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], 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, 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

81 important disease of rice (*Oryza sativa*), but it also causes blast disease on more than 50
82 cultivated and wild monocot plant species [9]. This includes other cereal crops such as
83 wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), finger millet (*Eleusine coracana*), and
84 foxtail millet (*Setaria italica*, *S. viridis*), as well as wild and cultivated grass hosts including
85 goosegrass (*Eleusine indica*), annual ryegrass (*Lolium multiflorum*), perennial ryegrass (*L.*
86 *perenne*), tall fescue (*Festuca arundinacea*) and St. Augustine grass (*Stenotaphrum*
87 *secundatum*) [10]. Previous studies based on multilocus sequence typing showed that *M.*
88 *oryzae* is subdivided into multiple clades, each found on only a limited number of host
89 species, with pathogenicity testing revealing host-specificity as a plausible driver of genetic
90 divergence [11, 12]. More recently, comparative genomics of eight isolates infecting wheat,
91 goosegrass, rice, foxtail millet, finger millet and barley revealed deep subdivision of *M.*
92 *oryzae* into three groups infecting finger millet or wheat, foxtail millet, and rice or barley
93 [13, 14]. Subsequent analysis of genomic data from nine wheat-infecting isolates, two
94 ryegrass infecting isolates, and one weeping lovegrass-infecting isolate subdivided lineages
95 infecting wheat only on the one hand and wheat or ryegrass on the other hand, and
96 revealed an additional lineage associated with the weeping lovegrass strain [15]. Together,
97 these studies suggest a history of host-range expansion or host-shifts and limited gene flow
98 between lineages within *M. oryzae*.

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

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

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

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

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

135 Results

136 Re-assessing the validity of the proposed *P. graminis-tritici* species by analyzing the 137 original published data according to Phylogenetic Species Recognition by 138 Genealogical Concordance (GCPSR)

139 To test the previous delineation of a subgroup of wheat-infecting isolates as a new
140 phylogenetic species, we re-analyzed the Castroagudin et al. dataset [24], which mostly
141 included sequences from Brazilian isolates. However, instead of using bootstrap support in
142 a total evidence genealogy inferred from concatenated sequences for species delineation,
143 we properly applied the GCPSR test [25, 26]. This test identifies a group as an independent
144 evolutionary lineage (i.e. phylogenetic species) if it satisfies two conditions: (1)
145 Genealogical concordance: the group is present in the majority of the single-locus
146 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, treatment 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 data for strains collected on rice (*Oryza sativa*), finger
166 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*). Also included were representatives of
169 previously unstudied host-specialized populations from foxtails (*Setaria* sp.), St. Augustine

170 grass (*Stenotaphrum secundatum*), weeping lovegrass (*Eragrostis curvula*), signalgrass
171 (*Brachiaria* sp.), panicgrass (*Bromus tectorum*) and oat (*Avena sativa*). Analyses were based
172 on 2,682 orthologous coding sequences that were single-copy in all *M. oryzae* genomes (in
173 total ~6.6 Mb of sequence data), and from whole-genome SNPs identified from pairwise
174 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); (2)
181 isolates from *Setaria* sp. (light green); (3) isolate Bd8401 from *Brachiaria distachya*
182 (brown); (4) isolate Bm88324 from *Brachiaria mutica* (olive); (5) isolates from
183 *Stenotaphrum* (red); (6) 17 of the 22 isolates from wheat and an isolate from *Bromus*
184 (blue); (7) the remaining 3/22 isolates from wheat together with isolates from Lolium,
185 Festuca, oat and a second isolate from Bromus (purple); (8 & 9) isolates from *Eleusine* that
186 formed two distinct clusters (orange and light orange); and (10) an isolate from *Eragrostis*
187 (yellow) (Figure 1). Increasing K mostly resulted in further subdivision among the isolates
188 from wheat, rice and *Lolium* sp. The discovery of three wheat blast isolates that grouped
189 with the Festuca-Lolium pathogens in cluster 7 was important because it supports the idea
190 that wheat-infecting isolates belong to at least two distinct populations.

191 Next, we inferred gene genealogies using maximum-likelihood and distance-based methods.
192 Both approaches produced trees that corresponded well with the subdivisions identified in
193 DAPC. The tree generated using maximum likelihood (ML) analysis of orthologous genes
194 displayed a topology with eight well-supported lineages (Figure 2). Six of these lineages
195 had 100% bootstrap support and showed one-to-one correspondence with the K clusters
196 from DAPC (Figure 1). Another well-supported lineage comprised sister clades made up of
197 the *Eleusine* and *Eragrostis* pathogens (Figure 1; Figure S3). The final lineage corresponded
198 perfectly with the "blue" DAPC cluster which contained the 17 isolates from wheat and
199 isolate P29 from *Bromus*. This last lineage had poor bootstrap support (50%), however,
200 and showed evidence of further population subdivision among the constituent isolates (i.e.
201 high confidence subclades).

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

206 **Levels of polymorphism within and divergence between lineages/species**

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

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

215 Genome-wide nucleotide divergence was one order of magnitude higher between *M. oryzae*
216 and its closest relatives, *M. grisea* and *M. pennisetigena*, than it was among isolates within
217 *M. oryzae.* The maximum pairwise distance (SNPs per kilobase) between any two *M. oryzae*
218 isolates was less than 1%, genome-wide (Figure S4; Table S1), compared with *M. oryzae* vs
219 *M. grisea*, *M. oryzae* vs. *M. pennisetigena,* or *M. grisea* vs *M. pennisetigena*, all of which were
220 consistently greater than 10%. The low level of genetic divergence among *M. oryzae*
221 isolates, compared with that observed against isolates with established separate species
222 status, provides good evidence against the relatively ancient cryptic species within
223 *M. oryzae*. (Table S1).

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

225 While the 10 loci utilized in the Castroagudin et al. [24] study do not support the *Pgt* split
226 based on GCPSR criteria, our DAPC and whole genome ML and NJ analyses supported the
227 partitioning of wheat blast isolates into two, genetically-distinct groups: one consisting
228 almost exclusively of wheat-infecting isolates, the other comprising largely *Festuca*- and
229 *Lolium*-infecting isolates as well as few wheat-infecting isolates (Figure 2; Figure 3).
230 Unfortunately, the Castroagudin et al. study did not include *Festuca*- and *Lolium*-infecting
231 isolates and genome sequences are not available for their strains. Therefore, to test for
232 possible correspondence between the proposed *Pgt* species and the *Festuca-Lolium* group
233 (or indeed the *Triticum* group), we extended the 10 loci analysis to the *M. oryzae* genome
234 sequences used in the present study. For reference, we included the multilocus data for 16

235 Castroagudin et al. strains representing all the major clades from that study. Nine of the 10
236 loci were successfully recovered from 68 of our *M. oryzae* genome sequences. The
237 remaining locus, CH7-BAC9, was absent from too many genome assemblies and, as a result,
238 was excluded from the analysis.

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

255 The basis for the erroneous designation of *Pgt* as a novel species was clearly revealed when
256 *MPG1* alleles were mapped onto the ML and NJ trees because there was a remarkable lack

257 of correspondence between allelic distribution and the phylogenetic groupings determined
258 using whole genome evidences. The *Pgt*-type allele was distributed among three different
259 clades but showed a discontinuous distribution. As an example, the *Triticum* group
260 contained three *MPG1* alleles. Two of these (including the *Pgt*-type) were also present in
261 the *Festuca-Lolium* clade, while the third was shared by distantly-related isolates from
262 *Stenotaphrum* (Figure S5). The *Eleusine* group also contained the *Pgt* allele and two other
263 variants, while the *Setaria* and *Oryza* pathogens possessed an *MPG1* allele distinct from all
264 the others. Overall, the distribution of *MPG1* alleles points to the occurrence of incomplete
265 lineage sorting and/or gene flow during *M. oryzae* diversification. Importantly, at least six
266 of the other markers studied by Castroagudin et al. showed similar discontinuities in their
267 distributions (Figure S6). In addition, two of the Castroagudin et al. markers (*ACT1* and
268 *CHS1*) showed no sequence variation among the 68 *M. oryzae* isolates analyzed in the
269 present study (data not shown).

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

271 The total evidence genealogies generated using distance-based and maximum likelihood
272 approaches on 2,682 single-copy orthologs and 76 *M. oryzae* genomes (Figure 2; Figure 3)
273 were highly concordant in terms of lineage composition and branching order. However,
274 concatenation methods can be positively misleading, as they assume that all gene trees are
275 identical in topology and branch lengths and they do not explicitly model the relationship
276 between the species tree and gene trees [32]. To estimate the "species tree" and to re-
277 assess previous findings of cryptic species within *M. oryzae*, we therefore used a
278 combination of species inference using the multispecies coalescent method implemented in
279 ASTRAL [27-29] and a new implementation of the GCPSR that can handle genomic data.

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

294 As a formal test for the presence of cryptic species within *M. oryzae*, we applied the
295 phylogenetic species recognition criteria to the set of 2,241 single-copy orthologous genes
296 using an implementation of the GCPSR scalable to any number of loci. Applying the GCPSR
297 following the non-discordance criterion of Dettman et al. (a clade has to be well supported
298 by at least one single-locus genealogy and not contradicted by any other genealogy at the
299 same level of support; [25]) resulted in the recognition of no species within *M. oryzae*.

300 **Admixture and gene flow between lineages**

301 The existence of gene flow and/or incomplete lineage sorting was also supported by
302 phylogenetic network analysis. We used the network approach neighbor-net implemented
303 in SPLITSTREE 4.13 [25] to visualize evolutionary relationships, while taking into account the
304 possibility of recombination within or between lineages. The network inferred from
305 haplotypes identified using the 2,682 single-copy orthologs in the 76 *M. oryzae* strains,
306 showed extensive reticulation connecting all lineages, consistent with recombination or
307 incomplete lineage sorting (Figure 6). To disentangle the role of gene flow versus
308 incomplete lineage sorting in observed network reticulations, we used ABBA/BABA tests,
309 which compare numbers of two classes of shared derived alleles (the ABBA and BABA
310 classes). For three lineages P1, P2 and P3 and an outgroup with genealogical relationships
311 (((P1,P2),P3),O), and under conditions of no gene flow, shared derived alleles between P2
312 and P3 (ABBA alleles) and shared derived alleles between P1 and P3 (BABA alleles) can
313 only be produced by incomplete lineage sorting, and should be equally infrequent [33].
314 Differences in numbers of ABBA and BABA alleles are interpreted as gene flow assuming no
315 recurrent mutation and no deep ancestral population structure within lineages. The D
316 statistic measuring differences in counts of ABBA and BABA alleles was significantly
317 different from zero (Z-score>3) in 66 of 68 lineage triplets, consistent with a history of
318 gene flow between lineages within *M. oryzae* (Table S2).

319 We then used the program STRUCTURE [34-36] to gain further insight into the possibility of
320 recent admixture between lineages (Figure S7). STRUCTURE uses Markov chain Monte Carlo
321 simulations to infer the assignment of genotypes into K distinct clusters, minimizing
322 deviations from Hardy–Weinberg and linkage disequilibria within each cluster. STRUCTURE
323 analysis provided evidence for admixture at all K values (Figure S7), suggesting that recent

324 admixture events have recently shaped patterns of population subdivision within *M. oryzae*.
325 The patterns of clustering inferred with STRUCTURE were largely similar to those inferred
326 with DAPC.

327 Discussion

328 Population subdivision but no cryptic phylogenetic species

329 Using population- and phylogenomic analyses of single-copy orthologous genes and whole-
330 genome SNPs identified in *M. oryzae* genomes from multiple cereal and grass hosts, we
331 provide evidence that *M. oryzae* is subdivided in multiple lineages preferentially associated
332 with one host genus. Neither the re-analysis of previous data, nor the analysis of new data
333 using previous phylogenetic species recognition markers, confirmed the existence of a
334 wheat blast-associated species called *P. graminis-tritici* [24]. Marker *MPG1*, which holds
335 most of the divergence previously detected, does not stand as a diagnostic marker of the
336 wheat-infecting lineage of *M. oryzae* when tested in other lineages. Improper species
337 identification of cryptic species also stems from the fact that available information on *M.*
338 *oryzae* diversity has been insufficiently taken into account. In particular, isolates from the
339 lineages most closely related to wheat strains (i.e. the *Lolium* lineage; [11, 12, 15, 22]) were
340 not represented in species identification work [24]. Hence, phylogenetic species
341 recognition by genealogical concordance did not identify cryptic phylogenetic species and
342 thus *M. oryzae* is not, strictly speaking, a species complex. As a consequence, Pgt should be
343 dismissed as a novel species.

344 Incipient speciation by ecological specialization following host-shifts

345 Several features of the life cycle of *M. oryzae* are conducive to speciation by ecological

346 specialization following host shifts, suggesting that the observed pattern of population
347 subdivision in *M. oryzae* actually corresponds to ongoing speciation. Previous experimental
348 measurements of perithecia formation and ascospore production – two important
349 components of reproductive success- suggested inter-fertility in vitro between most pairs
350 of lineages with high levels of ascospore viability [37-40]. This suggests that intrinsic pre-
351 or post-mating reproductive barriers, such as assortative mating by mate choice or gametic
352 incompatibility, and zygotic mortality, are not responsible for the relative reproductive
353 isolation between lineages – which creates the observed pattern of population subdivision.
354 Instead, the relative reproductive isolation between lineages could be caused by one or
355 several pre- or post-mating barriers (Table 1 in [41]), such as mating-system isolation or
356 hybrid sterility (intrinsic barrier), or difference in mating times, difference in mating sites,
357 immigrant inviability, or ecologically-based hybrid inviability (extrinsic barriers).

358 Previous pathogenicity assays revealed extensive variability in the host range of *M. oryzae*
359 isolates, and both in terms of pathogenicity towards a set of host species or pathogenicity
360 towards a set of genotypes from a given host [38, 42]. Indeed, extensive genetic analyses
361 show that host species specificity in *M. oryzae*, similar to rice cultivar specificity, is
362 controlled by a gene-for-gene relationship in which one to three avirulence genes in the
363 fungus prevent infection of particular host species [40, 43, 44]. Loss of the avirulence genes
364 would allow infection of novel hosts to occur. Additionally, host species specificity is not
365 strictly maintained. Under controlled conditions, most lineages have at least one host in
366 common [42], and strains of within one lineage can still cause rare susceptible lesions on
367 naive hosts [21, 45]. Moreover, a single plant infected by a single genotype can produce
368 large numbers of spores in a single growing season [46], allowing the pathogen to persist

369 on alternative host even if selection is strong, and promoting the rapid and repeated
370 creation of genetic variation [6]. Although some of these features appear to be antagonistic
371 with the possibility of speciation by host-specialization within *M. oryzae*, our finding that
372 the different lineages within *M. oryzae* tend to be sampled on a single host suggests that
373 ecological barriers alone may in fact contribute to reduce gene flow substantially between
374 host-specific lineages.

375 Mating within host (i.e. reproduction between individuals infecting the same host), and to a
376 lesser extent mating system isolation (i.e. lack of outcrossing reproduction), may
377 contribute to further reduce gene flow between *M. oryzae* lineages. The fact that mating in
378 *M. oryzae* likely occurs within host tissues, such as dead stems [47], may participate in the
379 maintenance of the different lineages by decreasing the rate of reproduction between
380 isolates adapted to different hosts [6]. Loss of sexual fertility also appears to have a role in
381 lineage maintenance. The rice lineage, in particular, is single mating-type and female-sterile
382 throughout most of its range, which would reduce the chance of outcrossing sex with
383 members of other lineages [48]. Our analyses rejected the null hypothesis of clonality in all
384 lineages, but they provided no time frame for the detected recombination events.
385 Population-level studies and experimental measurements of mating type ratios and female
386 fertility are needed to assess the reproductive mode of the different lineages within *M.*
387 *oryzae* in the field.

388 **Inter-lineage gene flow**

389 Several potential barriers contribute to reduce genetic exchanges between *M. oryzae*
390 lineages (see previous paragraph), but not completely so, as evidenced by signal of gene

391 flow and admixture detected in our genomic data. We hypothesize that the lack of strict
392 host specialization of the different lineages is a key driver of inter-lineage gene flow. Many
393 of the grass or cereal species that are hosts to *M. oryzae* are widely cultivated as staple
394 crops or widely distributed as pasture or weeds, including "universal suscepts" such as
395 Italian ryegrass, tall fescue and weeping lovegrass [37], increasing the chance for
396 encounters and mating between isolates with overlapping host ranges. These shared hosts
397 may act as a platform facilitating encounters and mating between fertile and compatible
398 isolates from different lineages, thereby enabling inter-lineage gene flow [49]. Plant health
399 vigilance is therefore warranted for disease emergence via recombination in regions where
400 multiple lineages are in contact and shared hosts are present. This is particularly so, given
401 that once infection of novel host has taken place (i.e. host shift or host range expansion),
402 the fungus has the capacity to build inoculum levels very rapidly, facilitating spread of the
403 disease over considerable distances. It is striking, for example, that wheat blast has, within
404 a year, spread from Bangladesh into the West Bengal region of India where it emerged in
405 2017 (openwheatblast.org).

406 **Conclusion**

407 Using a population genomics framework, we show that *M. oryzae* is subdivided into
408 multiple lineages with limited host range and evidence of genetic exchanges between them.
409 Our findings provide greater understanding of the eco-evolutionary factors underlying the
410 diversification of *M. oryzae* and highlight the practicality of genomic data for
411 epidemiological surveillance of its different intraspecific lineages. Reappraisal of species
412 boundaries within *M. oryzae* refuted the existence of a novel cryptic phylogenetic species
413 named *P. graminis-tritici,* underlining that the use of node support in total evidence

414 genealogies based on a limited dataset in terms of number of loci and of range of variation
415 in origin (geography and host) of isolates can lead to erroneously identify fungal cryptic
416 species. Our work illustrates the growing divide between taxonomy that 'creates the
417 language of biodiversity' [50] based on limited sets of characters, and genomic data that
418 reveals more finely the complexity and continuous nature of the lineage divergence process
419 called speciation.

420 Materials and Methods

421 Fungal strains

422 Thirty-eight newly sequenced genomes were analyzed together with 43 published
423 genomes [13, 14, 22, 51-53] resulting in a total of 81 *Magnaporthe* strains, including 76 *M.*
424 *oryzae* genomes representing 12 different hosts available for analysis (Table 1). We also
425 included as outgroups one strain of *Pyricularia pennisetigena* from *Pennisetum* sp. and four
426 strains of *Pyricularia grisea* (aka *Magnaporthe grisea*) from crabgrass (*Digitaria*
427 s*anguinalis*). All newly sequenced strains were single-spored prior to DNA extraction.
428

429 Genome sequencing and assembly

430 New genome data were produced by an international collaborative effort. Characteristics of
431 genome assemblies are summarized in Table S3. For newly sequenced genomes provided
432 by MF and BV, sequences were acquired on a MiSeq machine (Illumina, Inc.). Sequences
433 were assembled using the paired-end mode in NEWBLER V2.9 (Roche Diagnostics,
434 Indianapolis, IN). A custom perl script was used to merge the resulting scaffolds and
435 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 [54] 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 [54].

Orthologous genes identification in genomic sequences

Protein-coding gene models were predicted using AUGUSTUS V3.0.3 [55]. Orthologous genes were identified in the 76-genomes *M. oryzae* or in the dataset including outgroups using PROTEINORTHO [56]. The v8 version of the 70-15 *M. oryzae* reference genome [57] 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 [58]. 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.

Population subdivision, summary statistics of polymorphism and divergence

458 Population subdivision was analyzed using DAPC and STRUCTURE [30, 34-36], based on
459 multilocus haplotype profiles identified from ortholog alignments using a custom Python
460 script. DAPC was performed using the ADEGENET package in R [13]. We retained the first 30
461 principal components, and the first 4 discriminant functions. Ten independent STRUCTURE
462 runs were carried out for each number of clusters K, with 100,000 MCMC iterations after a
463 burn-in of 50,000 steps.
464 Polymorphism statistics were computed using EGGLIB 3.0.0b10 [59] excluding sites with
465 >30% missing data. Divergence statistics were computed using a custom perl script.
466 To infer total evidence trees within the 76 *M. oryzae* strains (respectively within the 81
467 *Magnaporthe* strains), all sequences from the 2,682 (respectively 2,241) orthologous
468 sequences were concatened. The maximum-likelihood tree was inferred using RAxML [60]
469 with the GTR-gamma model, and bootstrap supports were estimated after 1000 replicates.

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

471 The 10 loci used by Castroagudin et al. [24], i.e. actin (ACT), beta-tubulin1 (βT-1),
472 calmodulin (CAL), chitin synthase 1 (CHS1), translation elongation factor 1-alpha (EF1-α),
473 hydrophobin (MPG1), nitrogen regulatory protein 1 (NUT1), and three anonymous
474 markers (CH6, CH7-BAC7, CH7-BAC9), were search in all genomes using BLASTn. Due to
475 heterogeneity in the quality of assemblies, 9 of the 10 loci could be full-length retrieved
476 without ambiguity in 68 out of the 81 available genomes, still representative of the
477 diversity of host plants.
478

479 **Secondary data analysis**

480 Species recognition based on multiple gene genealogies as described by Castroagudin et al.
481 [24] was repeated following the reported methods. The robustness of the Pgt species
482 inference was tested by re-iterating the study, omitting one marker at a time. Individual
483 genealogies were built using RAxML with the GTR-gamma model and 100 bootstrap
484 replicates.
485
486 **Inference of "species tree" using ASTRAL**
487 The ASTRAL method [27, 29] is based on the multi-species coalescent and allows taking
488 into account possible discrepancies among individual gene genealogies to infer the "species
489 tree". Individual genealogies inferred using RAxML with the GTR-gamma model and 100
490 bootstrap replicates, were used as input data for ASTRAL analysis. Local supports around
491 branches were evaluated with 100 multilocus bootstrapping using the bootstrap replicates
492 inferred from each individual gene tree as input data, and with local quartet supports (q1,
493 obtained using the –t option of ASTRAL) that represent the proportion of quartet trees
494 around the branch that agree with the topology of this branch in the species tree.
495
496 **MPG1-based classification**
497 The *MPG1* hydrophobin sequence is described as being diagnostic for the
498 *P. graminis-tritici/M. oryzae* species split [24]. *MPG1* sequences from one of each species
499 (GIs: KU952644.1 for *P. gramini-triticis*, KU952661.1 for *M. oryzae*) were used as BLAST
500 [61] queries to classify isolates as either *P. graminis-tritici* or *M. oryzae.*
501
502 **Signatures of gene flow and/or incomplete lineage sorting**

503 A phylogenetic network was built using SPLITSTREE 4.13 [62], based on the concatenation of
504 sequences at single-copy orthologs identified in *M. oryzae*, excluding sites with missing
505 data, sites with gaps, singletons, and monomorphic sites. The null hypothesis of no
506 recombination was tested using the PHI test implemented in SPLITSTREE.
507
508 **ABBA/BABA tests**
509 ABBA/BABA tests were performed using custom python scripts. The D statistic measuring
510 the normalized difference in counts of ABBA and BABA sites was computed using equation
511 (2) in ref. [63]. Significance was calculated using block jackknife approach (100 replicates,
512 1k SNPs blocks), to account for non-independence among sites.
513
514 **Phylogenetic species recognition**
515 We used an implementation of the GCPSR scalable to genomic data (https://github.com/b-
516 brankovics/GCPSR). The method works in two steps: (1) Concordance and non-
517 discordance analysis produces a genealogy that has clades that are both concordant and
518 non-discordant across single gene genealogies, with support value for each of the clades
519 being the number of single gene genealogies harboring the given clade at bootstrap support
520 above 95%; (2) Exhaustive subdivision places all the strains into the least inclusive clades,
521 by removing clades that would specify a species within potential phylogenetic species. We
522 kept only two outgroup sequences per gene (BR29, *M. grisea*; Pm1, *M. pennisetigena*) to
523 make sure to have the same isolate at the root of all genealogies (Pm1 isolate). Majority-
524 rule consensus trees were produced from 100 outgrouped RAxML bootstrap replicates for
525 all 2241 genes. The concordance and non-discordance analysis was carried out assuming

526 95 as the minimum bootstrap support value, and a discordance threshold of 1. Exhaustive
527 subdivision was carried out using a concordance threshold of 1121.
528
529 **Whole genome alignment & tree building**
530 A custom perl script was used to mask sequences that occur in multiple alignments when
531 the genome is BLASTed against itself. The masked genomes were then aligned in a pairwise
532 fashion against all other genomes using BLAST [61]. Regions that did not uniquely align in
533 each pair at a threshold of $1e^{-200}$, were excluded. SNPs were then identified for each
534 pairwise comparison and scaled by the total number of nucleotides aligned after excluding
535 repetitive and duplicate regions. This produced a distance metric of SNPs/Mb of uniquely
536 aligned DNA. The pairwise distances were used to construct phylogenetic trees with the
537 neighbor-joining method as implemented in the R package, Analyses of Phylogenetics and
538 Evolution (APE) [64].

539 Because alignments are in pairwise sets as opposed to a single orthologous set, assessment
540 of confidence values by traditional bootstrapping by resampling with replacement is not
541 possible. Instead, confidence values were assigned by creating 1,000 bootstrap trees with
542 noise added from a normal distribution with a mean of zero, and the standard deviation
543 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, the Southgreen and Migale computing facilities, A. Akhunova at the Kansas State University Integrated Genomics Facility, and L. Chen, 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

Strain (ID if different)	Host	Year	Place of Isolation	Sequence source
BdBar	*Triticum aestivum*	2016	Barisal, Bangladesh	this study
BdJes	*Triticum aestivum*	2016	Jessore, Bangladesh	this study
BdMeh	*Triticum aestivum*	2016	Mehepur, Bangladesh	this study
B2	*Triticum aestivum*	2011	Bolivia	this study
B71	*Triticum aestivum*	2011	Bolivia	this study
Br7	*Triticum aestivum*	1990	Parana, Brazil	this study
BR32 (BR0032)	*Triticum aestivum*	1991	Brazil	[13]
Br48	*Triticum aestivum*	1990	Mato Grosso do Sul (Brazil)	[14]
Br80	*Triticum aestivum*	1991	Brazil	this study
Br130	*Triticum aestivum*	1990	Mato Grosso do Sul, Brazil	this study
P3	*Triticum durum*	2012	Canindeyu, Paraguay	this study
PY0925	*Triticum aestivum*	2009	Predizes, Brazil	[15]

Strain (ID if different)	Host	Year	Place of Isolation	Sequence source
PY36-1	*Triticum aestivum*	2007	Brasilia, Brazil	[15]
PY5003	*Triticum aestivum*	2005	Londrina, Brazil	[15]
PY5010	*Triticum aestivum*	2005	Londrina, Brazil	[15]
PY5033	*Triticum aestivum*	2005	Londrina, Brazil	[15]
PY6017	*Triticum aestivum*	2006	Coromandel, Brazil	[15]
PY6045	*Triticum aestivum*	2006	Goiania, Brazil	[15]
PY86-1	*Triticum aestivum*	2008	Cascavel, Brazil	[15]
T25	*Triticum aestivum*	1988	Parana, Brazil	this study
WHTQ	*Triticum aestivum)*	?	Brazil	this study
WBKY11	*Triticum aestivum*	2011	Lexington, KY	[22]
P28	*Bromus tectorum*	2012	Paraguay	this study
P29	*Bromus tectorum*	2012	Paraguay	this study
CHRF	*Lolium perenne*	1996	Silver Spring, MD	this study
CHW	*Lolium perenne*	1996	Annapolis, MD	this study
FH	*Lolium perenne*	1997	Hagerstown, MD	[22]
GG11	*Lolium perenne*	1997	Lexington, KY	this study
HO	*Lolium perenne*	1996	Richmond, PA	this study
LpKY97	*Lolium perenne*	1997	Lexington, KY	[22]
PgKY4OV2.1 (PgKY)	*Lolium perenne*	2000	Lexington, KY	[15]
PgPA18C-02 (PGPA)	*Lolium perenne*	1998	Pennsylvania, USA	[15]
PL2-1	*Lolium multiflorum*	2002	Pulaski Co., KY	this study
PL3-1	*Lolium multiflorum*	2002	Pulaski Co., KY	[22]
Pg1213-22	*Festuca arundinaceum*	1999/ 2000	GA	this study
TF05-1	*Festuca arundinaceum*	2005	Lexington, KY, USA	this study

Strain (ID if different)	Host	Year	Place of Isolation	Sequence source
IB33	*Oryza sativa*	?	Texas, USA	A. Marchetti
FR13	*Oryza sativa*	1988	France	[13]
GY11	*Oryza sativa*	1988	French Guyana	[13]
IA1	*Oryza sativa*	2009	Arkansas, USA	this study
IB49	*Oryza sativa*	1992	Arkansas, USA	this study
IC17	*Oryza sativa*	1992	Arkansas, USA	this study
IE1K	*Oryza sativa*	2003	Arkansas, USA	this study
INA168	*Oryza sativa*	1958	Aichi, Japan	[14]
KEN53-33	*Oryza sativa*	1953	Aichi, Japan	[14]
ML33	*Oryza sativa*	1995	Mali	this study
P131	*Oryza sativa*	?	Japan	[52]
Y34	*Oryza sativa*	1982	China, Yunnan	[52]
P-2	*Oryza sativa*	1948	Aichi, Japan	[14]
PH14 (PH0014-rn)	*Oryza-sativa*		The Philippines	[13]
TH3	*Oryza sativa*	?	Thailand	[14]
87-120	*Oryza sativa*			
TH12 (TH0012-rn)	*Hordeum vulgare*		Thailand	[13]
TH16 (TH0016)	*Hordeum vlgare*		Thailand	[13]
Arcadia	*Setaria viridis*	1998	Lexington, KY, USA	this study
US71 (US0071)	*Setaria*		USA	[13]
GrF52	*Setaria viridis*	2001	Lexington, KY, USA	this study
KANSV1-4-1	*Setaria viridis*	1975	Kanagawa, Japan	[14]

Strain (ID if different)	Host	Year	Place of Isolation	Sequence source
SA05-43	*Setaria viridis*	2005	Nagasaki, Japan	[14]
Sv9610	*Setaria viridis*	1996	Zhejian, China	[53]
Sv9623	*Setaria viridis*	1996	Zhejian, China	[53]
GFSI1-7-2	*Setaria italica*	1977	Gifu, Japan	[14]
B51	*Eleusine indica*	2012	Quirusillas, Bolivia	this study
BR62	*Eleusine indica*	1991	Brazil	this study
CD156	*Eleusine indica*	1989	Ferkessedougou, Cote d'Ivoire	[13]
EI9411	*Eleusine indica*	1990	Fujian, China	[53]
EI9064	*Eleusine indica*	1996	Fujian, China	[53]
G22	*Eleusine coracana*	1976	Japan	this study
Z2-1	*Eleusine coracana*	1977	Kagawa, Japan	[14]
PH42	*Eleusine indica*	1983	The Philippines	this study
SSFL02	*Stenotaphrum secundatum*	2002	Disneyworld, FL, USA	this study
SSFL14-3	*Stenotaphrum secundatum*	2014	New Smyrna, FL, USA	this study
G17	*Eragrostis curvula*	1976	Japan	this study
Br58	*Avena sativa*	1990	Parana, Brazil	[14]
Bd8401	*Brachiaria distachya*	1984	The Philippines	this study
Bm88324	*Brachiaria mutica*	1988	The Philippines	this study
PM1	*Pennisetum americanum*	1990	Georgia, U.S.A	this study
BR29	*Digitaria sanguinalis*	1989	Brazil	[13]

Strain (ID if different)	Host	Year	Place of Isolation	Sequence source
Dig41	*Digitaria sanguinalis*	ND	Hyogo, Japan	[14]
DsLIZ	*Digitaria sanguinalis*	2000	Lexington, KY, USA	this study
VO107	*Digitaria sanguinalis*	1981	Texas, USA	this study

ND, no data

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.

Figure 2. Total evidence maximum likelihood tree based on the concatenation of 2,682 orthologous coding sequences extracted from 76 *M. oryzae* genome. Numbers above branches represent bootstrap supports after 100 bootstrap replicates. Only nodes with bootstrap support > 50 are labelled.

Figure 3. Total evidence neighbor-joining distance tree using pairwise distances (SNPs/kb) calculated from analysis of pairwise blast alignments between repeat-masked genomes. Only nodes with confidences > 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 "Castroagudin" loci 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 marker loci. Dark 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). Black arrows show 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. 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. Brachiaria_1: Bm88324. Brachiaria_2: Bd8401. Eleusine_1 and Eleusine_2 are, respectively, the light orange and orange clusters in DAPC analysis.

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. 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.

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* alleles among *M. oryzae* isolates as indicated by mapping onto the neighboring tree built using whole genome SNP data. *MPG1* alleles were identified by using a reference marker sequence to search all the genomes using BLAST.

Figure S6. Distribution of alleles among *M. oryzae* isolates as indicated by mapping onto the neighboring tree built using whole genome SNP data. Trees are shown for the following loci: A) BAC6, B) β-tubulin1, C) CAL, D) CH7BAC7, E) CH7BAC9, F) EF1α, G) NUT1. Sequence variants are noted above each tree using the BLAST backtrace operations (BTOP) format.

Figure S7. A) Log likelihood of data vs number of clusters assumed in STRUCTURE analysis. Error bars are standard deviations of likelihood across STRUCTURE repeats. **B) 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.

References

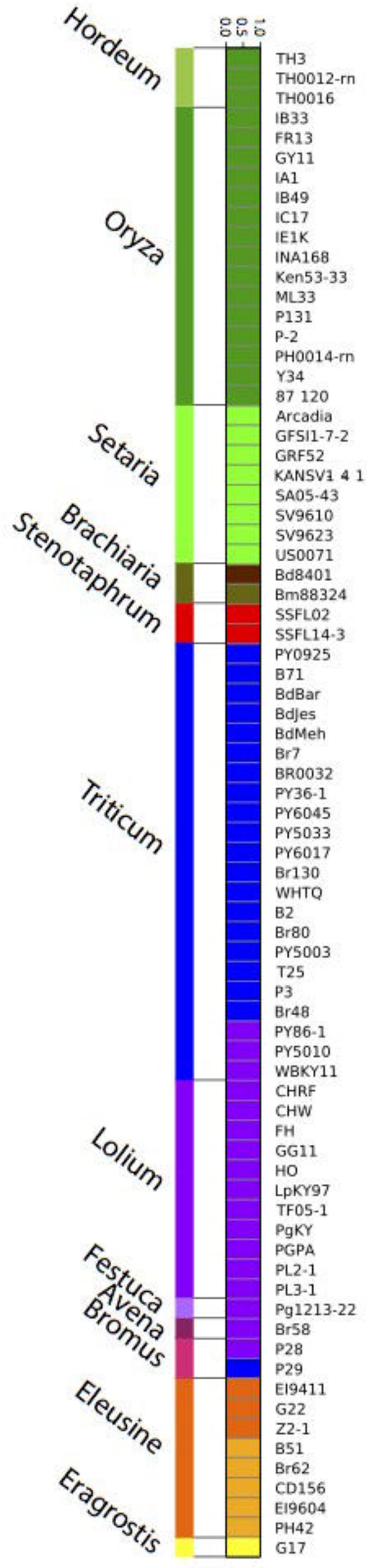
1. Fournier E, Giraud T, Albertini C, Brygoo Y. Partition of the *Botrytis cinerea* complex in France using multiple gene genealogies. Mycologia. 2005;97(6):1251-67. Epub 2006/05/26. PubMed PMID: 16722218.
2. Le Gac M, Hood ME, Fournier E, Giraud T. Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. Evolution; international journal of organic evolution. 2007;61(1):15-26. Epub 2007/02/16. doi: 10.1111/j.1558-5646.2007.00002.x. PubMed PMID: 17300424.
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*

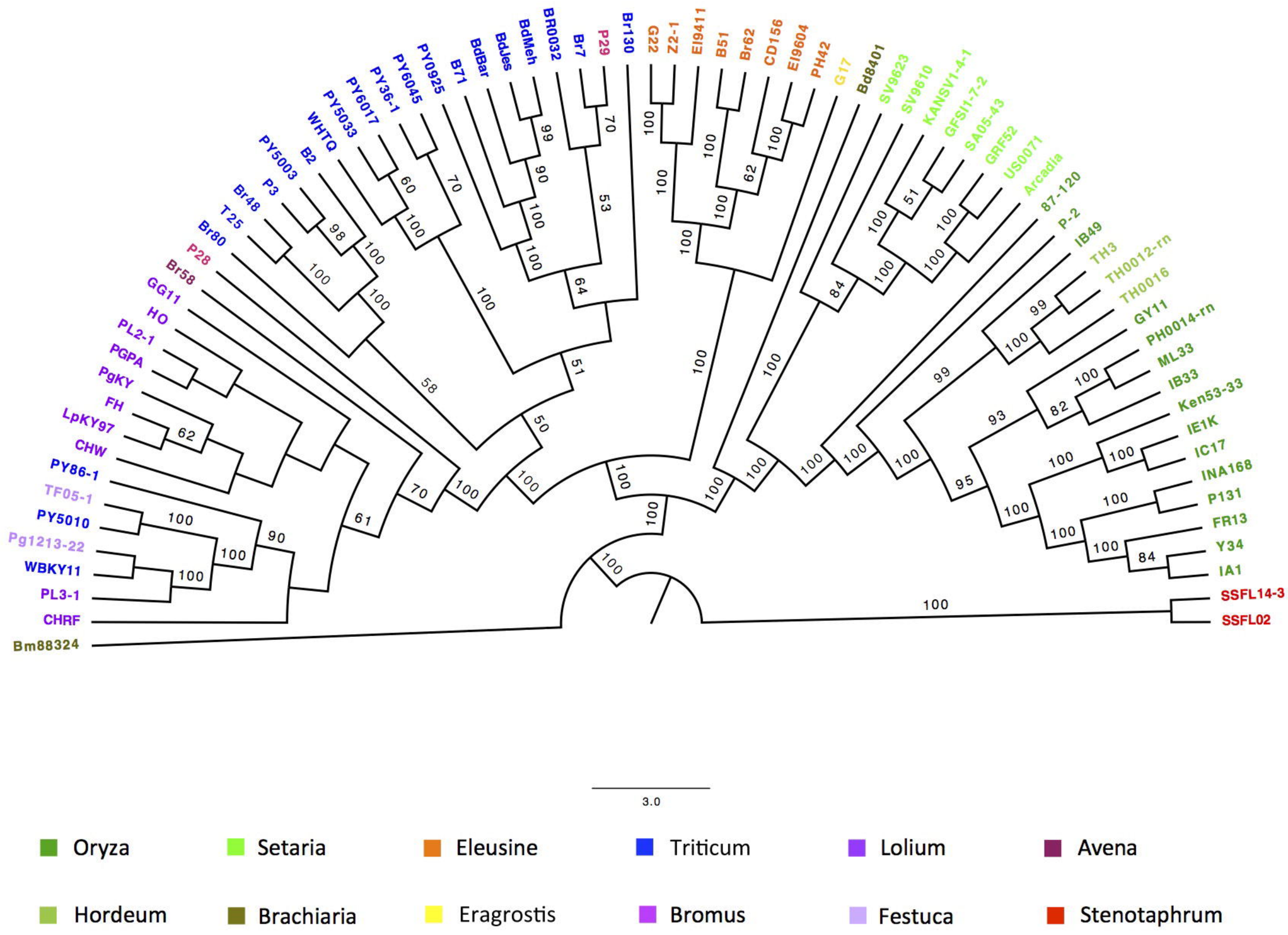
oryzae attacking different host plants. Genome Biol Evol. 2015;7(10):2896-912. doi: 10.1093/gbe/evv187. PubMed PMID: 26454013; PubMed Central PMCID: PMCPMC4684704.
14. Yoshida K, Saunders DG, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, et al. Host specialization of the blast fungus *Magnaporthe oryzae* 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. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.
40. 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.

41. Nosil P, Vines TH, Funk DJ. Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution; international journal of organic evolution. 2005;59(4):705-19. PubMed PMID: 15926683.
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.
47. Silué D, Notteghem JL. Production of perithecia of *Magnaporthe grisea* on rice plants. Mycol Res. 1990;94(8):1151-2.
48. 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.
49. 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.
50. 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.
51. 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*. Plant Cell. 2009;21(5):1573-91. doi: 10.1105/tpc.109.066324. PubMed PMID: 19454732; PubMed Central PMCID: PMCPMC2700537.
52. 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.
53. 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.
54. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research. 2008;18(5):821-9.

55. 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.
56. 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.
57. 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.
58. 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.
59. 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.
60. 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.
61. 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.
62. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics. 1998;14(1):68-73. Epub 1998/04/01. PubMed PMID: 9520503.
63. 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.
64. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004;20(2):289-90. PubMed PMID: 14734327.


0.0
0.5
1.0
Hordeum
Oryza
Setaria
Brachiaria
Stenotaphrum
Triticum
Lolium
Festuca
Avena
Bromus
Eleusine
Eragrostis
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
TF05-1
PgKY
PGPA
PL2-1
PL3-1
Pg1213-22
Br58
P28
P29
EI9411
G22
Z2-1
B51
Br62
CD156
EI9604
PH42
G17


Br130
G22
Z2-1
EI9411
B51
Br62
CD156
EI9604
PH42
G17
Bd8401
SV9623
SV9610
KANSV1-4-1
GFSI1-7-2
SA05-43
GRF52
US0071
Arcadia
87-120
P-2
IB49
TH3
TH0012-rn
TH0016
GY11
PH0014-rn
ML33
IB33
Ken53-33
IE1K
IC17
INA168
P131
FR13
Y34
IA1
SSFL14-3
SSFL02
BR0032
Br7
P29
BdMeh
BdJes
BdBar
B71
PY0925
PY6045
PY36-1
PY6017
PY5033
WHTQ
B2
PY5003
P3
Br48
T25
Br80
P28
Br58
GG11
HO
PL2-1
PGPA
PgKY
FH
LpKY97
CHW
PY86-1
TF05-1
PY5010
Pg1213-22
WBKY11
PL3-1
CHRF
Bm88324
3.0
Oryza
Setaria
Eleusine
Triticum
Lolium
Avena
Hordeum
Brachiaria
Eragrostis
Bromus
Festuca
Stenotaphrum

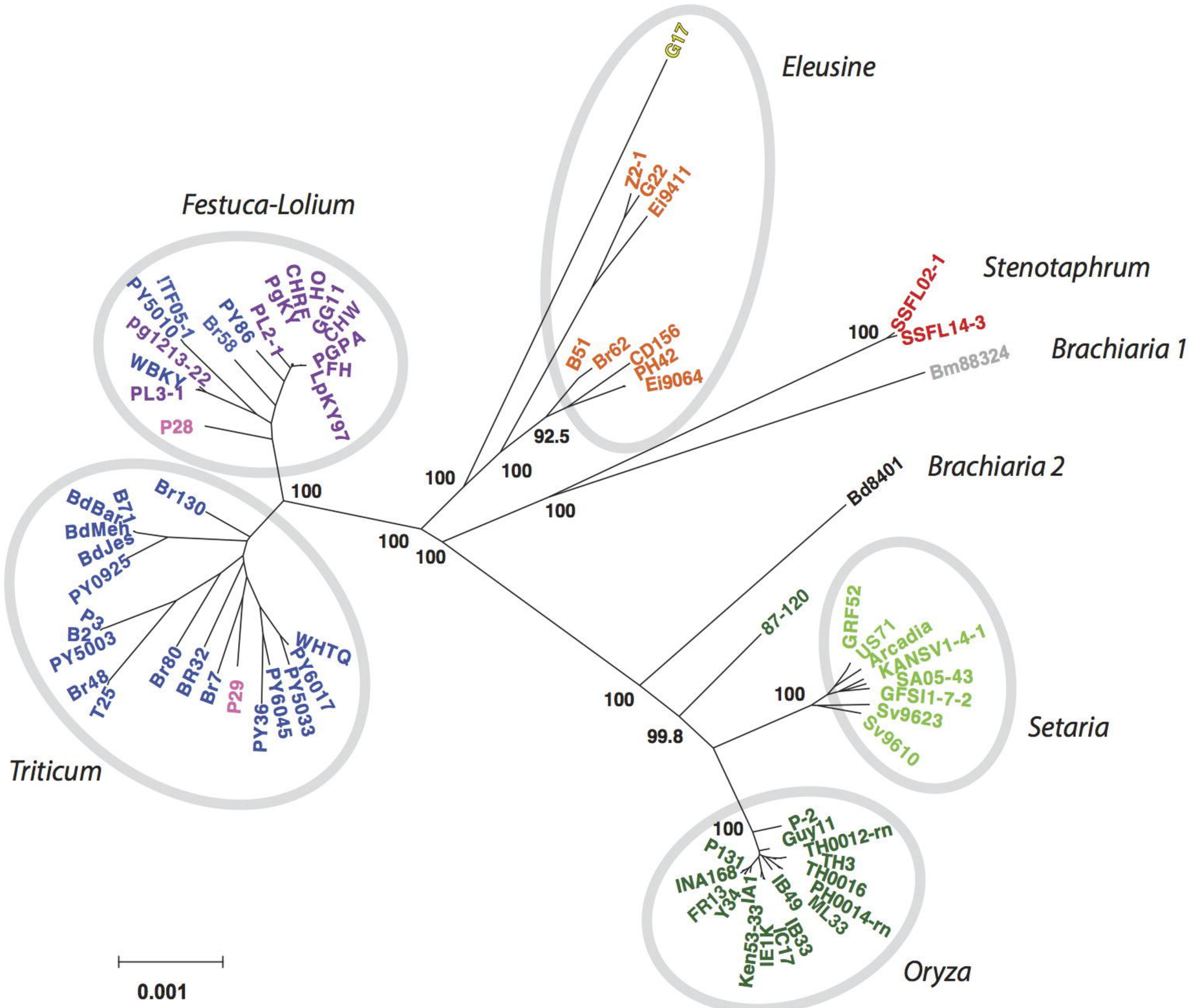
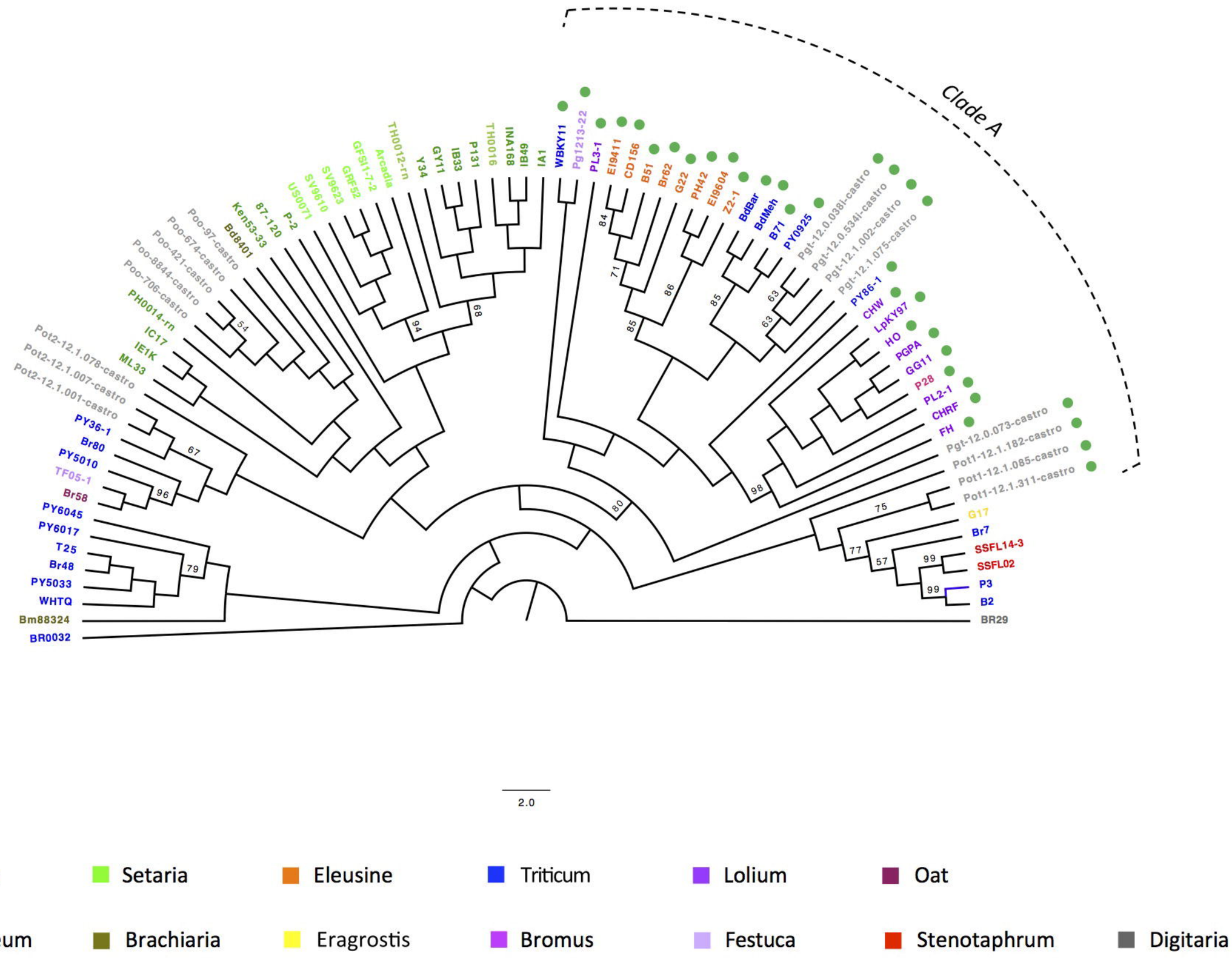
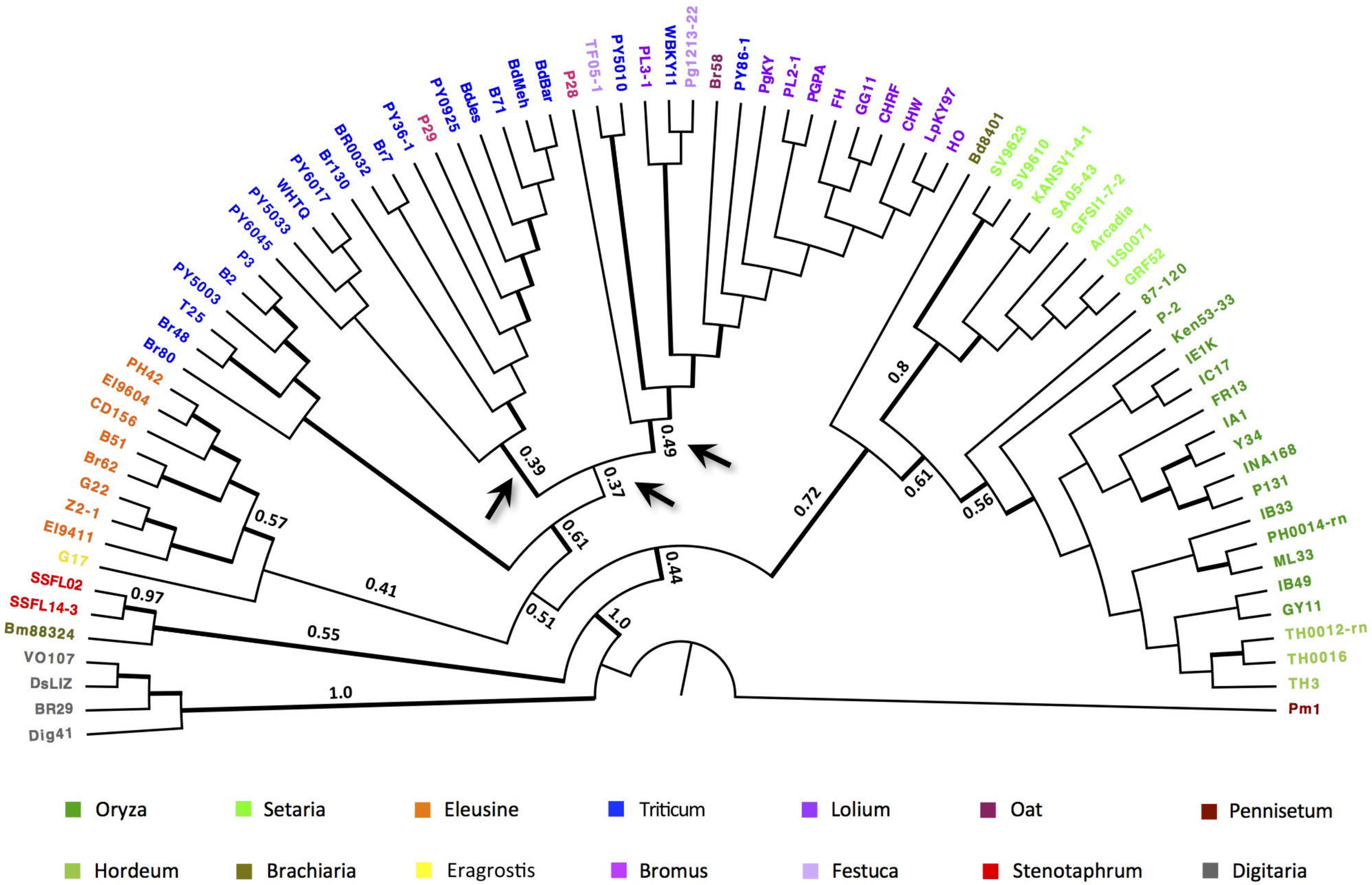
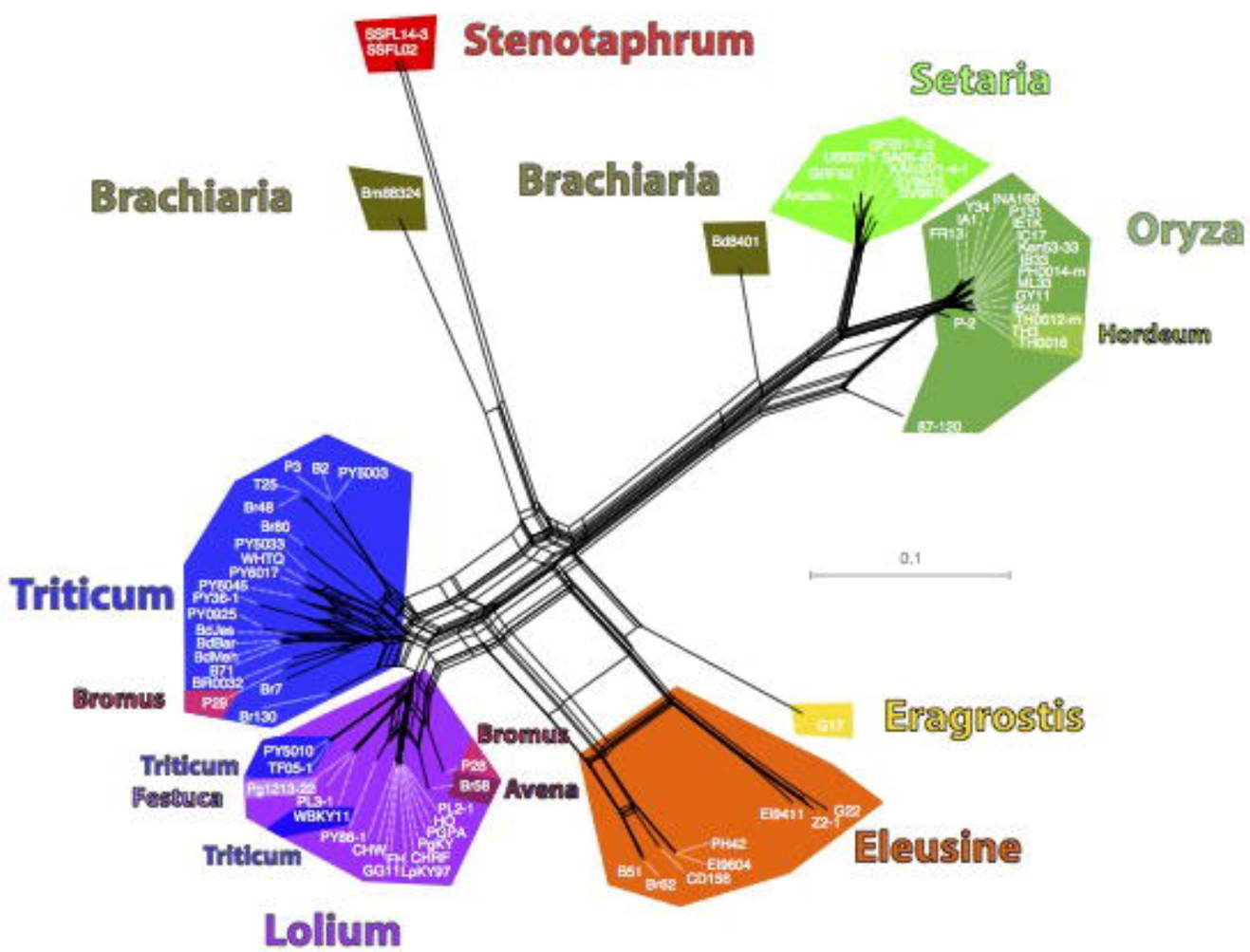

Eleusine
G17
Z2-1
G22
Ei9411
Festuca-Lolium
ITF05-1
PY5010
pg1213-22
WBKY
PL3-1
P28
Br58
PY86
PL2-1
PgKY
CHRE
HO
GG11
CHW
PGPA
FH
LpKY97
Stenotaphrum
SSFL02-1
SSFL14-3
Brachiaria 1
Bm88324
B51
Br62
CD156
PH42
Ei9064
92.5
100
100
100
100
100
100
Brachiaria 2
Bd8401
87-120
100
99.8
100
Setaria
GRF52
US71
Arcadia
KANSV1-4-1
SA05-43
GFSI1-7-2
Sv9623
Sv9610
Triticum
BdBar1
B71
Br130
BdMeh
BdJes
PY0925
P3
B2
PY5003
Br48
T25
Br80
BR32
Br7
P29
PY36
PY6045
PY5033
PY6017
WHTQ
Oryza
P-2
Guy11
TH0012-rn
TH3
TH0016
PH0014-rn
ML33
IB49
IB33
IC17
IE1K
Ken53-33
IA1
Y34
FR13
INA168
P131
0.001

Figure 3



Oryza
Setaria
Eleusine
Triticum
Lolium
Oat
Hordeum
Brachiaria
Eragrostis
Bromus
Festuca
Stenotaphrum
Digitaria


Dig41
BR29
DsLIZ
VO107
Bm88324
SSFL14-3
SSFL02
G17
EI9411
Z2-1
G22
Br62
B51
CD156
EI9604
PH42
Br80
Br48
T25
PY5003
B2
P3
PY6045
PY5033
WHTQ
PY6017
Br130
BR0032
Br7
PY36-1
P29
PY0925
BdJes
B71
BdMeh
BdBar
P28
TF05-1
PY5010
PL3-1
WBKY11
Pg1213-22
Br58
PY86-1
PgKY
PL2-1
PGPA
FH
GG11
CHRF
CHW
LpKY97
HO
Bd8401
SV9623
SV9610
KANSV1-4-1
SA05-43
GFSI1-7-2
Arcadia
US0071
GRF52
87-120
P-2
Ken53-33
IE1K
IC17
FR13
IA1
Y34
INA168
P131
IB33
PH0014-rn
ML33
IB49
GY11
TH0012-rn
TH0016
TH3
Pm1
1.0
0.55
0.97
0.41
0.57
0.51
0.61
0.37
0.39
0.49
0.44
1.0
0.72
0.61
0.56
0.8
Oryza
Setaria
Eleusine
Triticum
Lolium
Oat
Pennisetum
Hordeum
Brachiaria
Eragrostis
Bromus
Festuca
Stenotaphrum
Digitaria


Stenotaphrum
Setaria
Brachiaria
Brachiaria
Oryza
Hordeum
Triticum
Bromus
Triticum
Festuca
Triticum
Lolium
Bromus
Avena
Eragrostis
Eleusine
0.1