Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus Magnaporthe oryzae

Pierre Gladieux, Bradford Condon, Sébastien Ravel, Darren Soanes, Joao Leodato Nunes Maciel, Antonio Nhani, Li Chen, Ryohei Terauchi, Marc-Henri Lebrun, Didier Tharreau, et al.

RESEARCH ARTICLE

Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus *Magnaporthe oryzae*

Pierre Gladieux,[a] Bradford Condon,[b] Sebastien Ravel,[a] Darren Soanes,[c] Joao Leodato Nunes Maciel,[d] Antonio Nhani, Jr,[e] Li Chen,[b] Ryohei Terauchi,[f] Marc-Henri Lebrun,[g] Didier Tharreau,[a] Thomas Mitchell,[h] Kerry F. Pedley,[i] Barbara Valent,[j] Nicholas J. Talbot,[c] Mark Farman,[b] Elisabeth Fournier[a]

[a]UMR BGPI, Univ Montpellier, INRA, CIRAD, Montpellier SupAgro, Montpellier, France
[b]Department of Plant Pathology, University of Kentucky, Lexington, Kentucky, USA
[c]College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom
[d]Embrapa Wheat, Passo Fundo, Brazil
[e]Embrapa Agricultural Informatics, Campinas, Brazil
[f]Iwate Biotechnology Research Center, Kitakami, Iwate, Japan
[g]INRA, AgroParisTech UMR BIOGER, Thiverval-Grignon, France
[h]Department of Plant Pathology, Ohio State University, Columbus, Ohio, USA
[i]USDA, Agricultural Research Service, FDWSRU, Ft. Detrick, Maryland, USA
[j]Department of Plant Pathology, Kansas State University, Manhattan, Kansas, USA

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 multihost pathogen that infects multiple grasses and cereals, is responsible for the most damaging rice disease (rice blast), and is 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 ecoevolutionary factors that underlie the diversification of *M. oryzae* and highlight the practicality of genomic data for epidemiological surveillance in this important multihost pathogen.

IMPORTANCE Infection of novel hosts is a major route for disease emergence by pathogenic microorganisms. Understanding the evolutionary history of multihost pathogens is therefore important to better predict the likely spread and emergence of new diseases. *Magnaporthe oryzae* is a multihost 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 interlineage gene flow has contrib-

Received 18 July 2017 **Accepted** 20 November 2017 **Published** 27 February 2018

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

Editor John W. Taylor, University of California, Berkeley

Address correspondence to Elisabeth Fournier, elisabeth.fournier@inra.fr.

P.G. and B.C. contributed equally to the work.

This is contribution no. 18-005-J from the Kansas Agricultural Experiment Station and contribution number 17-12-107 from the Kentucky Agricultural Experiment Station.

uted 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.

KEYWORDS cryptic species, disease emergence, diversification, fungal pathogen, gene flow, population structure, rice, speciation, species recognition

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, and proximate (i.e., molecular) and ultimate (i.e., ecoevolutionary) 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 benefits of whole-genome approaches include 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 multihost 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*) (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* and *Setaria viridis*), as well as wild and cultivated grass hosts, including goosegrass (*Eleusine indica*), annual ryegrass (*Lolium multiflorum*), perennial ryegrass (*Lolium 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 only wheat 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 consider-

able 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 United States, 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*, which 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 (GCPSR [25, 26]) was not applied. 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 reassess the hypothesis that *P. graminis-tritici* 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* isolates from 12 host plant genera, four *Magnaporthe grisea* isolates from crabgrass [*Digitaria* spp.], and one *Magnaporthe pennisetigena* isolate 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

Reassessing the validity of the proposed *P. graminis-tritici* species by analyzing the original published data according to GCPSR. To test the previous delineation of a subgroup of wheat-infecting isolates as a new phylogenetic species, we reanalyzed the Castroagudin et al. data set (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: (i) genealogical concordance (the group is present in the majority of the single-locus genealogies) and (ii) genealogical nondiscordance (the group is well supported in at least one single-locus genealogy and is not contradicted in any other genealogy at the same level of support) (25). Visual inspection of the topologies and supports in each single-locus tree revealed that GCPSR condition 1 was not satisfied since isolates previously identified as belonging to the phylogenetic species *P. graminis-tritici* grouped together in only one maximum likelihood gene genealogy—the one produced using the *MPG1* locus (see Fig. S1A in the supplemental material). The *P. graminis-tritici* separation was not supported by any of the nine other single-locus genealogies (Fig. S1B to J).

Next, we used the multilocus data as input to the program ASTRAL with the goal of inferring a species tree that takes into account possible discrepancies among individual gene genealogies (27–29). The ASTRAL tree failed to provide strong support for the branch holding the isolates previously identified as *P. graminis-tritici* (Fig. S2). Thus, analysis of the Castroagudin et al. data according to GCPSR standards failed to support the existence of the newly described *P. graminis-tritici* species.

Inferring population subdivision within *M. oryzae* using whole-genome data. We sought to test whether a phylogenomic study could provide better insight into the possibility of speciation within *M. oryzae*. To this end, whole-genome sequence data were acquired for a comprehensive collection of 76 *M. oryzae* isolates from 12 host genera, four *M. grisea* isolates from *Digitaria* spp., and one *M. pennisetigena* isolate from *Pennisetum* (Table 1). The analysis included sequence data for strains collected on rice (*Oryza sativa*), finger millet and goosegrass (*Eleusine* spp.), wheat (*Triticum* spp.), tall fescue (*Festuca arundinaceum*), annual and perennial ryegrasses (*Lolium multiflorum* and *L. perenne*, respectively), and barley (*Hordeum vulgare*). Representatives of previously unstudied host-specialized populations from foxtails (*Setaria* sp.), St. Augustine grass (*Stenotaphrum secundatum*), weeping lovegrass (*Eragrostis curvula*), signalgrass (*Brachiaria* sp.), cheatgrass (*Bromus tectorum*), and oat (*Avena sativa*) were also included. Single nucleotide polymorphisms (SNPs) identified in aligned sequences of 2,682 orthologous single-copy genes were identified in all *M. oryzae* genomes (in total ~6.6 Mb of sequence data) and from whole-genome SNPs identified from pairwise BLAST alignments of repeat-masked genomes (average ~36 Mb aligned sequence).

First, we employed the multivariate approach implemented in discriminant analysis of principal components (DAPC [30]) to examine population subdivision within *M. oryzae*. Using the haplotypes identified from orthologous loci, the Bayesian information criterion plateaued at $K = 10$ in models varying in K from 2 to 20 clusters, indicating that $K = 10$ captures the most salient features of population subdivision (Fig. S3). Clusters identified at $K = 10$ were as follows: (i) isolates from rice and two isolates from barley (dark green in Fig. 1; referred to as the *Oryza* lineage); (ii) isolates from *Setaria* sp. (light green; referred to as the *Setaria* lineage); (iii) isolate Bm88324 from *Brachiaria mutica* (olive; referred to as the *Brachiaria*1 lineage); (iv) isolate Bd8401 from *Brachiaria distachya* (brown; referred to as the *Brachiaria*2 lineage); (v) isolates from *Stenotaphrum* (red; referred to as the *Stenotaphrum* lineage); (vi) 17 of the 22 isolates from wheat and an isolate from *Bromus* (blue; referred to as the *Triticum* lineage); (vii) the remaining 3/22 isolates from wheat together with isolates from *Lolium*, *Festuca*, and oat and a second isolate from *Bromus* (purple; referred to as the *Lolium* lineage); (viii and ix) isolates from *Eleusine* that formed two distinct clusters (light orange and orange; referred to as the *Eleusine*1 and *Eleusine*2 lineages, respectively); and (x) an isolate from *Eragrostis* (yellow; referred to as the *Eragrostis* lineage) (Fig. 1). 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 (ML) and distance-based methods. Both approaches produced trees that corresponded well with the subdivisions identified in DAPC. The tree generated using ML analysis of orthologous genes displayed a topology with 10 lineages (Fig. 2) showing one-to-one correspondence with the K clusters from DAPC (Fig. 1 and 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 (NJ) tree built using "total-genome" pairwise distances resolved very similar groupings as the DAPC (Fig. 1 and S3) and the ML ortholog tree (Fig. 3). The only major discrepancy between ML and NJ trees was the confident placement of 87-120—an isolate from rice—outside the rice clade in the NJ tree (Fig. 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 1 order of magnitude higher in the rice and wheat lineages (0.1% difference per site) than in the *Lolium* and *Setaria* lineages (other lineages were not included in the calculations

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

Isolate ID[c]	Synonym(s)	Host	Yr	Locality	NCBI accession no.	Sequence source	Reference(s)[b]
BdBar	BdBar16-1	*Triticum aestivum*	2016	Barisal, Bangladesh	SAMN04940126	23	
BdJes	BdJes16-1	*Triticum aestivum*	2016	Jessore, Bangladesh	SAMN04942531	23	
BdMeh	BdMeh16-1	*Triticum aestivum*	2016	Mehepur, Bangladesh	SAMN04942534	23	
B2		*Triticum aestivum*	2011	Bolivia	SAMN05580113	67	
B71		*Triticum aestivum*	2012	Bolivia	SAMN04942725	23	67
Br7		*Triticum aestivum*	1990	Parana, Brazil	SAMN08009545	22	67
BR0032	BR32	*Triticum aestivum*	1991	Brazil		13	15
Br48		*Triticum aestivum*	1990	Mato Grosso do Sul, Brazil		14	22
Br80		*Triticum aestivum*	1991	Brazil	SAMN08009546	22	67
Br130		*Triticum aestivum*	1990	Mato Grosso do Sul, Brazil	SAMN08009547	22	
P3		*Triticum durum*	2012	Canindeyu, Paraguay	SAMN08009568	67	
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	67
PY5033	PY05033	*Triticum aestivum*	2005	Londrina, Brazil		15	
PY6017	PY06017	*Triticum aestivum*	2006	Coromandel, Brazil		15	
PY6045	PY06045	*Triticum aestivum*	2006	Goiânia, Brazil		15	
PY86-1	PY86.1	*Triticum aestivum*	2008	Cascavel, Brazil		15	
T25		*Triticum aestivum*	1988	Parana, Brazil	SAMN08009575	22	67
WHTQ		*Triticum aestivum*	ND[a]	Brazil	SAMN08009580	22	
WBKY11	WBKY11-15	*Triticum aestivum*	2011	Lexington, KY	SAMN08009578	22	67
P28	P-0028	*Bromus tectorum*	2014	Paraguay	SAMN05864041	67	
P29	P-0029	*Bromus tectorum*	2014	Paraguay	SAMN05898532	67	
CHRF		*Lolium perenne*	1996	Silver Spring, MD	SAMN08009548	67	
CHW		*Lolium perenne*	1996	Annapolis, MD	SAMN08009549	67	
FH		*Lolium perenne*	1997	Hagerstown, MD	SAMN08009551	22	67
GG11		*Lolium perenne*	1997	Lexington, KY	SAMN08009555	22	
HO		*Lolium perenne*	1996	Richmond, PA	SAMN08009558	22	
LpKY97	LpKY97-1	*Lolium perenne*	1997	Lexington, KY	SAMN08009564	22	67
PgKY	PgKY4OV2.1	*Lolium perenne*	2000	Lexington, KY		15	
PGPA	PgPA18C-02, PgPA	*Lolium perenne*	1998	Pennsylvania, USA		15	
PL2-1		*Lolium multiflorum*	2002	Pulaski Co., KY	SAMN08009571	22	
PL3-1		*Lolium multiflorum*	2002	Pulaski Co., KY	SAMN08009572	22	67
Pg1213-22		*Festuca arundinaceum*	1999/2000	Georgia, USA	SAMN08009569	67	
TF05-1		*Festuca arundinaceum*	2005	Lexington, KY	SAMN08009576	This study	
IB33		*Oryza sativa*	ND	Texas, USA	SAMN08009560	This study	
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	SAMN08009559	67	
IB49	ZN61	*Oryza sativa*	1992	Arkansas, USA	SAMN08009561	67	
IC17	ZN57	*Oryza sativa*	1992	Arkansas, USA	SAMN08009562	67	
IE1K	TM2	*Oryza sativa*	2003	Arkansas, USA	SAMN08009563	67	
INA168	Ina168	*Oryza sativa*	1958	Aichi, Japan		14	68
KEN53-33	Ken53-33	*Oryza sativa*	1953	Aichi, Japan		14	
ML33		*Oryza sativa*	1995	Mali	SAMN08009565	This study	
P131		*Oryza sativa*	ND	Japan		54	22, 55
Y34		*Oryza sativa*	1982	Yunnan, China		54	22, 55
P-2	P2	*Oryza sativa*	1948	Aichi, Japan		14	
PH0014-rn	PH0014, PH14	*Oryza sativa*	ND	Philippines		13	15
TH3		*Oryza sativa*	ND	Thailand		14	14
87-120		*Oryza sativa*	ND		SAMN08377452	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	SAMN08009540	22	67
US0071	US71	*Setaria* spp.	ND	USA		13	15
GrF52		*Setaria viridis*	2001	Lexington, KY	SAMN08009556	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		55	
Sv9623		*Setaria viridis*	1996	Zhejian, China		55	
GFSI1-7-2	GFSI	*Setaria italica*	1977	Gifu, Japan		14	
B51		*Eleusine indica*	2012	Quirusillas, Bolivia	SAMN08009542	22	67
BR62		*Eleusine indica*	1991	Brazil		15	
CD156	CD0156	*Eleusine indica*	1989	Ferkessedougou, Ivory Coast		13	15

(Continued on next page)

TABLE 1 (Continued)

Isolate ID[c]	Synonym(s)	Host	Yr	Locality	NCBI accession no.	Sequence source	Reference(s)[b]
EI9411		*Eleusine indica*	1990	Fujian, China		55	
EI9064		*Eleusine indica*	1996	Fujian, China		55	
G22	WGG-FA40	*Eleusine coracana*	1976	Japan	SAMN08009554	This study	
Z2-1		*Eleusine coracana*	1977	Kagawa, Japan		14	68
PH42		*Eleusine coracana*	1983	Philippines	SAMN08009570	67	
SSFL02		*Stenotaphrum secundatum*	2002	Disney World, FL	SAMN08009573	67	
SSFL14-3		*Stenotaphrum secundatum*	2014	New Smyrna, FL	SAMN08009574	This study	
G17	K76-79	*Eragrostis curvula*	1976	Japan	SAMN08009553	15	
Br58		*Avena sativa*	1990	Parana, Brazil		14	68
Bd8401		*Brachiaria distachya*	1984	Philippines	SAMN08009543	This study	
Bm88324		*Brachiaria mutica*	1988	Philippines	SAMN08009544	This study	
PM1		*Pennisetum americanum*	1990	Georgia, USA	SAMN08377453	This study	
BR29	BR0029	*Digitaria sanguinalis*	1989	Brazil		13	
Dig41		*Digitaria sanguinalis*	ND	Hyogo, Japan		14	68
DsLIZ		*Digitaria sanguinalis*	2000	Lexington, KY	SAMN08009550	67	
VO107		*Digitaria sanguinalis*	1981	Texas, USA	SAMN08009577	This study	

[a]ND, no data.

[b]"Reference(s)" lists studies that used the sequencing data, besides the present study.

[c]Isolates Br116.5, Br118.2, TP2, MZ5-1-6, and Br35, sequenced by Inoue et al. (68); Bangladeshi isolates and isolates PY05002, PY06025, PY06047, PY25.1, PY35.3, and PY05035, sequenced by Islam et al. (15); isolate SA05-144, sequenced by Yoshida et al. (14); isolates PY5020 and PY22.1 from the work of Pieck et al. (67); and isolates DS9461 and DS0505, sequenced by Zhong et al. (55), were not included in the study.

due to small sample sizes—only lineages with $n > 6$ were 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 1 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 (Fig. S4; Table S1), compared with *M. oryzae* versus *M. grisea*, *M. oryzae* versus *M. pennisetigena*, or *M. grisea* versus *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,

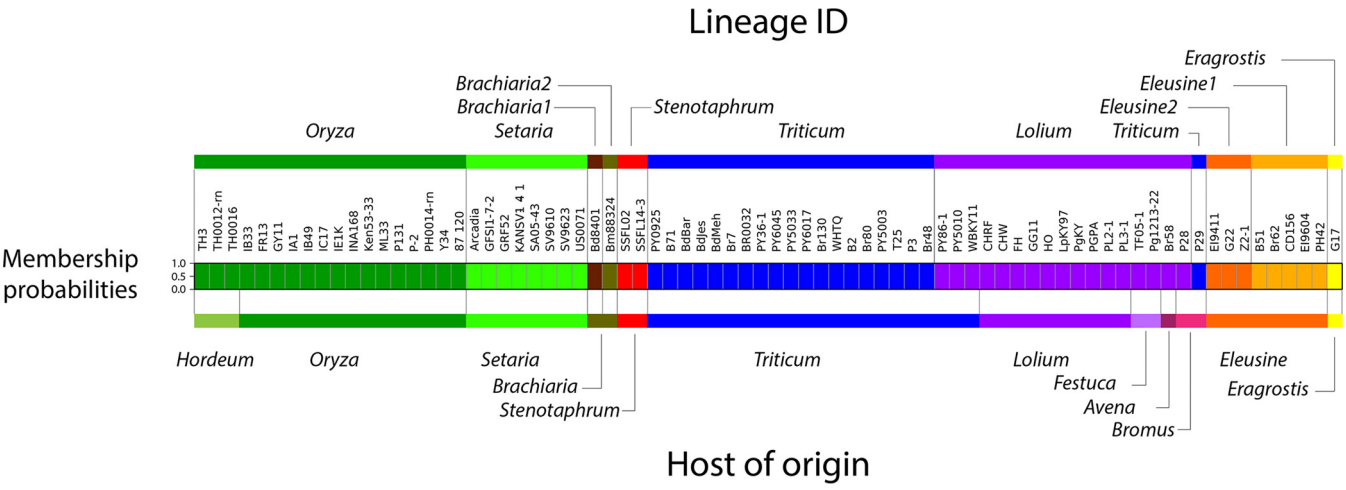


FIG 1 Discriminant analysis of principal components, assuming *K* of 10 clusters. Each isolate is represented by a thick vertical line divided into *K* segments that represent the isolate's estimated membership probabilities in the $K = 10$ clusters (note that all isolates have high membership probabilities in a single cluster, and hence, only a single segment is visible). The host of origin of samples is shown below the bar plot, and lineage IDs are shown above the bar plot.

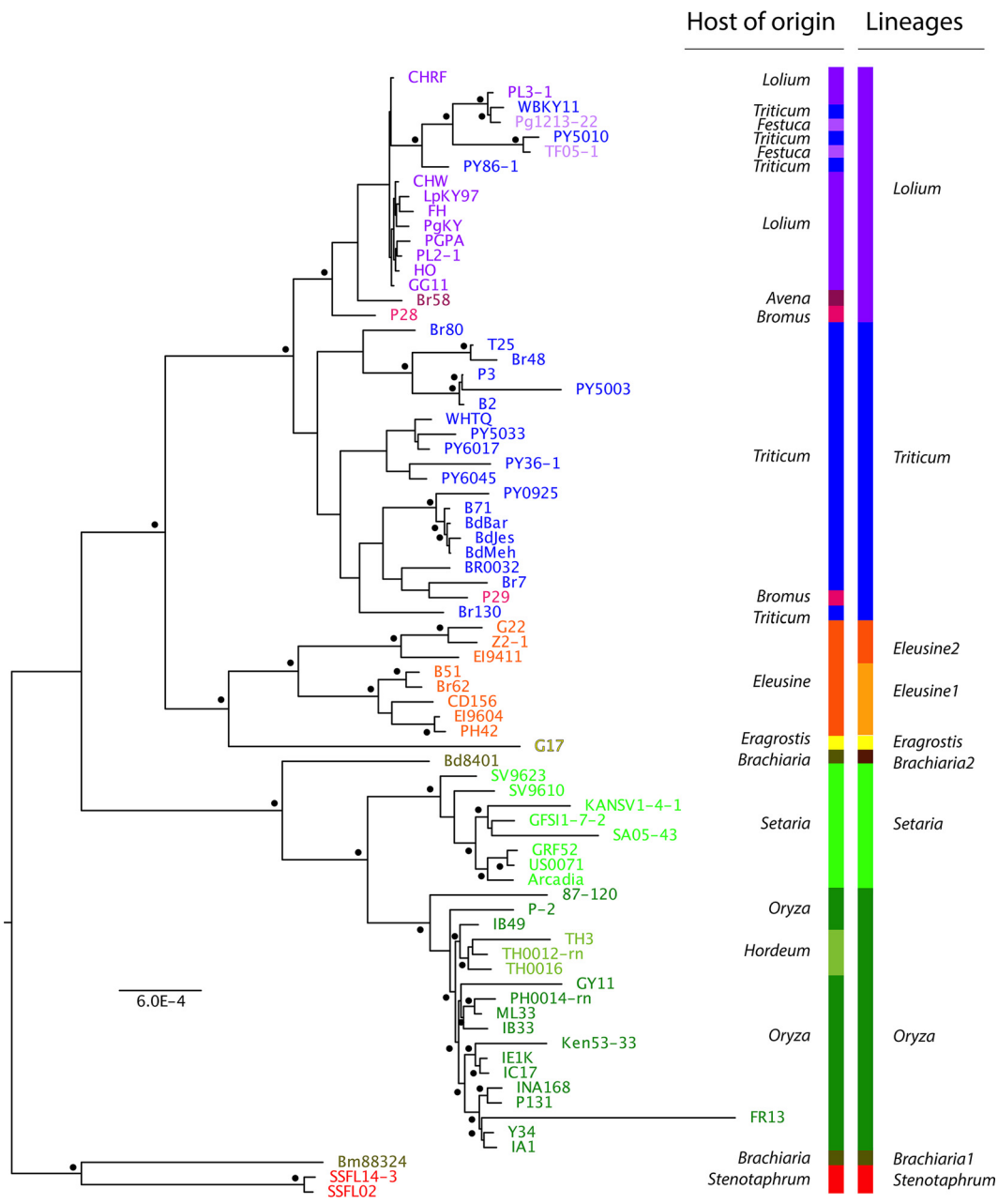
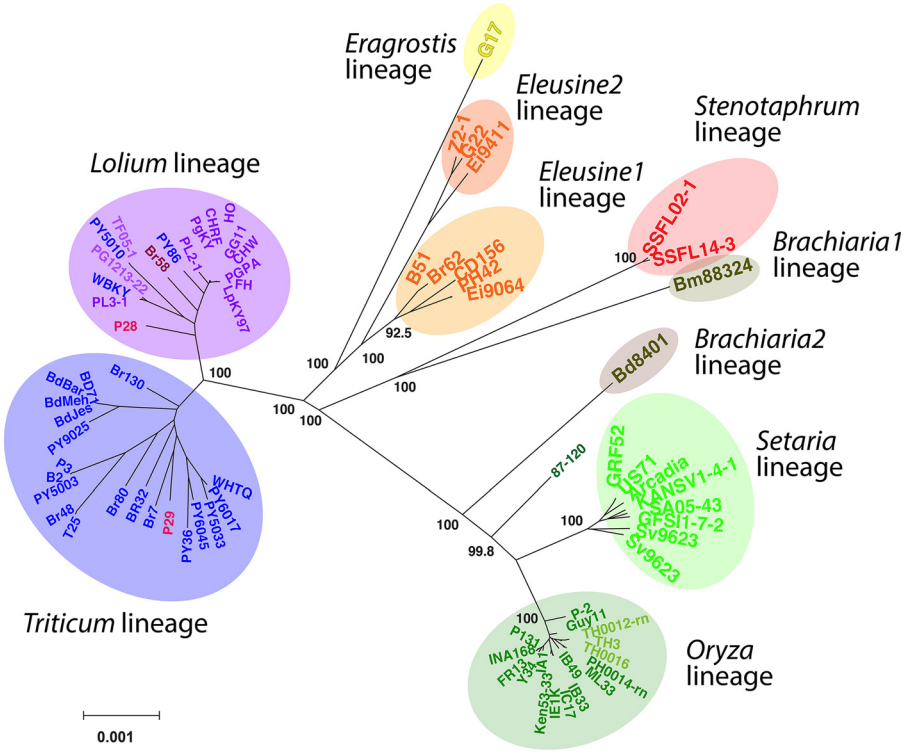


FIG 2 Maximum likelihood tree based on the concatenation of 2,682 orthologous coding sequences extracted from 76 *M. oryzae* genome. Nodes with bootstrap support of >90% are indicated by dots (100 bootstrap replicates).

provides good evidence against the existence of relatively ancient cryptic species within *M. oryzae* (Table S1).

Reassessment of *P. graminis-tritici* as a novel species using whole-genome data. While the 10 loci utilized in the Castroagudin et al. (24) study do not support the *P. graminis-tritici* 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 and the other comprising largely *Festuca*- and *Lolium*-infecting isolates as well as a few wheat-infecting isolates (Fig. 2 and 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 *P. graminis-tritici* species and the *Lolium* lineage (or indeed the *Triticum* lineage), we



Hosts of origin (tip colors)

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

FIG 3 Total-evidence neighbor-joining distance tree using pairwise distances (number of differences per kilobase) calculated from analysis of pairwise BLAST alignments between repeat-masked genomes. Only nodes with confidence of >80% (see Materials and Methods) are labeled. Ovals are drawn around the main lineages for clarity.

extended the 10-locus analysis to the *M. oryzae* genome sequences used in the present study. For reference, we included the multilocus data for 16 isolates from the Castroagudin et al. study (24), representing all the major clades from that study. Nine of the 10 loci were successfully recovered from 68 of our *M. oryzae* genome sequences. The remaining locus, CH7-BAC9, was absent from too many genome sequences and, as a result, was excluded from the analysis.

The nine concatenated loci produced a total-evidence RAxML tree in which very few branches had bootstrap support greater than 50% (Fig. 4). All of the *P. graminis-tritici*

TABLE 2 Summary of population genetic variation at 2,682 single-copy orthologous genes in wheat, *Lolium*, rice, and *Setaria* lineages of *Magnaporthe oryzae*[a]

Lineage	*n*	*S*	*K*	H_e	θ_w	π	PHI test (*P* value)
Wheat	20	5.8	1.9	0.17	1.28E−3	1.24E−3	0
Lolium	17	3.1	1.5	0.10	7.02E−4	6.54E−4	0
Rice	18	5.3	2.3	0.12	1.55E−3	7.75E−4	0
Setaria	8	2.6	1.8	0.18	9.10E−4	7.68E−4	0

[a]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 base pair; π is nucleotide diversity per base pair; H_e is haplotype diversity; *K* is the number of haplotypes; PHI test is the pairwise homoplasy test; *S* is the number of segregating sites. 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.

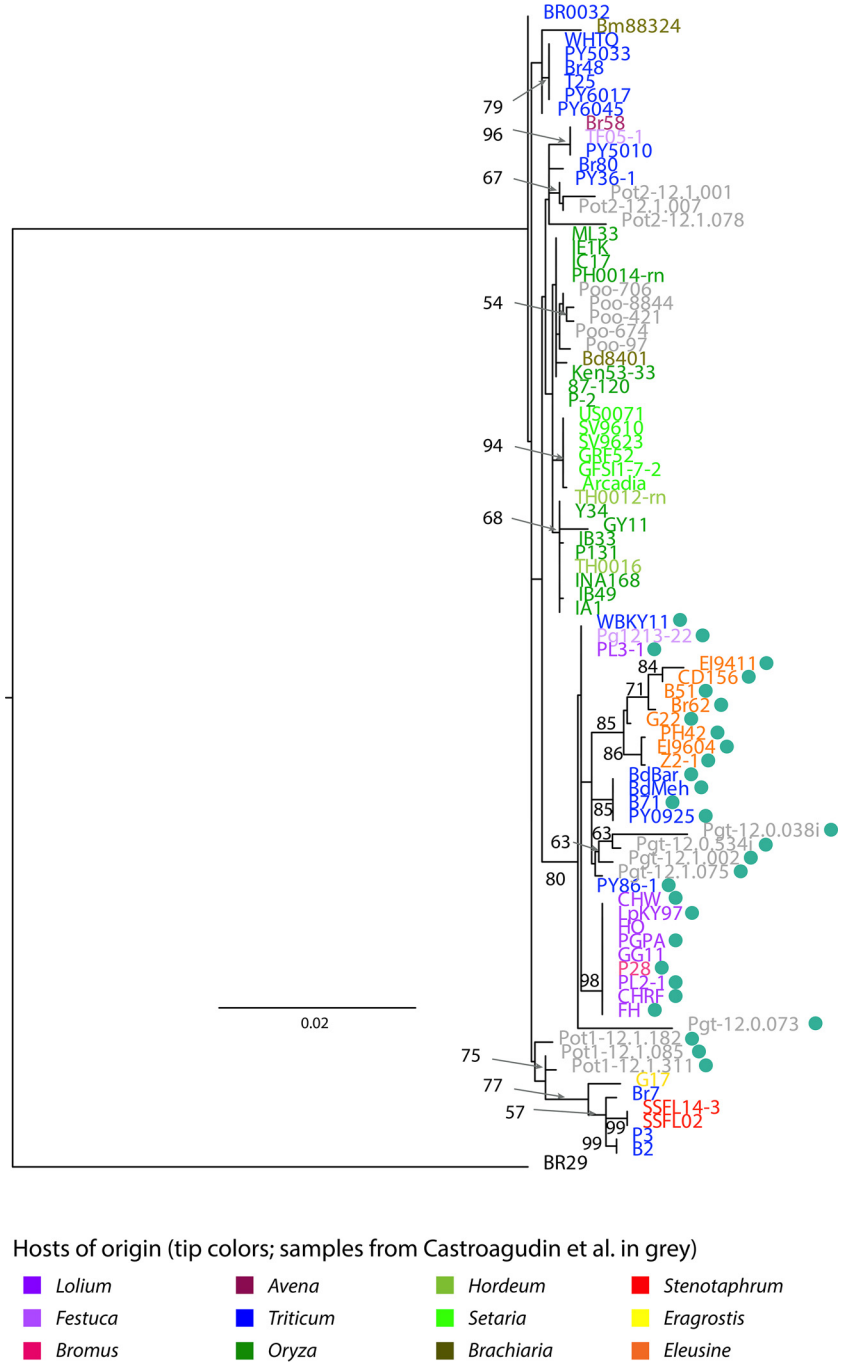


FIG 4 Maximum-likelihood tree based on concatenated data set comprising nine loci used in the work of Castroagudin et al. (24), retrieved from 76 *M. oryzae* genomes. Numbers above branches represent bootstrap supports after 100 bootstrap replicates. Only nodes with bootstrap support of >50 are labeled. Representatives of isolates used by Castroagudin et al. (24) in their study were included in the analysis and are colored in light gray. Green dots mark the strains containing the *P. graminis-tritici*-type allele according to the work of Castroagudin et al. (24).

isolates from the Castroagudin et al. study were contained in a clade with 80% support. Inspection of the *MPG1* marker that was reported to be diagnostic for *P. graminis-tritici* (Castroagudin et al. [24]) revealed that all of the isolates in this clade contained the *P. graminis-tritici*-type allele (green dots) and should therefore be classified as *P. graminis-tritici* (Fig. 4). Critically, however, a few isolates outside this clade also harbored the *P. graminis-tritici*-type allele. Moreover, the clade also included isolates from the present

study which came from wheat, annual ryegrass, perennial ryegrass, tall fescue, finger millet, and goosegrass—isolates that did not group together in the DAPC analysis (Fig. 1) or in the ML and NJ trees built using the orthologous genes or whole-genome SNP data (Fig. 2 and 3). Isolates carrying the *P. graminis-tritici*-type allele were in fact distributed among three genetically distinct and well-supported clades (Fig. 2 and 3). Furthermore, visual inspection of the topologies and bootstrap supports for each single-locus tree revealed that GCPSR criteria were not satisfied for the clade including all of the *P. graminis-tritici* isolates from the Castroagudin et al. study. Thus, isolates characterized by Castroagudin et al. (24) as *P. graminis-tritici* fail to constitute a phylogenetically cohesive group based on total genome evidence, and thus, the existence of the *P. graminis-tritici* species is not supported by our new genome-wide data and analyses.

The basis for the previous designation of *P. graminis-tritici* 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 (Fig. S5). As an example, isolates from the *Triticum* lineage carried three different *MPG1* alleles. Two of these (including the *P. graminis-tritici* type) were also present in the *Lolium* lineage, while the third *MPG1* (ACT17T-C-6CAA140, Fig. S5) was shared by distantly related isolates from the *Stenotaphrum* lineage (Fig. S5). Isolates from the *Eleusine* lineage also carried the *P. graminis-tritici*-type *MPG1* allele and two other variants, while isolates from the *Setaria* and *Oryza* lineages carried an *MPG1* allele distinct from all the others (Fig. 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 (Fig. 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 (Fig. 2 and 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 reassess 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 Fig. 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 data set 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).

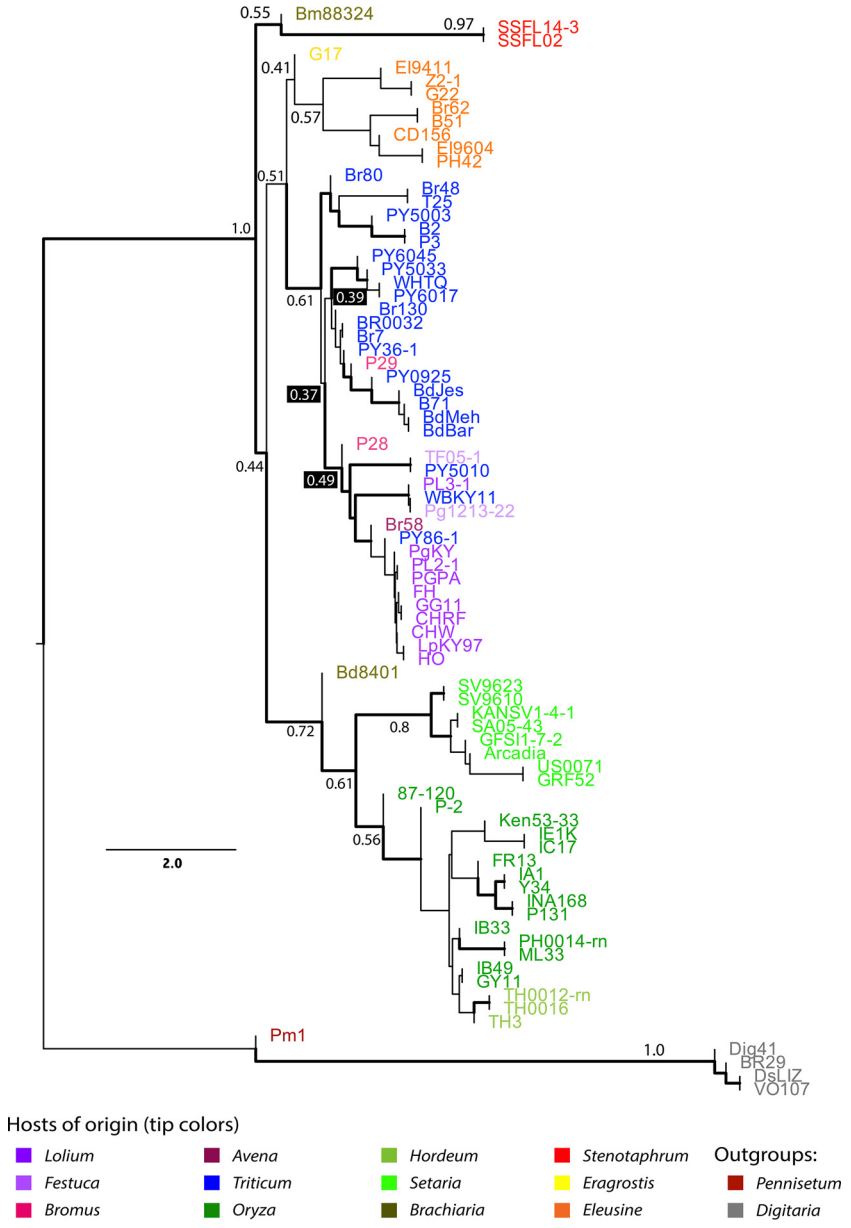


FIG 5 ASTRAL analysis to test for incomplete lineage sorting/gene flow among 81 *Magnaporthe* genomes, using 2,241 single-copy orthologous sequence loci. Thicker branches represent branches that have a bootstrap support of >50 after multilocus bootstrapping. Numbers above branches represent q1 local 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 shown on black background for branches holding wheat blast isolates.

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 according to the nondiscordance 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

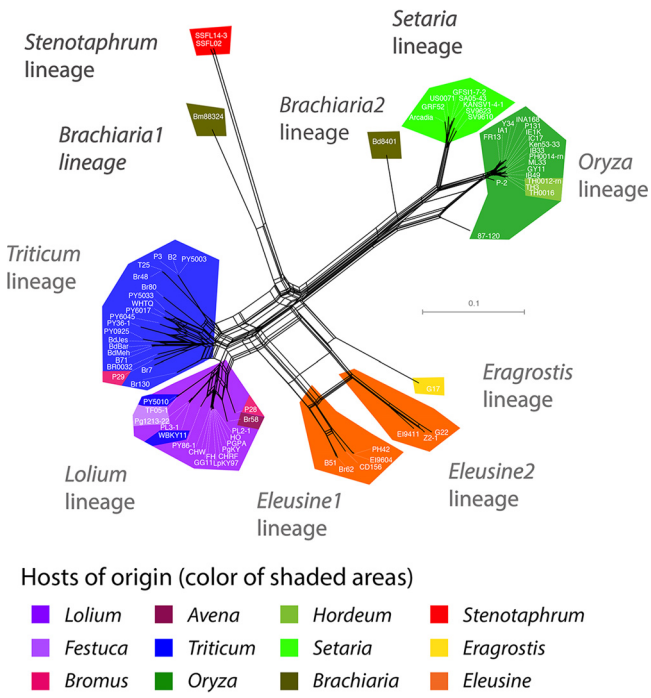


FIG 6 Neighbor-Net network built with SplitsTree. The figure shows relationships between haplotypes identified based on the full set of 25,078 SNPs identified in 2,682 single-copy orthologs, excluding sites with missing data, gaps, and singletons.

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 (Fig. 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 be produced only by incomplete lineage sorting and should be equally infrequent (34, 35). 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), according to the following 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 *Brachiaria*2/*Setaria*/*Oryza*, excluding (P1,P2) pairs already used in principle 1) 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 *Brachiaria*2/*Setaria*/*Oryza*) and *Stenotaphrum* or *Brachiaria*1 as P3 carried information about gene flow across a relatively long time period (Fig. 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 (Fig. 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 a private allele found no gene, among the 2,241 genes surveyed, carrying mutations exclusive to a single lineage. Together, these results indicate that gene flow was widespread, across both 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 (Fig. S3). Structure uses Markov chain Monte Carlo (MCMC) 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 (Fig. 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 into multiple lineages preferentially associated with one host plant genus. Neither the reanalysis 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 the 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 a valid name for wheat-infecting strains, because (i) 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, and (ii) 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 a 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 ecologic specialization following host shifts, suggesting that the observed pattern of population subdivision in *M. oryzae* actually corresponds to ongoing speciation. Previous experimental measurements of perithecium formation and ascospore production—two important components of reproductive success—suggested interfertility *in vitro* between most pairs of lineages with high levels of ascospore viability (40–43). This suggests that intrinsic pre- or postmating reproductive barriers, such as assortative mating by mate choice or gametic incompatibility, and zygotic mortality, are not responsible for the relative reproductive isolation between lineages—which creates the observed pattern of population subdivision. Instead, the relative reproductive isolation between lineages could be caused by one or several pre- or postmating barriers (see Table 1 in reference 44), such as mating-system isolation or hybrid sterility (intrinsic barrier), or difference in mating times, difference in mating sites, immigrant inviability, or ecologically based hybrid inviability (extrinsic barriers).

Previous pathogenicity assays revealed extensive variability in the host range of *M. oryzae* isolates, in terms of both pathogenicity toward a set of host species and pathogenicity toward a set of genotypes from a given host (40, 41). Indeed, extensive genetic analyses show that host species specificity in *M. oryzae*, similar to rice cultivar specificity, could be controlled by a gene-for-gene relationship in which one to three avirulence genes in the fungus prevent infection of particular host species (43, 45, 46). Loss of the avirulence genes would allow infection of novel hosts to occur. Additionally, host species specificity is not strictly maintained. Under controlled conditions, most lineages have at least one host in common (40), and strains within one lineage can still cause rare susceptible lesions on naive hosts (21, 47). Moreover, a single plant infected by a single genotype can produce large numbers of spores in a single growing season (48), allowing the pathogen to persist on an 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 to 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 ecologic 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 reducing gene flow between *M. oryzae* lineages. The fact that mating in *M. oryzae* likely occurs within host tissues, such as dead stems (49), 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 (50). 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.

Interlineage gene flow. Several potential barriers contribute to reduce genetic exchanges between *M. oryzae* lineages (see above), 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 interlineage 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 interlineage gene flow (51). 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 a 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 (http://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 ecoevolutionary 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 data set in terms of number of loci and of range of variation in origin (geography and host) of isolates can lead to erroneous identification of fungal cryptic species. Our work illustrates the growing divide between taxonomy that "creates the language of biodiversity" (52) 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, 53–55), 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 in the supplemental material. For newly sequenced genomes provided by M.F. and B.V., 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 contig files in a nonredundant fashion to generate a final assembly. Newly sequenced genomes BR130 and WHTQ, provided by T.M., were sequenced using an Illumina paired-end sequencing approach at >50× depth. Short reads were assembled *de novo* using Velvet 1.2.10 (56), resulting in a 41.5-Mb genome for BR130 with an N_{50} of 44.8 kb and a 43.7-Mb genome for WHTQ with an N_{50} of 36.2 kb. For newly sequenced genomes provided by D.S. and N.J.T., DNA was sequenced on the Illumina HiSeq 2500 sequencer, producing 100 nucleotide 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 (56).

Orthologous gene identification in genomic sequences. Protein-coding gene models were predicted using Augustus V3.0.3 (57). Orthologous genes were identified in the 76 genomes of *M. oryzae* or in the data set including outgroups using ProteinOrtho (58). The v8 version of the 70-15 *M. oryzae* reference genome (59) 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 ortholog sets were aligned using MACSE (60). 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 data set and 2,682 alignments for the 76 *M. oryzae* strains.

Population subdivision and summary statistics of polymorphism and divergence. Population subdivision was analyzed using DAPC and Structure (30, 36–38), based on multilocus haplotype profiles identified from ortholog alignments using a custom Python script. DAPC was performed using the Adegenet package in R (13). We retained the first 30 principal components and the first 4 discriminant functions. Ten independent Structure runs were carried out for each number of clusters *K*, with 100,000 MCMC iterations after a burn-in of 50,000 steps.

Polymorphism statistics were computed using EggLib 3.0.0b10 (61) excluding sites with >30% missing data. Divergence statistics were computed using a custom perl script.

To infer total-evidence trees within the 76 *M. oryzae* strains (respectively within the 81 *Magnaporthe* strains), all sequences from the 2,682 (respectively 2,241) orthologous sequences were concatenated. The maximum-likelihood tree was inferred using RAxML (62) with the general time reversible (GTR)-gamma model, and bootstrap supports were estimated after 1,000 replicates.

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

Secondary data analysis. Species recognition based on multiple gene genealogies as described by Castroagudin et al. (24) was repeated according to the reported methods. The robustness of the *P. graminis-tritici* species inference was tested by reiterating the study, omitting one marker at a time. Individual genealogies were built using RAxML with the GTR-gamma model and 100 bootstrap replicates.

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

***MPG1*-based classification.** The *MPG1* hydrophobin sequence is described as being diagnostic for the *P. graminis-tritici*/*M. oryzae* species split (24). *MPG1* sequences from one of each species (gene identifiers [GIs] KU952644.1 for *P. graminis-tritici* and KU952661.1 for *M. oryzae*) were used as BLAST (63) 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 (64), 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 reference 35. Significance was calculated using the block jackknife approach (100 replicates, 1,000 SNP blocks), to account for nonindependence among sites.

Probabilistic chromosome painting. We used the 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, and rounded to the nearest thousand for N_e ($N_e = 5{,}000$; $\mu = 0.0009$). The file *recom_rate_infile* detailing the recombination rate between SNPs was built using the Interval program in LDhat version 2.2 (65) based on the whole data set 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 three times to check for convergence.

Phylogenetic species recognition. We used an implementation of the GCPSR scalable to genomic data (https://github.com/b-brankovics/GCPSR) (69). The method works in two steps. (i) Concordance and nondiscordance analysis produces a genealogy that has clades that are both concordant and nondiscordant 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%. (ii) 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 ensure having 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 2,241 genes. The concordance and nondiscordance 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 1,121.

Whole-genome alignment and tree building. A custom perl script was used to mask sequences that occur in multiple alignments when the genome is subjected to BLAST analysis against itself. The masked genomes were then aligned in a pairwise fashion against all other genomes using BLAST (63). Regions that did not uniquely align in each pair at a threshold of 1e−200 were excluded. SNPs were then identified for each pairwise comparison and scaled by the total number of nucleotides aligned after excluding repetitive and duplicate regions. This produced a distance metric of SNPs per megabase of uniquely aligned DNA. The pairwise distances were used to construct phylogenetic trees with the

neighbor-joining method as implemented in the R package Analyses of Phylogenetics and Evolution (APE) (66).

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

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.01219-17.

TEXT S1, PDF file, 0.9 MB.
FIG S1, PDF file, 0.5 MB.
FIG S2, PDF file, 0.1 MB.
FIG S3, PDF file, 1 MB.
FIG S4, PDF file, 0.1 MB.
FIG S5, PDF file, 0.9 MB.
FIG S6, PDF file, 0.3 MB.
TABLE S1, XLSX file, 0.1 MB.
TABLE S2, XLSX file, 0.1 MB.
TABLE S3, PDF file, 0.1 MB.

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 by the Agriculture and Food Research Initiative Competitive grant no. 2013-68004-20378 from the USDA National Institute of Food and Agriculture is acknowledged.

REFERENCES

1. Fournier E, Giraud T, Albertini C, Brygoo Y. 2005. Partition of the *Botrytis cinerea* complex in France using multiple gene genealogies. Mycologia 97:1251–1267. https://doi.org/10.3852/mycologia.97.6.1251.
2. Le Gac M, Hood ME, Fournier E, Giraud T. 2007. Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. Evolution 61:15–26. https://doi.org/10.1111/j.1558-5646.2007.00002.x.
3. Taylor JW, Fisher MC. 2003. Fungal multilocus sequence typing—it's not just for bacteria. Curr Opin Microbiol 6:351–356. https://doi.org/10.1016/S1369-5274(03)00088-2.
4. Planet PJ, Narechania A, Chen L, Mathema B, Boundy S, Archer G, Kreiswirth B. 2017. Architecture of a species: phylogenomics of *Staphylococcus aureus*. Trends Microbiol 25:153–166. https://doi.org/10.1016/j.tim.2016.09.009.
5. Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre GP, Bank C, Brännström A, Brelsford A, Clarkson CS, Eroukhmanoff F, Feder JL, Fischer MC, Foote AD, Franchini P, Jiggins CD, Jones FC, Lindholm AK, Lucek K, Maan ME, Marques DA, Martin SH, Matthews B, Meier JI, Möst M, Nachman MW, Nonaka E, Rennison DJ, Schwarzer J, Watson ET, Westram AM, Widmer A. 2014. Genomics and the origin of species. Nat Rev Genet 15:176–192. https://doi.org/10.1038/nrg3644.
6. Giraud T, Gladieux P, Gavrilets S. 2010. Linking the emergence of fungal plant diseases with ecological speciation. Trends Ecol Evol 25:387–395. https://doi.org/10.1016/j.tree.2010.03.006.
7. Giraud T, Villaréal LM, Austerlitz F, Le Gac M, Lavigne C. 2006. Importance of the life cycle in sympatric host race formation and speciation of pathogens. Phytopathology 96:280–287. https://doi.org/10.1094/PHYTO-96-0280.
8. Klaubauf S, Tharreau D, Fournier E, Groenewald JZ, Crous PW, de Vries RP, Lebrun MH. 2014. Resolving the polyphyletic nature of *Pyricularia* (Pyriculariaceae). Stud Mycol 79:85–120. https://doi.org/10.1016/j.simyco.2014.09.004.
9. Ou SH. 1980. A look at worldwide rice blast disease control. Plant Dis 64:439–445. https://doi.org/10.1094/PD-64-439.
10. Ou SH. 1985. Rice diseases. Commonwealth Agricultural Bureau, Slough, United Kingdom.
11. Couch BC, Fudal I, Lebrun MH, Tharreau D, Valent B, van Kim P, Notteghem JL, Kohn LM. 2005. Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. Genetics 170:613–630. https://doi.org/10.1534/genetics.105.041780.
12. Farman ML. 2002. *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 92:245–254. https://doi.org/10.1094/PHYTO.2002.92.3.245.
13. Chiapello H, Mallet L, Guerin C, Aguileta G, Amselem J, Kroj T, Ortega-Abboud E, Lebrun MH, Henrissat B, Gendrault A, Rodolphe F, Tharreau D, Fournier E. 2015. Deciphering genome content and evolutionary relationships of isolates from the fungus *Magnaporthe oryzae* attacking different host plants. Genome Biol Evol 7:2896–2912. https://doi.org/10.1093/gbe/evv187.
14. Yoshida K, Saunders DG, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, Inoue Y, Chuma I, Tosa Y, Cano LM, Kamoun S, Terauchi R. 2016. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics 17:370. https://doi.org/10.1186/s12864-016-2690-6.
15. Islam MT, Croll D, Gladieux P, Soanes DM, Persoons A, Bhattacharjee P, Hossain MS, Gupta DR, Rahman MM, Mahboob MG, Cook N, Salam MU, Surovy MZ, Sancho VB, Maciel JL, Nhani Júnior A, Castroagudín VL, Reges JT, Ceresini PC, Ravel S, Kellner R, Fournier E, Tharreau D, Lebrun MH,

McDonald BA, Stitt T, Swan D, Talbot NJ, Saunders DG, Win J, Kamoun S. 2016. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. BMC Biol 14:84. https://doi.org/10.1186/s12915-016-0309-7.

16. Igarashi S. 1990. Update on wheat blast (Pyricularia oryzae) in Brazil, p 480–483. *In* Saunders D (ed), Proceedings of the International Conference—Wheat for the Nontraditional Warm Areas. CIMMYT, Mexico City, Mexico.
17. Cabrera M, Gutiérrez S. 2007. Primer registro de *Pyricularia grisea* en cultivos de trigo del NE de Argentina. Depto. Protección Vegetal, Facultad de Ciencias Agrarias, Universidad Nacional del Nordeste, Corrientes, Argentina. http://www.agr.unne.edu.ar/images/documentos/extension/Res2007/SanVegetal/SanVegetal_06.pdf.
18. Viedma L. 2005. Wheat blast occurrence in Paraguay. Phytopathology 95:S152.
19. Kohli MM, Mehta YR, Guzman E, De Viedma L, Cubilla LE. 2011. *Pyricularia* blast—a threat to wheat cultivation. Czech J Genet Plant Breed 47:S130–S134.
20. Maciel JLN, Ceresini PC, Castroagudin VL, Zala M, Kema GHJ, McDonald BA. 2014. Population structure and pathotype diversity of the wheat blast pathogen *Magnaporthe oryzae* 25 years after its emergence in Brazil. Phytopathology 104:95–107. https://doi.org/10.1094/PHYTO-11-12-0294-R.
21. Cruz CD, Valent B. 2017. Wheat blast disease: danger on the move. Trop Plant Pathol 42:210–222. https://doi.org/10.1007/s40858-017-0159-z.
22. Farman ML, Peterson GL, Chen L, Starnes JH, Valent B, Bachi P, Murdock L, Hershman D, Pedley K, Fernandes JM, Bavaresco J. 2017. The Lolium pathotype of *Magnaporthe oryzae* recovered from a single blasted wheat plant in the United States. Plant Dis 101:684–692. https://doi.org/10.1094/PDIS-05-16-0700-RE.
23. Malaker PK, Barma NCD, Tiwari TP, Collis WJ, Duveiller E, Singh PK, Joshi AK, Singh RP, Braun H-J, Peterson GL, Pedley KF, Farman ML, Valent B. 2016. First report of wheat blast caused by *Magnaporthe oryzae* pathotype *triticum* in Bangladesh. Plant Dis 100:2330. https://doi.org/10.1094/PDIS-05-16-0666-PDN.
24. Castroagudin VL, Moreira SI, Pereira DAS, Moreira SS, Brunner PC, Maciel JLN, Crous PW, McDonald B, Alves E, Ceresini PC. 2016. Wheat blast disease caused by *Pyricularia graminis-tritici* sp. nov. Persoonia 37:199–216. https://doi.org/10.3767/003158516X692149.
25. Dettman JR, Jacobson DJ, Taylor JW. 2003. A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. Evolution 57:2703–2720. https://doi.org/10.1111/j.0014-3820.2003.tb01514.x.
26. Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet Biol 31:21–32. https://doi.org/10.1006/fgbi.2000.1228.
27. Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541–i548. https://doi.org/10.1093/bioinformatics/btu462.
28. Mirarab S, Warnow T. 2015. Astral II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44–i52. https://doi.org/10.1093/bioinformatics/btv234.
29. Sayyari E, Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol Biol Evol 33:1654–1668. https://doi.org/10.1093/molbev/msw079.
30. Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94. https://doi.org/10.1186/1471-2156-11-94.
31. Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267. https://doi.org/10.1093/molbev/msj030.
32. Edwards SV, Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, Zhong B, Wu S, Lemmon EM, Lemmon AR, Leaché AD, Liu L, Davis CC. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol Phylogenet Evol 94:447–462. https://doi.org/10.1016/j.ympev.2015.10.027.
33. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res 23:1817–1828. https://doi.org/10.1101/gr.159426.113.
34. Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol Biol Evol 32:244–257. https://doi.org/10.1093/molbev/msu269.
35. Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. Mol Biol Evol 28:2239–2252. https://doi.org/10.1093/molbev/msr048.
36. Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.
37. Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.
38. Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x.
39. Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. PLoS Genet 8:e1002453. https://doi.org/10.1371/journal.pgen.1002453.
40. Kato H, Yamamoto M, Yamaguchi-Ozaki T, Kadouchi H, Iwamoto Y, Nakayashiki H, Tosa Y, Mayama S, Mori N. 2000. Pathogenicity, mating ability and DNA restriction fragment length polymorphisms of *Pyricularia* populations isolated from Gramineae, Bambusideae and Zingiberaceae plants. J Gen Plant Pathol 66:30–47. https://doi.org/10.1007/PL00012919.
41. Urashima AS, Igarashi S, Kato H. 1993. Host range, mating type and fertility of *Pyricularia grisea* from wheat in Brazil. Plant Dis 77:1211–1216. https://doi.org/10.1094/PD-77-1211.
42. Orbach MJ, Chumley FG, Valent B. 1996. Electrophoretic karyotypes of Magnaporthe grisea pathogens of diverse grasses. Mol Plant Microbe Interact 9:261–271.
43. Thi T, Vy P, Hyon G-S, Thi N, Nga T, Inoue Y, Chuma I, Tosa Y. 2014. Genetic analysis of host-pathogen incompatibility between Lolium isolates of *Pyricularia oryzae* and wheat. J Gen Plant Pathol 80:59. https://doi.org/10.1007/s10327-013-0478-y.
44. Nosil P, Vines TH, Funk DJ. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719. https://doi.org/10.1111/j.0014-3820.2005.tb01747.x.
45. Takabayashi N, Tosa Y, Oh HS, Mayama S. 2002. A gene-for-gene relationship underlying the species-specific parasitism of Avena/triticum isolates of *Magnaporthe grisea* on wheat cultivars. Phytopathology 92:1182–1188. https://doi.org/10.1094/PHYTO.2002.92.11.1182.
46. Tosa Y, Tamba H, Tanaka K, Mayama S. 2006. Genetic analysis of host species specificity of *Magnaporthe oryzae* isolates from rice and wheat. Phytopathology 96:480–484. https://doi.org/10.1094/PHYTO-96-0480.
47. Heath MC, Valent B, Howard RJ, Chumley FG. 1990. Interactions of two strains of *Magnaporthe grisea* with rice, goosegrass, and weeping lovegrass. Can J Bot 68:1627–1637. https://doi.org/10.1139/b90-209.
48. Gurr S, Samalova M, Fisher M. 2011. The rise and rise of emerging infectious fungi challenges food security and ecosystem health. Fungal Biol Rev 25:181–188. https://doi.org/10.1016/j.fbr.2011.10.004.
49. Silué D, Notteghem JL. 1990. Production of perithecia of *Magnaporthe grisea* on rice plants. Mycol Res 94:1151–1152. https://doi.org/10.1016/S0953-7562(09)81351-8.
50. Saleh D, Xu P, Shen Y, Li C, Adreit H, Milazzo J, Ravigné V, Bazin E, Nottéghem JL, Fournier E, Tharreau D. 2012. Sex at the origin: an Asian population of the rice blast fungus *Magnaporthe oryzae* reproduces sexually. Mol Ecol 21:1330–1344. https://doi.org/10.1111/j.1365-294X.2012.05469.x.
51. Lemaire C, De Gracia M, Leroy T, Michalecka M, Lindhard-Pedersen H, Guerin F, Gladieux P, Le Cam B. 2016. Emergence of new virulent populations of apple scab from nonagricultural disease reservoirs. New Phytol 209:1220–1229. https://doi.org/10.1111/nph.13658.
52. Hibbett DS, Taylor JW. 2013. Fungal systematics: is a new age of enlightenment at hand? Nat Rev Microbiol 11:129–133. https://doi.org/10.1038/nrmicro2963.
53. Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, Tosa Y, Chuma I, Takano Y, Win J, Kamoun S, Terauchi R. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. Plant Cell 21:1573–1591. https://doi.org/10.1105/tpc.109.066324.
54. Xue M, Yang J, Li Z, Hu S, Yao N, Dean RA, Zhao W, Shen M, Zhang H, Li C, Liu L, Cao L, Xu X, Xing Y, Hsiang T, Zhang Z, Xu JR, Peng YL. 2012. Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. PLoS Genet 8:e1002869. https://doi.org/10.1371/journal.pgen.1002869.
55. Zhong Z, Norvienyeku J, Chen M, Bao J, Lin L, Chen L, Lin Y, Wu X, Cai Z, Zhang Q, Wang B, Wang Z. 2016. Directional selection from host

plants is a major force driving host specificity in *Magnaporthe* species. Sci Rep 6:25591. https://doi.org/10.1038/srep25591.
56. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.
57. Keller O, Kollmar M, Stanke M, Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics 27:757–763. https://doi.org/10.1093/bioinformatics/btr010.
58. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 12:124. https://doi.org/10.1186/1471-2105-12-124.
59. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND, Lee YH, Carbone I, Brown D, Oh YY, Donofrio N, Jeong JS, Soanes DM, Djonovic S, Kolomiets E, Rehmeyer C, Li W, Harding M, Kim S, Lebrun MH, Bohnert H, Coughlan S, Butler J, Calvo S, Ma LJ, Nicol R, Purcell S, Nusbaum C, Galagan JE, Birren BW. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434:980–986. https://doi.org/10.1038/nature03449.
60. Ranwez V, Harispe S, Delsuc F, Douzery EJ. 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PLoS One 6:e22594. https://doi.org/10.1371/journal.pone.0022594.
61. De Mita S, Siol M. 2012. EggLib: processing, analysis and simulation tools for population genetics and genomics. BMC Genet 13:27. https://doi.org/10.1186/1471-2156-13-27.
62. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
63. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.
64. Huson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14:68–73. https://doi.org/10.1093/bioinformatics/14.1.68.
65. Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots. Genome Res 17:1219–1227. https://doi.org/10.1101/gr.6386707.
66. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290. https://doi.org/10.1093/bioinformatics/btg412.
67. Pieck ML, Ruck A, Farman ML, Peterson GL, Stack JP, Valent B, Pedley KF. 2017. Genomics-based marker discovery and diagnostic assay development for wheat blast. Plant Dis 101:103–109. https://doi.org/10.1094/PDIS-04-16-0500-RE.
68. Inoue Y, Vy TTP, Yoshida K, Asano H, Mitsuoka C, Asuke S, Anh VL, Cumagun CJR, Chuma I, Terauchi R, Kato K, Mitchell T, Valent B, Farman M, Tosa Y. 2017. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. Science 357:80–83. https://doi.org/10.1126/science.aam9654.
69. Brankovics B, van Dam P, Rep M, de Hoog GS, van der Lee TA, Waalwijk C, van Diepeningen AD. 2017. Mitochondrial genomes reveal recombination in the presumed asexual Fusarium oxysporum species complex. BMC Genomics 18:735.