* [24]. Marker *MPG1*, which holds

most of the divergence previously detected, does not stand as a diagnostic marker of the wheat-infecting lineage of *M. oryzae* when tested in other lineages. Previous conclusions about the existence of cryptic species *P. graminis-tritici* also stem from the fact that available information on *M. oryzae* diversity had been insufficiently taken into account. In particular, isolates from the lineages most closely related to wheat strains (i.e. isolates from the *Lolium* lineage; [11, 12, 15, 22]) were not represented in previous species identification work [24]. Using phylogenetic species recognition by genealogical concordance we could not identify cryptic phylogenetic species and thus *M. oryzae* is not, strictly speaking, a species complex. As a consequence, *Pyricularia graminis-tritici* cannot - and should not – be considered as a valid name for wheat-infecting strains, because (1) it refers to a subset of wheat-infecting strains, and quarantine on *P. graminis-tritici* alone would not prevent introduction of aggressive wheat blast pathogens (2) it groups very aggressive wheat pathogens from South America and South Asia with *Eleusine*-infecting strains that are largely distributed in the world. Given the devastating potential of wheat blast disease, it is vital that accurate strain identification and species assignment can be carried out by plant health agencies in order to safeguard against importation and spread of the disease. Correct species assignment is therefore a critical consideration. Hence, although the formal rules of taxonomy would imply treating *P. graminis-tritici* as synonym of *Magnaporthe oryzae*, we strongly recommend dismissal of *P. graminis-tritici* as a valid name to refer to wheat-infecting strains of *M. oryzae*.

Incipient speciation by ecological specialization following host-shifts

Several features of the life cycle of *M. oryzae* are conducive to speciation by ecological specialization following host shifts, suggesting that the observed pattern of population

393 subdivision in *M. oryzae* actually corresponds to ongoing speciation. Previous experimental
394 measurements of perithecia formation and ascospore production – two important
395 components of reproductive success- suggested inter-fertility in vitro between most pairs
396 of lineages with high levels of ascospore viability [40-43]. This suggests that intrinsic pre-
397 or post-mating reproductive barriers, such as assortative mating by mate choice or gametic
398 incompatibility, and zygotic mortality, are not responsible for the relative reproductive
399 isolation between lineages – which creates the observed pattern of population subdivision.
400 Instead, the relative reproductive isolation between lineages could be caused by one or
401 several pre- or post-mating barriers (Table 1 in [44]), such as mating-system isolation or
402 hybrid sterility (intrinsic barrier), or difference in mating times, difference in mating sites,
403 immigrant inviability, or ecologically-based hybrid inviability (extrinsic barriers).

404 Previous pathogenicity assays revealed extensive variability in the host range of *M. oryzae*
405 isolates, and both in terms of pathogenicity towards a set of host species or pathogenicity
406 towards a set of genotypes from a given host [41, 45]. Indeed, extensive genetic analyses
407 show that host species specificity in *M. oryzae*, similar to rice cultivar specificity, could be
408 controlled by a gene-for-gene relationship in which one to three avirulence genes in the
409 fungus prevent infection of particular host species [43, 46, 47]. Loss of the avirulence genes
410 would allow infection of novel hosts to occur. Additionally, host species specificity is not
411 strictly maintained. Under controlled conditions, most lineages have at least one host in
412 common [45], and strains within one lineage can still cause rare susceptible lesions on
413 naive hosts [21, 48]. Moreover, a single plant infected by a single genotype can produce
414 large numbers of spores in a single growing season [49], allowing the pathogen to persist
415 on alternative host even if selection is strong, and promoting the rapid and repeated

creation of genetic variation [6]. Although some of these features appear to be antagonistic with the possibility of divergence by host-specialization within *M. oryzae*, our finding that the different lineages within *M. oryzae* tend to be sampled on a single host suggests that ecological barriers alone may in fact contribute to reduce gene flow substantially between host-specific lineages. Differences in the geographic distribution of hosts, for which the level of sympatry has varied -and still varies- in space and time, might also contribute to reduced gene flow between lineages infecting different hosts, although some level of sympatry at some time is required so that new hosts could become infected, triggering host-range expansion or host-shifting.

Mating within host (i.e. reproduction between individuals infecting the same host), and to a lesser extent mating system isolation (i.e. lack of outcrossing reproduction), may contribute to further reduce gene flow between *M. oryzae* lineages. The fact that mating in *M. oryzae* likely occurs within host tissues, such as dead stems [50], may participate in the maintenance of the different lineages by decreasing the rate of reproduction between isolates adapted to different hosts [6]. Loss of sexual fertility also appears to have a role in lineage maintenance. The rice lineage, in particular, is single mating-type and female-sterile throughout most of its range, which would reduce the chance of outcrossing sex with members of other lineages [51]. Our analyses rejected the null hypothesis of clonality in all lineages, but they provided no time frame for the detected recombination events. Population-level studies and experimental measurements of mating type ratios and female fertility are needed to assess the reproductive mode of the different lineages within *M. oryzae* in the field.

Inter-lineage gene flow

Several potential barriers contribute to reduce genetic exchanges between *M. oryzae* lineages (see previous paragraph), but not completely so, as evidenced by signal of gene flow and admixture detected in our genomic data. We hypothesize that the lack of strict host specialization of the different lineages is a key driver of inter-lineage gene flow. Many of the grass or cereal species that are hosts to *M. oryzae* are widely cultivated as staple crops or widely distributed as pasture or weeds, including "universal suscepts" such as barley, Italian ryegrass, tall fescue and weeping lovegrass [40], increasing the chance for encounters and mating between isolates with overlapping host ranges. These shared hosts may act as a platform facilitating encounters and mating between fertile and compatible isolates from different lineages, thereby enabling inter-lineage gene flow [52]. Plant health vigilance is therefore warranted for disease emergence via recombination in regions where multiple lineages are in contact and shared hosts are present. This is particularly so, given that once infection of novel host has taken place (i.e. host shift or host range expansion), the fungus has the capacity to build inoculum levels very rapidly, facilitating spread of the disease over considerable distances. It is striking, for example, that wheat blast has, within a year, spread from Bangladesh into the West Bengal region of India where it emerged in 2017 (openwheatblast.org).

Conclusion

Using a population genomics framework, we show that *M. oryzae* is subdivided into multiple lineages with limited host range and present evidence of genetic exchanges between them. Our findings provide greater understanding of the eco-evolutionary factors underlying the diversification of *M. oryzae* and highlight the practicality of genomic data for

epidemiological surveillance of its different intraspecific lineages. Reappraisal of species boundaries within *M. oryzae* refuted the existence of a novel cryptic phylogenetic species named *P. graminis-tritici,* underlining that the use of node support in total evidence genealogies based on a limited dataset in terms of number of loci and of range of variation in origin (geography and host) of isolates can lead to erroneously identify fungal cryptic species. Our work illustrates the growing divide between taxonomy that 'creates the language of biodiversity' [53] based on limited sets of characters, and genomic data that reveals more finely the complexity and continuous nature of the lineage divergence process called speciation.

Materials and Methods

Fungal strains

Thirty-eight newly sequenced genomes were analyzed together with 43 published genomes [13, 14, 22, 54-56] resulting in a total of 81 *Magnaporthe* strains, including 76 *M. oryzae* genomes representing 12 different hosts available for analysis (Table 1). We also included as outgroups one strain of *Pyricularia pennisetigena* from *Pennisetum* sp. and four strains of *Pyricularia grisea* (syn. *Magnaporthe grisea*) from crabgrass (*Digitaria sanguinalis*). All newly sequenced strains were single-spored prior to DNA extraction.

Genome sequencing and assembly

New genome data were produced by an international collaborative effort. Characteristics of genome assemblies are summarized in Table S3. For newly sequenced genomes provided by MF and BV, sequences were acquired on a MiSeq machine (Illumina, Inc.). Sequences

were assembled using the paired-end mode in NEWBLER V2.9 (Roche Diagnostics, Indianapolis, IN). A custom perl script was used to merge the resulting scaffolds and contigs files in a non-redundant fashion to generate a final assembly. Newly sequenced genomes BR130 and WHTQ provided by TM were sequenced using an Illumina paired-end sequencing approach at >50X depth. Short reads were assembled *de-novo* using VELVET 1.2.10 [57] resulting in a 41.5Mb genome for BR130 with N50 44.8Kb, and 43.7Mb for WHTQ with N50 36.2Kb. For newly sequenced genomes provided by DS and NT, DNA was sequenced on the Illumina HiSeq 2500 producing 100 base paired-end reads, except in the case of VO107 which was sequenced on the Illumina Genome Analyzer II producing 36 base paired-end reads. Reads were filtered using FASTQ-MCF and assembled 'de novo' using VELVET 1.2.10 [57].

Orthologous genes identification in genomic sequences

Protein-coding gene models were predicted using AUGUSTUS V3.0.3 [58]. Orthologous genes were identified in the 76-genomes *M. oryzae* or in the dataset including outgroups using PROTEINORTHO [59]. The v8 version of the 70-15 *M. oryzae* reference genome [60] was added at this step in order to validate the predicted sets of orthologs. Only orthologs that were single-copy in all genomes were included in subsequent analyses. Genes of each single-copy orthologs sets were aligned using MACSE [61]. Sequences from the lab strain 70-15 were removed and not included in further analyses due to previously shown hybrid origin [13]. Only alignments containing polymorphic sites within *M. oryzae* strains were kept for further analyses. This resulted in 2,241 alignments for the whole dataset, and 2,682 alignments for the 76 *M. oryzae* strains.

506

507 **Population subdivision, summary statistics of polymorphism and divergence**

508 Population subdivision was analyzed using DAPC and STRUCTURE [30, 36-38], based on
509 multilocus haplotype profiles identified from ortholog alignments using a custom Python
510 script. DAPC was performed using the ADEGENET package in R [13]. We retained the first 30
511 principal components, and the first 4 discriminant functions. Ten independent STRUCTURE
512 runs were carried out for each number of clusters K, with 100,000 MCMC iterations after a
513 burn-in of 50,000 steps.

514 Polymorphism statistics were computed using EGGLIB 3.0.0b10 [62] excluding sites with
515 >30% missing data. Divergence statistics were computed using a custom perl script.

516 To infer total evidence trees within the 76 *M. oryzae* strains (respectively within the 81
517 *Magnaporthe* strains), all sequences from the 2,682 (respectively 2,241) orthologous
518 sequences were concatened. The maximum-likelihood tree was inferred using RAxML [63]
519 with the GTR-gamma model, and bootstrap supports were estimated after 1000 replicates.

520 **Retrieval of loci used in the Castroagudin et al. (2016) study**

521 The 10 loci used by Castroagudin et al. [24], i.e. actin (ACT), beta-tubulin1 (βT-1),
522 calmodulin (CAL), chitin synthase 1 (CHS1), translation elongation factor 1-alpha (EF1-α),
523 hydrophobin (MPG1), nitrogen regulatory protein 1 (NUT1), and three anonymous
524 markers (CH6, CH7-BAC7, CH7-BAC9), were search in all genomes using BLASTn. Due to
525 heterogeneity in the quality of assemblies, 9 of the 10 loci could be full-length retrieved
526 without ambiguity in 68 out of the 81 available genomes, still representative of the
527 diversity of host plants.

528

529 **Secondary data analysis**

530 Species recognition based on multiple gene genealogies as described by Castroagudin et al.
531 [24] was repeated following the reported methods. The robustness of the Pgt species
532 inference was tested by re-iterating the study, omitting one marker at a time. Individual
533 genealogies were built using RAxML with the GTR-gamma model and 100 bootstrap
534 replicates.

535

536 **Inference of "species tree" using ASTRAL**

537 The ASTRAL method [27, 29] is based on the multi-species coalescent and allows taking
538 into account possible discrepancies among individual gene genealogies to infer the "species
539 tree". Individual genealogies inferred using RAxML with the GTR-gamma model and 100
540 bootstrap replicates, were used as input data for ASTRAL analysis. Local supports around
541 branches were evaluated with 100 multilocus bootstrapping using the bootstrap replicates
542 inferred from each individual gene tree as input data, and with local quartet supports (q1,
543 obtained using the –t option of ASTRAL) that represent the proportion of quartets
544 recovered from the whole set of individual gene trees that agree with the local topology
545 around the branch in the species tree.

546

547 **MPG1-based classification**

548 The *MPG1* hydrophobin sequence is described as being diagnostic for the
549 *P. graminis-tritici/M. oryzae* species split [24]. *MPG1* sequences from one of each species

(GIs: KU952644.1 for *P. gramini-triticis*, KU952661.1 for *M. oryzae*) were used as BLAST [64] queries to classify isolates as either *P. graminis-tritici* or *M. oryzae.*

Signatures of gene flow and/or incomplete lineage sorting

A phylogenetic network was built using SPLITSTREE 4.13 [65], based on the concatenation of sequences at single-copy orthologs identified in *M. oryzae*, excluding sites with missing data, sites with gaps, singletons, and monomorphic sites. The null hypothesis of no recombination was tested using the PHI test implemented in SPLITSTREE.

ABBA/BABA tests

ABBA/BABA tests were performed using custom python scripts. The D statistic measuring the normalized difference in counts of ABBA and BABA sites was computed using equation (2) in ref. [66]. Significance was calculated using block jackknife approach (100 replicates, 1k SNPs blocks), to account for non-independence among sites.

Probabilistic chromosome painting.

We used CHROMOPAINTER program version 0.0.4 for probabilistic chromosome painting. This analysis was based on biallelic SNPs without missing data identified in the set of 2,682 single-copy orthologs, ordered according to their position in the reference genome of the rice-infecting strain 70-15. We initially estimated the recombination scaling constant N_e and emission probabilities (μ) by running the expectation-maximization algorithm with 200 iterations for each lineage and chromosome. Estimates of N_e and μ were then computed as averages across lineages, weighted by chromosome length, rounded to the

nearest thousand for N_e (N_e=5000; μ=0.0009). The file *recom_rate_infile* detailing the recombination rate between SNPs was built using the INTERVAL program in LDHAT version 2.2 [67] based on the whole dataset combining isolates from all lineages, with 10 repeats by chromosome to check for convergence. Estimated N_e and μ values and the per-chromosome recombination maps estimated using LDHAT were then used to paint the chromosomes of each lineage, considering the remaining lineages as donors, using 200 expectation-maximization iterations. For each lineage and each chromosome, CHROMOPAINTER was run thrice to check for convergence.

Phylogenetic species recognition

We used an implementation of the GCPSR scalable to genomic data (https://github.com/b-brankovics/GCPSR). The method works in two steps: (1) Concordance and non-discordance analysis produces a genealogy that has clades that are both concordant and non-discordant across single gene genealogies, with support value for each of the clades being the number of single gene genealogies harboring the given clade at bootstrap support above 95%; (2) Exhaustive subdivision places all the strains into the least inclusive clades, by removing clades that would specify a species within potential phylogenetic species. We kept only two outgroup sequences per gene (BR29, *M. grisea*; Pm1, *M. pennisetigena*) to make sure to have the same isolate at the root of all genealogies (Pm1 isolate). Majority-rule consensus trees were produced from 100 outgrouped RAxML bootstrap replicates for all 2241 genes. The concordance and non-discordance analysis was carried out assuming 95 as the minimum bootstrap support value, and a discordance threshold of 1. Exhaustive subdivision was carried out using a concordance threshold of 1121.

596

597 **Whole genome alignment & tree building**

598 A custom perl script was used to mask sequences that occur in multiple alignments when
599 the genome is BLASTed against itself. The masked genomes were then aligned in a pairwise
600 fashion against all other genomes using BLAST [64]. Regions that did not uniquely align in
601 each pair at a threshold of $1e^{-200}$, were excluded. SNPs were then identified for each
602 pairwise comparison and scaled by the total number of nucleotides aligned after excluding
603 repetitive and duplicate regions. This produced a distance metric of SNPs/Mb of uniquely
604 aligned DNA. The pairwise distances were used to construct phylogenetic trees with the
605 neighbor-joining method as implemented in the R package, Analyses of Phylogenetics and
606 Evolution (APE) [68].

607 Because alignments are in pairwise sets as opposed to a single orthologous set, assessment
608 of confidence values by traditional bootstrapping by resampling with replacement is not
609 possible. Instead, confidence values were assigned by creating 1,000 bootstrap trees with
610 noise added from a normal distribution with a mean of zero, and the standard deviation
611 derived from the pairwise distances between or within groups.

Acknowledgments

We thank Sophien Kamoun for inspiration and for providing critical input on a previous version of the manuscript, Alfredo Urshimura for collecting and supplying us with the DNA of Brazilian isolates, and the Southgreen and Migale computing facilities. We also thank A. Akhunova at the Kansas State University Integrated Genomics Facility, and J. Webb, M. Heist and R. Ellsworth at the University of Kentucky for their technical assistance. Support is acknowledged by the Agriculture and Food Research Initiative Competitive Grant No. 2013-68004-20378 from the USDA National Institute of Food and Agriculture. This is contribution number 18-005-J from the Kansas Agricultural Experiment Station and contribution number XX-XX-XXX from the Kentucky Agricultural Experiment Station.

Tables

Table 1. *Magnaporthe oryzae*, *M. grisea* and *M. pennisetigena* strains used in this study

Isolate ID	Synonym	Host	Year	Locality	NCBI accession	Sequence source	References
BdBar	BdBar16-1	*Triticum aestivum*	2016	Barisal, Bangladesh		[23]	
BdJes	BdJes16-1	*Triticum aestivum*	2016	Jessore, Bangladesh		[23]	
BdMeh	BdMeh16-1	*Triticum aestivum*	2016	Mehepur, Bangladesh		[23]	
B2		*Triticum aestivum*	2011	Bolivia	XXXXXXXX	[69]	
B71		*Triticum aestivum*	2012	Bolivia		[23]	[69]
Br7		*Triticum aestivum*	1990	Parana, Brazil	XXXXXXXX	[22]	[69]
BR0032	BR32	*Triticum aestivum*	1991	Brazil		[13]	[15]
Br48		*Triticum aestivum*	1990	Mato Grosso do Sul, Brazil		[14]	[22]
Br80		*Triticum aestivum*	1991	Brazil	XXXXXXXX	[22]	[69]
Br130		*Triticum aestivum*	1990	Mato Grosso do Sul, Brazil	XXXXXXXX	[22]	
P3		*Triticum durum*	2012	Canindeyu, Paraguay	XXXXXXXX	[69]	
PY0925		*Triticum aestivum*	2009	Predizes, Brazil		[15]	
PY36-1	PY36.1	*Triticum aestivum*	2007	Brasilia, Brazil		[15]	
PY5003	PY05003	*Triticum aestivum*	2005	Londrina, Brazil		[15]	
PY5010	PY05010	*Triticum aestivum*	2005	Londrina, Brazil		[15]	[69]
PY5033	PY05033	*Triticum aestivum*	2005	Londrina, Brazil		[15]	
PY6017	PY06017	*Triticum aestivum*	2006	Coromandel, Brazil		[15]	
PY6045	PY06045	*Triticum aestivum*	2006	Goiania, Brazil		[15]	
PY86-1	PY86.1	*Triticum aestivum*	2008	Cascavel, Brazil		[15]	
T25		*Triticum aestivum*	1988	Parana, Brazil	XXXXXXXX	[22]	[69]
WHTQ		*Triticum aestivum*	ND	Brazil	XXXXXXXX	[22]	
WBKY11	WBKY11-15	*Triticum aestivum*	2011	Lexington, KY, USA	XXXXXXXX	[22]	[69]
P28	P-0028	*Bromus tectorum*	2014	Paraguay	XXXXXXXX	[69]	
P29	P-0029	*Bromus tectorum*	2014	Paraguay	XXXXXXXX	[69]	
CHRF		*Lolium perenne*	1996	Silver Spring, MD, USA	XXXXXXXX	[69]	
CHW		*Lolium perenne*	1996	Annapolis, MD, USA	XXXXXXXX	[69]	
FH		*Lolium perenne*	1997	Hagerstown, MD, USA	XXXXXXXX	[22]	[69]
GG11		*Lolium perenne*	1997	Lexington, KY, USA	XXXXXXXX	[22]	
HO		*Lolium perenne*	1996	Richmond, PA, USA	XXXXXXXX	[22]	
LpKY97	LpKY97-1	*Lolium perenne*	1997	Lexington, KY, USA	XXXXXXXX	[22]	[69]

PgKY	PgKY4OV2.1	*Lolium perenne*	2000	Lexington, KY, USA		[15]	
PGPA	PgPA18C-02, PgPA	*Lolium perenne*	1998	Pennsylvania, USA		[15]	
PL2-1		*Lolium multiflorum*	2002	Pulaski Co., KY, USA	XXXXXXXX	[22]	
PL3-1		*Lolium multiflorum*	2002	Pulaski Co., KY, USA	XXXXXXXX	[22]	[69]
Pg1213-22		*Festuca arundinaceum*	1999/ 2000	GA	XXXXXXXX	[69]	
TF05-1		*Festuca arundinaceum*	2005	Lexington, KY, USA	XXXXXXXX	this study	
IB33		*Oryza sativa*	ND	Texas, USA	XXXXXXXX	A. Marchetti	
FR13	FR0013	*Oryza sativa*	1988	France		[13]	[15]
GY11	GY0011, Guy11	*Oryza sativa*	1988	French Guyana		[13]	[15, 22]
IA1	ARB114	*Oryza sativa*	2009	Arkansas, USA	XXXXXXXX	[69]	
IB49	ZN61	*Oryza sativa*	1992	Arkansas, USA	XXXXXXXX	[69]	
IC17	ZN57	*Oryza sativa*	1992	Arkansas, USA	XXXXXXXX	[69]	
IE1K	TM2	*Oryza sativa*	2003	Arkansas, USA	XXXXXXXX	[69]	
INA168	Ina168	*Oryza sativa*	1958	Aichi, Japan		[14]	[70]
KEN53-33	Ken53-33	*Oryza sativa*	1953	Aichi, Japan		[14]	
ML33		*Oryza sativa*	1995	Mali	XXXXXXXX	this study	
P131		*Oryza sativa*	ND	Japan		[55]	[22, 56]
Y34		*Oryza sativa*	1982	China, Yunnan		[55]	[22, 56]
P-2	P2	*Oryza sativa*	1948	Aichi, Japan		[14]	
PH0014-rn	PH0014, PH14	*Oryza sativa*	ND	The Philippines		[13]	[15]
TH3		*Oryza sativa*	ND	Thailand		[14]	[14]
87-120		*Oryza sativa*	ND		XXXXXXXX	this study	
TH0012-rn	TH0012, TH12	*Hordeum vulgare*	ND	Thailand		[13]	[15]
TH0016	TH16	*Hordeum vulgare*	ND	Thailand		[13]	[15]
Arcadia		*Setaria viridis*	1998	Lexington, KY, USA	XXXXXXXX	[22]	[69]
US0071	US71	*Setaria spp.*	ND	USA		[13]	[15]
GrF52		*Setaria viridis*	2001	Lexington, KY, USA	XXXXXXXX	this study	
KANSV1-4-1	KNSV	*Setaria viridis*	1975	Kanagawa, Japan		[14]	
SA05-43		*Setaria viridis*	2005	Nagasaki, Japan		[14]	
Sv9610		*Setaria viridis*	1996	Zhejian, China		[56]	
Sv9623		*Setaria viridis*	1996	Zhejian, China		[56]	
GFSI1-7-2	GFSI	*Setaria italica*	1977	Gifu, Japan		[14]	

B51		*Eleusine indica*	2012	Quirusillas, Bolivia	XXXXXXXX	[22]	[56, 69]
BR62		*Eleusine indica*	1991	Brazil		[15]	
CD156	CD0156	*Eleusine indica*	1989	Ferkessedougou, Ivory Coast		[13]	[15]
EI9411		*Eleusine indica*	1990	Fujian, China		[56]	
EI9064		*Eleusine indica*	1996	Fujian, China		[56]	
G22	WGG-FA40	*Eleusine coracana*	1976	Japan	XXXXXXXX	this study	
Z2-1		*Eleusine coracana*	1977	Kagawa, Japan		[14]	[70]
PH42		*Eleusine indica*	1983	The Philippines	XXXXXXXX	[69]	
SSFL02		*Stenotaphrum secundatum*	2002	Disneyworld, FL, USA	XXXXXXXX	[69]	
SSFL14-3		*Stenotaphrum secundatum*	2014	New Smyrna, FL, USA	XXXXXXXX	this study	
G17	K76-79	*Eragrostis curvula*	1976	Japan	XXXXXXXX	[15]	
Br58		*Avena sativa*	1990	Parana, Brazil		[14]	[70]
Bd8401		*Brachiaria distachya*	1984	The Philippines	XXXXXXXX	this study	
Bm88324		*Brachiaria mutica*	1988	The Philippines	XXXXXXXX	this study	
PM1		*Pennisetum americanum*	1990	Georgia, USA	XXXXXXXX	this study	
BR29	BR0029	*Digitaria sanguinalis*	1989	Brazil		[13]	
Dig41		*Digitaria sanguinalis*	ND	Hyogo, Japan		[14]	[70]
DsLIZ		*Digitaria sanguinalis*	2000	Lexington, KY, USA	XXXXXXXX	[69]	
VO107		*Digitaria sanguinalis*	1981	Texas, USA	XXXXXXXX	this study	

ND, no data. "References" lists studies that used the sequencing data, besides present study. Isolates Br116.5, Br118.2, TP2, MZ5-1-6, and Br35 sequenced by Inoue et al. [70], Bangladeshi isolates and isolates PY05002, PY06025, PY06047, PY25.1, PY35.3, PY05035 sequenced by Islam et al. [15], isolate SA05-144 sequenced by Yoshida et al. [14], and isolates DS9461 and DS0505 sequenced by Zhong et al. [56] were not included in the study.

Table 2. Summary of population genetic variation at 2682 single-copy orthologous genes in wheat, lolium, rice and setaria lineages of *Magnaporthe oryzae*

Lineage	n	S	K	H_e	θ_w	π	PHI test (p-value)
Wheat	20	5.8	1.9	0.17	1.28E-03	1.24E-03	0
Lolium	17	3.1	1.5	0.10	7.02E-04	6.54E-04	0
Rice	18	5.3	2.3	0.12	1.55E-03	7.75E-04	0
Setaria	8	2.6	1.8	0.18	9.10E-04	7.68E-04	0

Other lineages were not included in calculations because of too small a sample size (n<6); n is sample size; θ_w is Watterson's θ per bp; π is nucleotide diversity per bp; H_e is haplotype diversity; K is the number of haplotypes. PHI test is the Pairwise Homoplasy Test. The PHI test is implemented in SPLITSTREE. The null hypothesis of no recombination was tested, for the PHI test using random permutations of the positions of the SNPs based on the expectation that sites are exchangeable if there is no recombination

Figure legends

Figure 1. Discriminant Analysis of Principal Components, assuming K=10 clusters

Each isolate is represented by a thick vertical line divided into K segments that represent the isolate's estimated membership probabilities in the K=10 clusters (note that all isolates have high membership probabilities in a single cluster, hence only a single segment is visible). The host of origin of samples is shown below the barplot, and lineage IDs are shown above the barplot.

Figure 2. Maximum likelihood tree based on the concatenation of 2,682 orthologous coding sequences extracted from 76 *M. oryzae* genome. Nodes with bootstrap support >90% are indicated by dots (100 bootstrap replicates).

Figure 3. Total evidence neighbor-joining distance tree using pairwise distances (number of differences/kb) calculated from analysis of pairwise blast alignments between repeat-masked genomes. Only nodes with confidence > 80% (see methods) are labeled. Gray ovals are drawn around the main host-specialized populations for clarity.

Figure 4. Maximum likelihood tree based on concatenated dataset comprising nine loci used in Castroagudin et al. [24], **retrieved from 76 *M. oryzae* genomes.** Numbers above branches represent bootstrap supports after 100 bootstrap replicates. Only nodes with bootstrap support > 50 are labelled. Representatives of isolates used by Castroagudin et al. [24] in their study were included in the analysis and are colored in light grey. Green dots mark the strains containing the *Pgt*-type allele according to Castroagudin et al. [24].

Figure 5. ASTRAL analysis to test for incomplete lineage sorting/gene flow among 81 *Magnaporthe* genomes, using 2,241 single-copy orthologous sequence loci. Thicker

branches represent branches that have a bootstrap support > 50 after multilocus bootstrapping. Number above branches represent q1 local support (i.e. the proportion of quartet trees in individual genealogies that agree with the topology recovered by the ASTRAL analysis around the branch), with q1 values showed on black background for branches holding wheat blast isolates.

Figure 6. Neighbor-Net network built with SPLITSTREE. The figure shows relationships between haplotypes identified based on the full set of 25,078 SNPs identified in 2,682 single-copy orthologs, excluding sites missing data, gaps and singletons.

Supplementary Tables legends

Table S1. Pairwise distances measured in SNPs per megabase of uniquely aligned DNA

Table S2. (A) Gene flow signatures from ABBA/BABA tests. P1, P2, P3 refer to the three lineages used for the tests. The D statistic tests for an overrepresentation of ABBA versus BABA patterns. SE is the standard error. Z-score and P-value for the test of whether D differs significantly from zero, calculated using 1000 block jackknifes of 100 SNPs. Analyses were based on 354,848 biallelic SNPs identified in 2241 single-copy orthologous genes, with *M. grisea* as the outgroup. Brachiaria1: Bm88324. Brachiaria2: Bd8401. Eleusine1 and Eleusine2 are, respectively, the light orange and orange clusters in DAPC analysis. Boldface labels indicate pairs connected by gene flow, as indicated by the sign of D. Time period represents the time scale over which gene flow is measured, as described in Figure S6. **(B) Timing of gene flow.** For all pairs of lineages belonging to triplets that yielded D values significantly different from zero, the corresponding time period over which gene flow was measured (as defined in Figure S6) is indicated. Each pair belongs to multiple triplets, spanning different time periods, and the reported time period is therefore the consensus of the corresponding time scales. For instance, the pair (brachiaria2, eleusine1) was included in triplets measuring gene flow at both intermediate (time periods 1 and 2) and recent (time period 1) time scales, and the consensus time period is therefore "1+2".

Table S3. Summary statistics of genome assemblies used in this study.

Supplementary Figure legends

Figures S1A-S1J Maximum likelihood tree based on *MPG1, ACT1, b-tubulin1, BAC6, CAL, CH7BAC7, CH7BAC9, CHS, Ef1a* and *NUT1* marker, respectively. Trees are represented as unrooted cladograms. Dark branches represent branches with bootstrap support > 50 after 100 bootstrap replicates (corresponding support are indicated). Clades are labeled according to the convention used by Castroagudin et al. [24]. Green dots: representatives of Pgt (*Pyricularia graminis-tritici sp. nov.*). Red dots: representatives of Pot (*Pyricularia oryzae* pathotype Triticum) clade 1. Blue dots: representatives of Pot clade 2. Orange dots: representatives of Poo (*Pyricularia oryzae* pathotype Oryza).

Figure S2. Species tree inference based on the dataset of Castroagudin et al. [24] using ASTRAL. The tree is represented as an unrooted cladogram.

Multilocus bootstrap supports above 50 are indicates above branches. Dark branches represent branches with bootstrap support > 50 after 100 bootstrap replicates (corresponding support are indicated). Number in brackets are q1 local quartet supports (i.e. the proportion of quartet trees in individual genealogies that agree with the topology recovered by the ASTRAL analysis around the branch). Clades are labeled according to the convention used by Castroagudin et al. [24] as in S1-S10 Figures.

Figure S3. Analyses of population subdivision using clustering algorithms. (A) Bayesian Information Criterion vs number of clusters assumed in DAPC analysis. The Bayesian Information Criterion assesses the fit of models of population structure assuming

different K values. **(B) DAPC analysis of population subdivision, assuming K=2 to K=15 clusters.** Each isolate is represented by a thick vertical line divided into K segments that represent the isolate's estimated membership probabilities in the K clusters. The host of origin of samples is shown below the barplot. **(C) Log likelihood of data vs number of clusters assumed in STRUCTURE analysis.** Error bars are standard deviations of likelihood across STRUCTURE repeats. **(D) STRUCTURE analysis of population subdivision, assuming K=2 to K=15 clusters.** Each isolate is represented by a thick vertical line divided into K segments that represent the isolate's estimated membership proportions in the K clusters (note that two to seven clusters are empty, i.e. represented by no isolates, for models with K>9). The host of origin of samples is shown below the barplot.

Figure S4. Neighbor joining tree showing the genetic distance separating the *M. oryzae* strains from *M. grisea* and *M. pennisetigena*. Distances are in SNPs/kb.

Figure S5. Distribution of MPG1, BAC6, β-tubulin1, CAL, CH7BAC7, CH7BAC9, EF1α and NUT1 alleles among *M. oryzae* isolates as indicated by mapping onto the neighbor-joining tree built using whole genome SNP data. Alleles were identified by using a reference marker sequence for each gene, to search all the genomes using BLAST. Sequence variants are noted above each tree using the BLAST backtrace operations (BTOP) format.

Figure S6. Time periods for gene flow covered by different triplets of lineages in ABBA/BABA tests. Heterogeneity in divergence time between members of ((P1,P2),P3) triplets allows to examine gene flow at three time scales [33]: (A) triplets including the most recently diverged lineages as P1 and P2 (i.e. the *Triticum* and *Lolium* lineages, the two

Eleusine lineages, or the *Oryza* and *Setaria* lineages) carry information about gene flow across relatively recent times, (B) triplets including, as P1 and P2, two lineages from the same main group of lineages (i.e. *Eragrostis* / *Eleusine*1 / *Eleusine*2 / *Triticum* / *Lolium* or *Brachiaria2* / *Setaria* / *Oryza*, excluding (P1,P2) pairs already used in (A)) carry information about gene flow across intermediate times, and (C) triplets including, as P1 and P2, two lineages from different groups of lineages (i.e. *Eragrostis* / *Eleusine*1 / *Eleusine*2 / *Triticum* / *Lolium* and *Brachiaria2* / *Setaria* / *Oryza*) and *Stenotaphrum* or *Brachiaria1* as P3 carry information about gene flow across a relatively long time period. All three graphs correspond to hypothetical cases in which the D statistic that measures imbalance between ABBA and BABA types indicates gene flow between P2 and P3 (i.e. positive D values). In (A) and (B) multiple possible topologies are shown, as P1, P2 and P3 can either belong to the same group of lineages or to different groups of lineages.

Supplementary Text legends

Text S1. Probabilistic "chromosome painting" analyses.

References

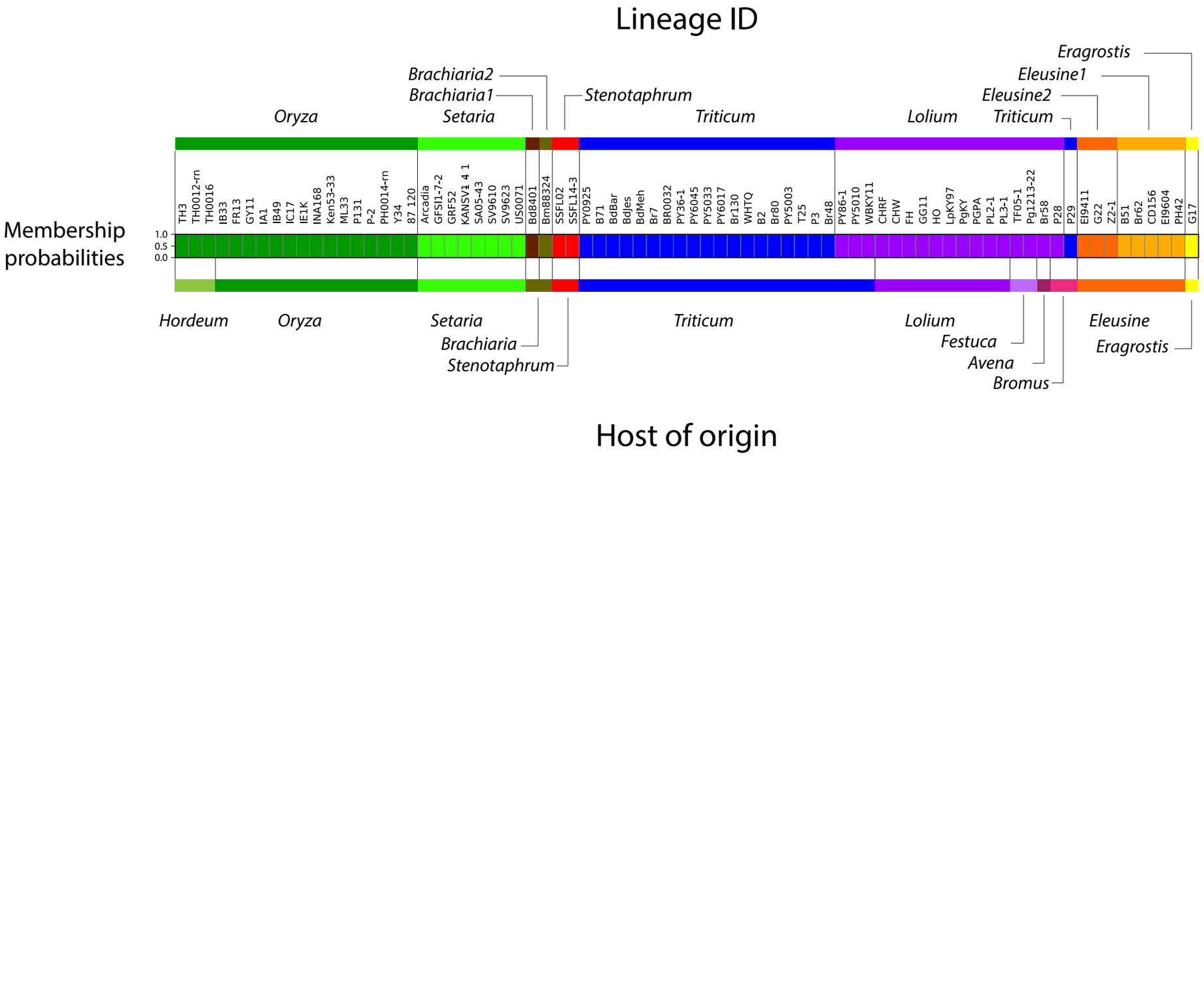
1. Fournier E, Giraud T, Albertini C, Brygoo Y. Partition of the *Botrytis cinerea* complex in France using multiple gene genealogies. Mycologia. 2005;97(6):1251-67. Epub 2006/05/26. PubMed PMID: 16722218.
2. Le Gac M, Hood ME, Fournier E, Giraud T. Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. Evolution. 2007;61(1):15-26. doi: 10.1111/j.1558-5646.2007.00002.x. PubMed PMID: WOS:000244285000003.
3. Taylor JW, Fisher MC. Fungal multilocus sequence typing--it's not just for bacteria. Current opinion in microbiology. 2003;6(4):351-6. Epub 2003/08/28. PubMed PMID: 12941403.
4. Planet PJ, Narechania A, Chen L, Mathema B, Boundy S, Archer G, et al. Architecture of a Species: Phylogenomics of *Staphylococcus aureus*. Trends in microbiology. 2017;25(2):153-66. Epub 2016/10/19. doi: 10.1016/j.tim.2016.09.009. PubMed PMID: 27751626.
5. Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, et al. Genomics and the origin of species. Nature reviews Genetics. 2014;15(3):176-92. Epub 2014/02/19. doi: 10.1038/nrg3644. PubMed PMID: 24535286.
6. 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. Epub 2010/05/04. doi: 10.1016/j.tree.2010.03.006. PubMed PMID: 20434790; PubMed Central PMCID: PMCPMC2885483.
7. Giraud T, Villareal LM, 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. Epub 2008/10/24. doi: 10.1094/phyto-96-0280. PubMed PMID: 18944443.
8. Klaubauf S, Tharreau D, Fournier E, Groenewald JZ, Crous PW, de Vries RP, et al. Resolving the polyphyletic nature of *Pyricularia* (Pyriculariaceae). Studies in mycology. 2014;79:85-120. Epub 2014/12/11. doi: 10.1016/j.simyco.2014.09.004. PubMed PMID: 25492987; PubMed Central PMCID: PMCPMC4255532.
9. Ou SH. A look at worldwide rice blast disease control. Plant disease. 1980;64(5):439-45.
10. Ou SH. Rice Diseases. Ou SH, editor. Slough, UK: Commonwealth Agricultural Bureaux; 1985.
11. Couch BC, Fudal I, Lebrun MH, 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. Epub 2005/04/02. doi: 10.1534/genetics.105.041780. PubMed PMID: 15802503; PubMed Central PMCID: PMCPMC1450392.
12. Farman ML. *Pyricularia grisea* isolates causing gray leaf spot on perennial ryegrass (*Lolium perenne*) in the United States: relationship to *P. grisea* isolates from other host plants. Phytopathology. 2002;92. doi: 10.1094/phyto.2002.92.3.245.
13. 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. Epub

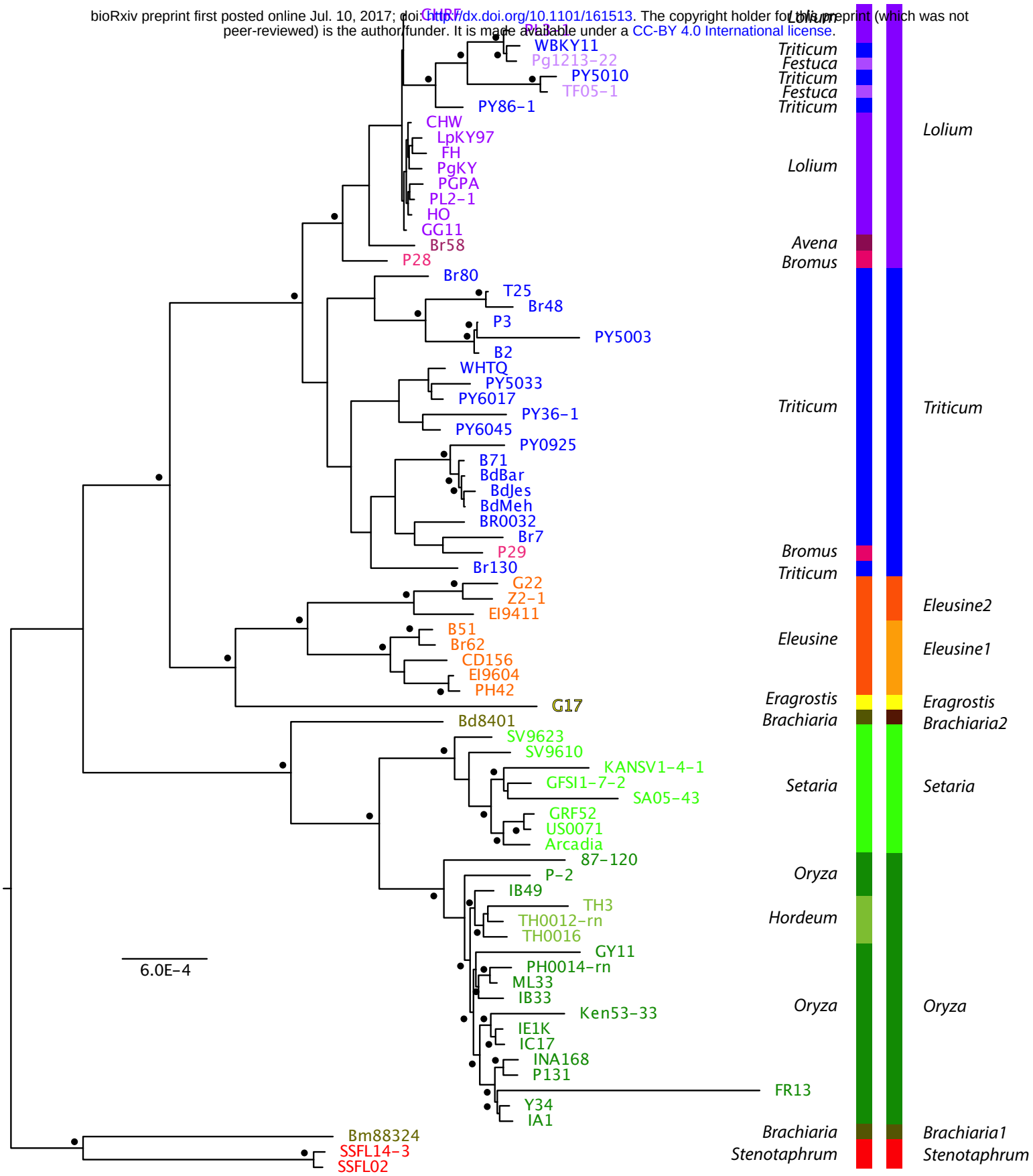
2015/10/11. doi: 10.1093/gbe/evv187. PubMed PMID: 26454013; PubMed Central PMCID: PMCPMC4684704.
14. Yoshida K, Saunders DG, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, et al. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics. 2016;17:370. doi: 10.1186/s12864-016-2690-6. PubMed PMID: 27194050; PubMed Central PMCID: PMCPMC4870811.
15. 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 Biol. 2016;14(1):84. doi: 10.1186/s12915-016-0309-7. PubMed PMID: 27716181; PubMed Central PMCID: PMCPMC5047043.
16. Igarashi S. Update on wheat blast (Pyricularia oryzae) in Brazil. In: Saunders D, editor. Proceedings of the International Conference – Wheat for the Nontraditional Warm Areas. Mexico DF, Mexico: CIMMYT; 1990.
17. Cabrera M, Gutiérrez S. Primer registro de *Pyricularia grisea* en cultivos de trigo del NE de Argentina. Depto2007. Available from: agr.unne.edu.ar/ Extension/Res2007/SanVegetal/SanVegetal_06.pdf.
18. Viedma L. Wheat blast occurrence in Paraguay. Phytopathology. 2005;95:S152.
19. Kohli MM, Mehta YR, Guzman E, De Viedma L, Cubilla LE. *Pyricularia* blast – a threat to wheat cultivation. Czech J Genet Plant Breed. 2011;47:S130-4.
20. Maciel JLN, Ceresini PC, Castroagudin VL, Zala M, Kema GHJ, McDonald BA. Population structure and pathotype diversity of the wheat blast pathogen *Magnaporthe oryzae* 25 years after its emergence in Brazil. Phytopathology. 2014;104. doi: 10.1094/phyto-11-12-0294-r.
21. Cruz CD, Valent B. Wheat blast disease: danger on the move. Tropical Plant Pathology. 2017:1-13.
22. Farman ML, Peterson GL, Chen L, Starnes JH, Valent B, Bachi P, et al. The Lolium pathotype of *Magnaporthe oryzae* recovered from a single blasted wheat plant in the United States. Plant Dis. 2017;101:684-92.
23. Malaker PK, Barma NCD, Tiwari TP, Collis WJ, Duveiller E, Singh PK, et al. First report of wheat blast caused by *Magnaporthe oryzae* pathotype *triticum* in Bangladesh. Plant Dis. 2016;100(11):2330-. doi: 10.1094/PDIS-05-16-0666-PDN.
24. Castroagudin VL, Moreira SI, Pereira DAS, Moreira SS, Brunner PC, Maciel JLN, et al. Wheat blast disease caused by *Pyricularia graminis-tritici* sp. nov. Persoonia. 2016;37.
25. Dettman JR, Jacobson DJ, Taylor JW. A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. Evolution; international journal of organic evolution. 2003;57(12):2703-20. Epub 2004/02/06. PubMed PMID: 14761051.
26. Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, et al. Phylogenetic species recognition and species concepts in fungi. Fungal genetics and biology : FG & B. 2000;31(1):21-32. Epub 2000/12/16. doi: 10.1006/fgbi.2000.1228. PubMed PMID: 11118132.
27. Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics. 2014;30(17):i541-8. doi: 10.1093/bioinformatics/btu462. PubMed PMID: 25161245; PubMed Central PMCID: PMCPMC4147915.

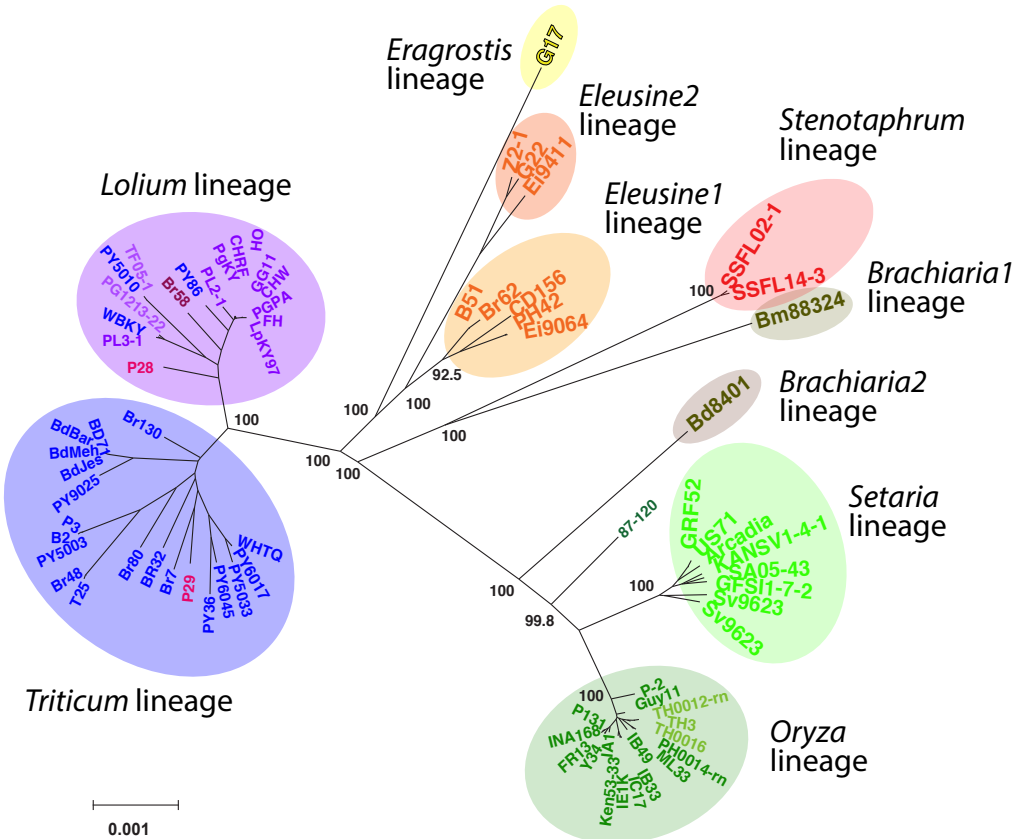
28. Mirarab S, Warnow T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 2015;31(12):i44-52. doi: 10.1093/bioinformatics/btv234. PubMed PMID: 26072508; PubMed Central PMCID: PMCPMC4765870.
29. Sayyari E, Mirarab S. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. Mol Biol Evol. 2016;33(7):1654-68. doi: 10.1093/molbev/msw079. PubMed PMID: 27189547; PubMed Central PMCID: PMCPMC4915361.
30. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC genetics. 2010;11:94. doi: 10.1186/1471-2156-11-94. PubMed PMID: 20950446; PubMed Central PMCID: PMCPMC2973851.
31. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution. 2005;23(2):254-67.
32. Edwards SV, Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, et al. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Molecular phylogenetics and evolution. 2016;94:447-62.
33. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 2013;23(11):1817-28. doi: 10.1101/gr.159426.113. PubMed PMID: PMC3814882.
34. Martin SH, Davey JW, Jiggins CD. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. Molecular biology and evolution. 2014;32(1):244-57.
35. Durand EY, Patterson N, Reich D, Slatkin M. Testing for Ancient Admixture between Closely Related Populations. Molecular Biology and Evolution. 2011;28(8):2239-52.
36. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003;164(4):1567-87. PubMed PMID: 12930761; PubMed Central PMCID: PMCPMC1462648.
37. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945-59. PubMed PMID: 10835412; PubMed Central PMCID: PMCPMC1461096.
38. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. Molecular ecology resources. 2009;9(5):1322-32.
39. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. PLoS Genet. 2012;8(1):e1002453.
40. Kato H, Yamamoto M, Yamaguchi-Ozaki T. Pathogenicity, mating ability and DNA restriction fragment length polymorphisms of *Pyricularia* populations isolated from Gramineae, Bambusideae and Zingiberaceae plants. J Gen Plant Pathol. 2000;66. doi: 10.1007/pl00012919.
41. Urashima AS, Igarashi S, Kato H. Host range, mating type and fertility of *Pyricularia grisea* from wheat in Brazil. Plant Disease. 1993;77. doi: 10.1094/pd-77-1211.
42. Orbach MJ, Chumley FG, Valent B. Electrophoretic karyotypes of Magnaporthe grisea pathogens of diverse grasses. MPMI-Molecular Plant Microbe Interactions. 1996;9(4):261-71.

43. Thi T, Vy P, Hyon G-s, Thi N, Nga T, Inoue Y, et al. Genetic analysis of host-pathogen incompatibility between Lolium isolates of Pyriculariaoryzae and wheat. Journal of General Plant Pathology: JGPP. 2014;80(1):59.
44. Nosil P, Vines TH, Funk DJ. Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution. 2005;59(4):705-19. PubMed PMID: ISI:000228734300001.
45. Kato H, Yamamoto M, Yamaguchi-Ozaki T, Kadouchi H, Iwamoto Y, Nakayashiki H, et al. Pathogenicity, mating ability and DNA restriction fragment length polymorphisms of Pyricularia populations isolated from Gramineae, Bambusideae and Zingiberaceae plants. Journal of General Plant Pathology. 2000;66(1):30-47.
46. 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. Epub 2008/10/24. doi: 10.1094/phyto.2002.92.11.1182. PubMed PMID: 18944243.
47. Tosa Y, Tamba H, Tanaka K, Mayama S. Genetic analysis of host species specificity of Magnaporthe oryzae isolates from rice and wheat. Phytopathology. 2006;96(5):480-4.
48. Heath MC, Valent B, Howard RJ, Chumley FG. Interactions of two strains of Magnaporthe grisea with rice, goosegrass, and weeping lovegrass. Canadian Journal of Botany. 1990;68(8):1627-37.
49. Gurr S, Samalova M, Fisher M. The rise and rise of emerging infectious fungi challenges food security and ecosystem health. Fungal Biol Rev. 2011;25(4):181-8. doi: https://doi.org/10.1016/j.fbr.2011.10.004.
50. Silué D, Notteghem JL. Production of perithecia of *Magnaporthe grisea* on rice plants. Mycol Res. 1990;94(8):1151-2.
51. 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. Epub 2012/02/09. doi: 10.1111/j.1365-294X.2012.05469.x. PubMed PMID: 22313491.
52. 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. The New phytologist. 2016;209(3):1220-9. Epub 2015/10/03. doi: 10.1111/nph.13658. PubMed PMID: 26428268.
53. Hibbett DS, Taylor JW. Fungal systematics: is a new age of enlightenment at hand? Nature reviews Microbiology. 2013;11(2):129-33. Epub 2013/01/05. doi: 10.1038/nrmicro2963. PubMed PMID: 23288349.
54. Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, et al. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen Magnaporthe oryzae. The Plant cell. 2009;21(5):1573-91. Epub 2009/05/21. doi: 10.1105/tpc.109.066324. PubMed PMID: 19454732; PubMed Central PMCID: PMCPMC2700537.
55. Xue M, Yang J, Li Z, Hu S, Yao N, Dean RA, et al. Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. PLoS Genet. 2012;8(8):e1002869. doi: 10.1371/journal.pgen.1002869. PubMed PMID: 22876203; PubMed Central PMCID: PMCPMC3410873.
56. Zhong Z, Norvienyeku J, Chen M, Bao J, Lin L, Chen L, et al. Directional selection from host plants is a major force driving host specificity in *Magnaporthe* species. Sci Rep.

2016;6:25591. doi: 10.1038/srep25591. PubMed PMID: 27151494; PubMed Central PMCID: PMCPMC4858695.
57. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research. 2008;18(5):821-9.
58. Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics. 2011;27(6):757-63. Epub 2011/01/11. doi: 10.1093/bioinformatics/btr010. PubMed PMID: 21216780.
59. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011;12:124. Epub 2011/04/30. doi: 10.1186/1471-2105-12-124. PubMed PMID: 21526987; PubMed Central PMCID: PMCPMC3114741.
60. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature. 2005;434. doi: 10.1038/nature03449.
61. Ranwez V, Harispe S, Delsuc F, Douzery EJ. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PloS one. 2011;6(9):e22594. Epub 2011/09/29. doi: 10.1371/journal.pone.0022594. PubMed PMID: 21949676; PubMed Central PMCID: PMCPMC3174933.
62. De Mita S, Siol M. EggLib: processing, analysis and simulation tools for population genetics and genomics. BMC genetics. 2012;13:27. Epub 2012/04/13. doi: 10.1186/1471-2156-13-27. PubMed PMID: 22494792; PubMed Central PMCID: PMCPMC3350404.
63. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3. Epub 2014/01/24. doi: 10.1093/bioinformatics/btu033. PubMed PMID: 24451623; PubMed Central PMCID: PMCPMC3998144.
64. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389-402. PubMed PMID: 9254694; PubMed Central PMCID: PMCPMC146917.
65. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics. 1998;14(1):68-73. PubMed PMID: ISI:000074405900008.
66. Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. Mol Biol Evol. 2011;28(8):2239-52. doi: 10.1093/molbev/msr048. PubMed PMID: 21325092; PubMed Central PMCID: PMCPMC3144383.
67. Auton A, McVean G. Recombination rate estimation in the presence of hotspots. Genome research. 2007;17(8):1219-27.
68. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004;20(2):289-90. PubMed PMID: 14734327.
69. Pieck ML, Ruck A, Farman ML, Peterson GL, Stack JP, Valent B, et al. Genomics-based marker discovery and diagnostic assay development for wheat blast. Plant Disease. 2017;101(1):103-9.
70. 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.

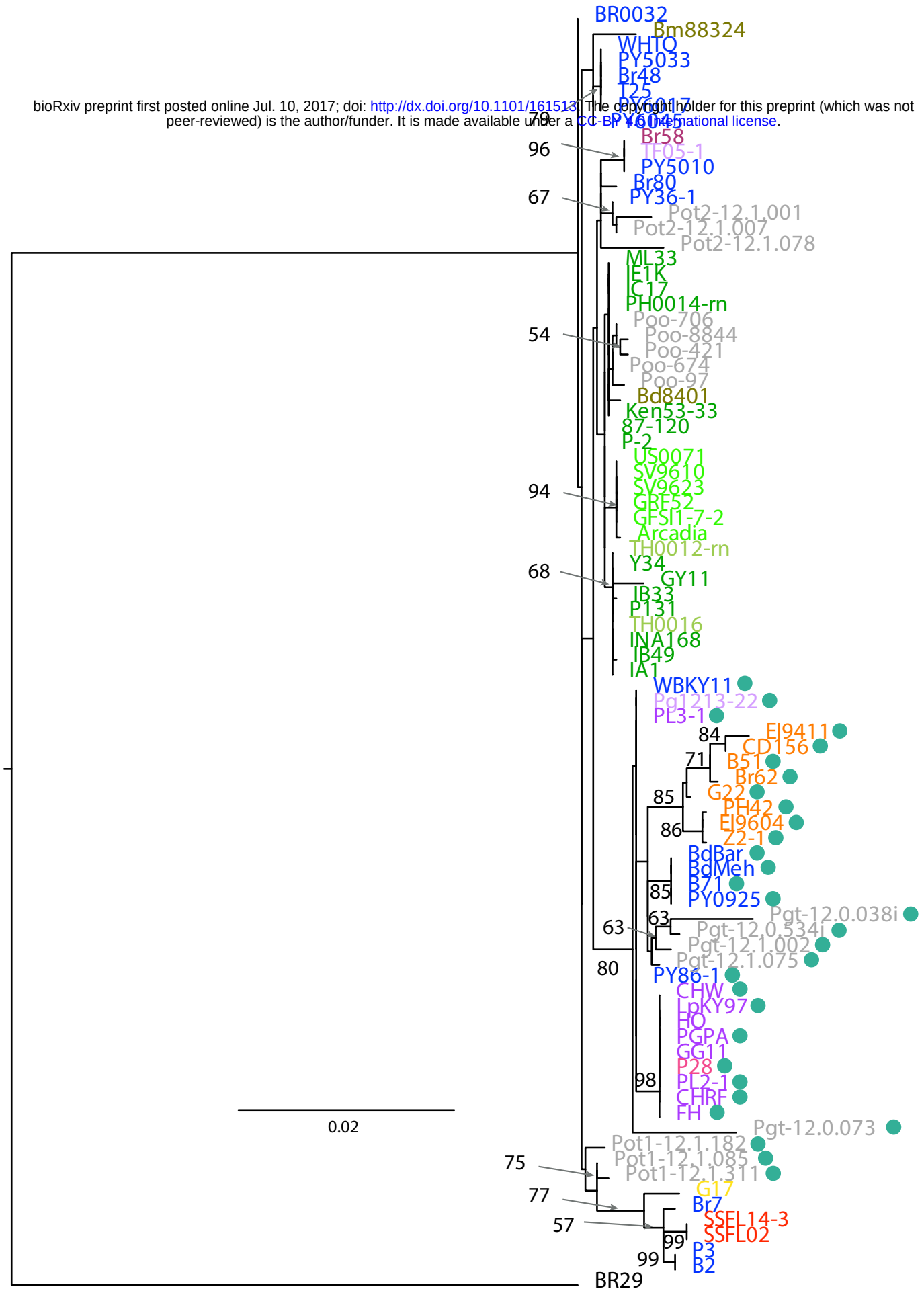

Lineage ID
Eragrostis
Eleusine1
Eleusine2
Triticum
Brachiaria2
Brachiaria1
Stenotaphrum
Oryza
Setaria
Triticum
Lolium
Membership probabilities
1.0
0.5
0.0
TH3
TH0012-rn
TH0016
IB33
FR13
GY11
IA1
IB49
IC17
IE1K
INA168
Ken53-33
ML33
P131
P-2
PH0014-rn
Y34
87 120
Arcadia
GFSI1-7-2
GRF52
KANSV1 4 1
SA05-43
SV9610
SV9623
US0071
Bd8401
Bm88324
SSFL02
SSFL14-3
PY0925
B71
BdBar
BdJes
BdMeh
Br7
BR0032
PY36-1
PY6045
PY5033
PY6017
Br130
WHTQ
B2
Br80
PY5003
T25
P3
Br48
PY86-1
PY5010
WBKY11
CHRF
CHW
FH
GG11
HO
LpKY97
PgKY
PGPA
PL2-1
PL3-1
TF05-1
Pg1213-22
Br58
P28
P29
EI9411
G22
Z2-1
B51
Br62
CD156
EI9604
PH42
G17
Hordeum
Oryza
Setaria
Triticum
Lolium
Eleusine
Brachiaria
Stenotaphrum
Festuca
Avena
Bromus
Eragrostis
Host of origin





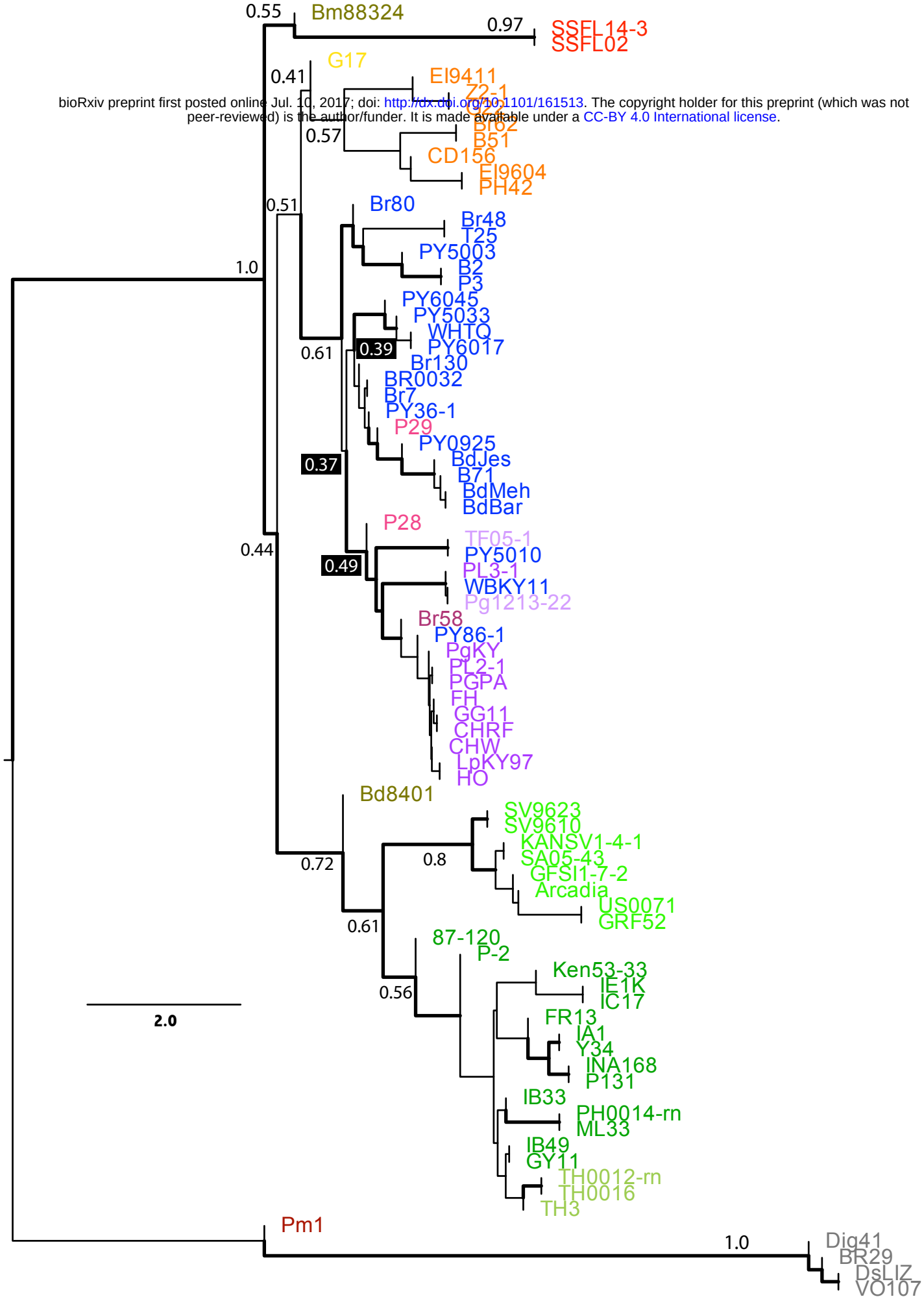
Hosts of origin (tip colors)

Lolium	*Avena*	*Hordeum*	*Stenotaphrum*
Festuca	*Triticum*	*Setaria*	*Eragrostis*
Bromus	*Oryza*	*Brachiaria*	*Eleusine*



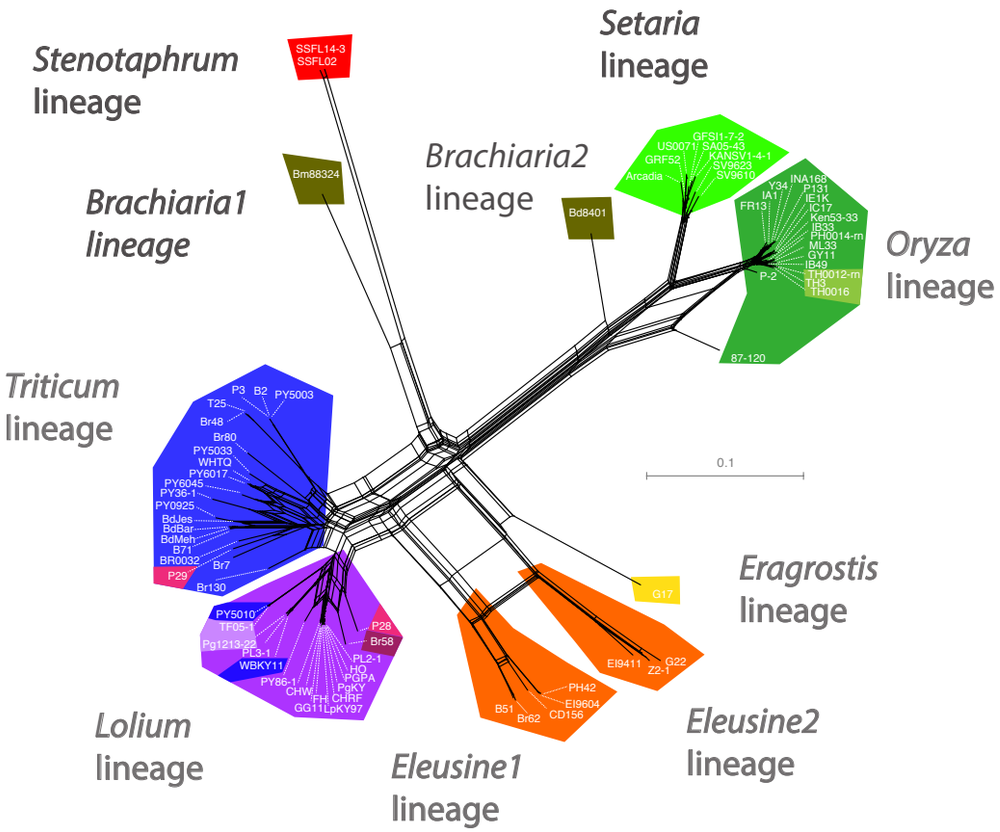
Hosts of origin (tip colors; samples from Castroagudin et al. in grey)

Lolium	*Avena*	*Hordeum*	*Stenotaphrum*
Festuca	*Triticum*	*Setaria*	*Eragrostis*
Bromus	*Oryza*	*Brachiaria*	*Eleusine*



Hosts of origin (tip colors)

Lolium	*Avena*	*Hordeum*	*Stenotaphrum*	Outgroups:
Festuca	*Triticum*	*Setaria*	*Eragrostis*	*Pennisetum*
Bromus	*Oryza*	*Brachiaria*	*Eleusine*	*Digitaria*



Hosts of origin (color of shaded areas)

Lolium	*Avena*	*Hordeum*	*Stenotaphrum*
Festuca	*Triticum*	*Setaria*	*Eragrostis*
Bromus	*Oryza*	*Brachiaria*	*Eleusine*