A Re-evaluation of Phylogenomic Data Reveals that Current Understanding in Wheat Blast Population Biology and Epidemiology is Obfuscated by Oversights in Population Sampling

Mark L. Farman[1], Joao P. Ascari[2], Mostafa Rahnama[1], Emerson M. Del Ponte[2], Kerry F. Pedley[3], Sebastián Martinez[4], José Maurício C. Fernandes[5], and Barbara Valent[6]

[1] Department of Plant Pathology, University of Kentucky, Lexington, KY, 40546 USA
[2] Departamento de Fitopatologia, Universidade Federal de Viçosa, Viçosa, MG, 36570-900 Brazil
[3] USDA/ARS/Foreign Disease Weed Science Research Unit, Fort Detrick, MD 21702
[4] Instituto Nacional de Investigación Agropecuaria, INIA Treinta y Tres, 33000 Treinta y Tres, Uruguay
[5] Embrapa Trigo, Passo Fundo, RS, 99050-970 Brazil
[6] Department of Plant Pathology, Kansas State University, Manhattan, KS, 66506, USA

* Corresponding author: farman@uky.edu

ABSTRACT

Wheat blast, caused by the *Pyricularia oryzae Triticum* lineage (PoT), first emerged in Brazil and quickly spread to neighboring countries. Its recent appearance in Bangladesh and Zambia highlights a need to understand the disease's population biology and epidemiology so as to mitigate pandemic outbreaks. Current knowledge is mostly based on characterizations of Brazilian wheat blast isolates and comparison with isolates from non-wheat, endemic grasses. These foregoing studies concluded that the wheat blast population lacks host specificity and, as a result, undergoes extensive gene flow with populations infecting non-wheat hosts. Additionally, based on genetic similarity between wheat blast and isolates infecting *Urochloa* species, it was proposed that the disease originally emerged via a host jump from this grass, and that *Urochloa* likely plays a central role in wheat blast epidemiology, owing to its widespread use as a pasture grass. However, due to inconsistencies with broader phylogenetic studies, we suspected that these seminal studies hadn't actually sampled the populations normally found on endemic grasses and, instead, had repeatedly isolated members of PoT and the related *Lolium* pathogen lineage (PoL1). Re-analysis of the Brazilian data as part of a comprehensive, global, phylogenomic dataset that included a small number of S. American isolates sampled away from wheat confirmed our suspicion and identified four new *P. oryza*e lineages on grass hosts. As a result, the conclusions underpinning current understanding in wheat blast's evolution, population biology and epidemiology are unsubstantiated and could be equivocal.

INTRODUCTION

Wheat blast (also known as brusone) is a serious disease of wheat caused by the fungus *Pyricularia oryzae*. Having recently spread to Asia and Africa (Malaker et al. 2016; Tembo et al. 2020), it now poses a serious threat to global wheat production (Ceresini et al. 2018; Cruz and Valent 2017; Singh et al. 2021). As a recently emerged disease that first surfaced in Paraná state, Brazil in 1985 (Igarashi et al. 1986), wheat blast promises to be a valuable model for understanding how new diseases evolve; how newly-evolved populations are structured and change over time; how they interact with endemic *P. oryzae* populations in the field; and, then, finally, what are the population genetic consequences of invading new continents. Intimate knowledge of pathogen population structure/genetics also has practical benefits in disease control as it can provide deep insights into disease epidemiology, and underpins a solid theoretical foundation for the development of diagnostic tools and quarantine guidelines. Here, it should go without saying that accurate inference of population structure is absolutely critical to success in these endeavors (Milgroom 2017).

Despite a long history of research in blast diseases (most notably rice blast), the overall low level of sequence divergence between *P. oryzae* isolates (~1%) prevented significant progress in understanding its population structure until the development of hypervariable molecular markers (simple sequence repeats - SSRs), and next generation sequencing. Ceresini and coworkers were pioneers with these technologies and used SSRs (Maciel et al. 2014), multilocus sequencing (Castroagudín et al. 2016), and eventually whole genome sequencing (Ceresini et al. 2018, 2019) to characterize the wheat blast population. They showed that fungal isolates causing wheat blast are phylogenetically distinct from those found on rice, *Digitaria*, *Eleusine*, *Setaria* and other grasses, and renamed the population as a new species, *Pyricularia graminis tritici* (Pygt) to reflect this fact (Castroagudín et al. 2016; Ceresini et al. 2019). More recently, however, evidence of gene flow (Gladieux et al. 2018) among other concerns (Valent et al. 2019), has raised this conclusion into doubt.

Early population studies of wheat blast using simple sequence repeats (SSRs) found these markers to be in equilibrium, while at the same time clonal lineages were identified. This implied that the population has a mixed reproductive mode, where sexual cycles generate diversity and well-adapted clones are then propagated vegetatively (Maciel et al. 2014). Ceresini and coworkers then compared wheat blast strains with fungal isolates found on neighboring grasses and weeds. Using SSRs, and then genome sequence data, they found that many grass-infecting isolates exhibited a high degree of genetic similarity to isolates found on wheat (Castroagudin et

al. 2017; Ceresini et al. 2018, 2019). Also detected was evidence for significant gene flow between the grass- and wheat-infecting populations, with the predominant migration being from the former to the latter (Castroagudin et al. 2017). When combined with the observations that several isolates from grasses were capable of causing disease on wheat in inoculation assays; virulence phenotypes were often shared between the wheat and non-wheat host groups (Castroagudin et al. 2017); and the discovery of inter-fertility between isolates from wheat and other Poaceae (Bruno and Urashima 2001; Galbieri and Urashima 2008; Urashima et al. 1993); this led Ceresini and coworkers to propose that the wheat blast fungus mates preferentially on non-wheat hosts, producing an ascospore population with high diversity which then infects nearby wheat (Ceresini et al., 2018, 2019). An extended conclusion from these findings was that wheat blast (Pygt) is not a wheat-specialized pathogen so that *P. oryzae* - being predominantly host-specialized - is not a good model for studying Pygt biology (Ceresini et al., 2018, 2019).

With regard to the grass-infecting populations, significant focus has been placed on the one infecting *Urochloa* because the wheat blast pathogen reportedly bore the strongest similarity to isolates from this host. This led to the conclusion that a host jump by a *Urochloa* pathogen was a key step in the evolution of wheat blast (Stukenbrock and McDonald 2008). Furthermore, because *Urochloa* is widely used as a forage grass in Brazil, and is often grown in close proximity to wheat, it has been proposed that signalgrass pastures are significant inoculum reservoirs for pathogen over-wintering, and serve as a bridge that facilitates pathogen spread and gene flow between regions (Ceresini et al. 2019).

Unfortunately, if one peruses the phylogenomic data that underpin the foregoing conclusions, and considers them in the light of data produced by the broader research community, major inconsistencies quickly appear with potentially far-reaching consequences for foundational knowledge on wheat blast. Ceresini and coworkers' earliest studies found that wheat blast isolates could be grouped in two main clades, with only one (Pygt) being clearly resolved from the rice blast pathogens (Castroagudín et al. 2016). Later, as genomic sequence data provided greater resolution, additional wheat blast isolates were brought under the Pygt umbrella, along with some of the isolates from grasses (Castroagudin et al. 2016, 2017; Ceresini et al. 2018, 2019). However, the *Pygt/P. oryzae* boundary shifted every time new data were added (compare Castroagudin et al., 2017 vs. Ceresini et al., 2018 vs. Ceresini et al., 2019). The shifting boundary not only raises doubt about Pygt's status as a new species but it also highlights a major problem with the reported relationship between Pygt and grass-infecting *P. oryzae*. In particular, the lack of phylogenetic resolution between isolates from many different host genera goes against prior work which showed that virtually all *P. oryzae* from non-wheat hosts not only grouped according to their host

of origin, but also showed complete and unequivocal phylogenetic resolution from PoT/PoL1 (Gladieux et al. 2018). Indeed, this tendency for groups of phylogenetically-related isolates (lineages) to be found on the same host (with few exceptions) is what gives *P. oryzae* its reputation as a host-specialized pathogen.

That virtually all grass-infecting isolates collected elsewhere around the globe group according to host-of-origin and are genetically very distinct from isolates found on wheat (Gladieux et al. 2018) led us to suspect that Ceresini and coworkers - having only collected isolates from grasses growing near diseased wheat (suppl. Fig. S1) - might have failed to sample the true endemic grass-infecting populations in Brazil, and instead, repeatedly recovered cross-infecting PoT (or PoL1) . To test this possibility, we performed a new phylogenomic analysis that integrated their data with those generated by the broader *P. oryzae* research community - with the prediction that most of their isolates would group with PoT/PoL1. Then, to rule out the possibility that the *P. oryzae* populations found on grasses in South America are simply different to those found everywhere else around the globe, we used genome sequencing to survey a modest number of isolates collected from endemic grasses in Uruguay, and at locations distant from wheat-growing regions in Minas Gerais, Brazil.

MATERIAL AND METHODS

Fungal isolates. To maximize our chances of characterizing the *P. oryzae* populations normally found in association with endemic grasses, we performed sampling at varying distances from regions affected by wheat blast (0 m to ~750 km), and during different seasons (Ascari et al. submitted). A geographic map was produced to indicate where the non-wheat isolates were collected, together with information on their hosts of origin and distance to wheat fields. The map was produced using the R packages rnaturalearth, rnaturalearthhires and scatterpie (Yu 2021; South 2021, 2022) (Fig. S1).

Diseased tissues were placed in a dew chamber overnight to induce sporulation. Spores were then collected by brushing the conidiophores lightly with a sterile, sealed Pasteur pipette and then spread across 3% water agar. A single, germinated spore was then transferred to oatmeal agar supplemented with ampicillin (100 µg/ml). The colony was allowed to colonize whatman 3M paper squares placed on the agar surface, which were then collected, placed in glassine envelopes and dried in a containment hood for 2 d. The cultures were stored desiccated at -20°. The isolates that are the central focus of this paper are listed in Table 1.

DNA extraction. Single-spored isolates were cultured for 7 d with shaking on 10 ml of liquid complete medium (Valent et al. 1984). The mycelial ball was grabbed with forceps, patted dry on paper towels, placed in a 15 ml plastic, conical tube, frozen at -20 °C, and then freeze-dried overnight (https://youtu.be/h9TZANDMnd8). A glass rod was then used to grind the freeze-dried pellet to a fine powder against the side of the tube (https://youtu.be/RR0qwc3liEI). One milliliter of lysis buffer (10 mM Tris-HCL, pH 8.0; 10% SDS; 100 mM NaCl; 10 mM EDTA) was then added and mixed in by shaking and tube inversion. The buffer was incubated at room temperature for 15 min, after which 1.5 ml of phenol:chloroform:isoamyl alcohol was added and mixed in by shaking and tube inversion. The tube was allowed to sit at room temperature for 15 min with periodic re-mixing and then the mixture was fully emulsified by shaking before adding equal amounts to two tubes of Phase Lock gel (5PRIME, Gaithersburg, MD). The tubes were centrifuged for 20 min at 21,000 g in a benchtop centrifuge and the supernatant was then decanted into two separate microfuge tubes. The DNA was precipitated by mixing in 0.54 volumes of room temperature isopropanol, followed by immediate centrifugation at 21,000 g for 5 min. The pellets were rinsed with 70% ethanol and then allowed to air dry. Pellets were re-dissolved in TE + 10 µg/ml RNAse (Qiagen Corp.).

Library Construction and Genome Sequencing. Sequence-ready libraries were prepared using the Nextera (Illumina Corp., San Diego, CA) and HyperPlus DNA kits (Roche Diagnostics, Indianapolis, IN). Nextera libraries were generated according to the manufacturer's protocol with the lone modification being an extension of the tagmentation reaction time to 60 min. HyperPlus libraries were generated precisely according to the manufacturer's protocol. Libraries were submitted to NovoGene for sequence acquisition (150 bp paired-end reads) on the HiSeq2500 platform.

Genome assembly. Raw sequence reads were quality filtered and adapters were trimmed using Trimmomatic 0.39 (Bolger et al. 2014) with the following options: ILLUMINACLIP:adapters.fa:2:30:10 SLIDINGWINDOW:20:20 MINLEN:90. Filtered reads were then assembled using velvet 1.2.10 (Zerbino and Birney, 2008) with the velvetoptimiser wrapper (https://github.com/tseemann/VelvetOptimiser) being used to iterate through kmer values of 89 to 129, with a step size of 2, to find the optimal assembly (default optimization parameters).

Genome masking and SNP calling. SNPs were called from BLAST alignments of masked genomes using iSNPcaller (https://github.com/drdna/iSNPcaller), according to two basic strategies: the first involved comparing genomes in all possible pairwise combinations with the goal of minimizing information loss due to extensive presence/absence polymorphism; the second involved alignments to the B71 reference genome (GCA_001675625.1) and was used to generate fasta alignments of variant sites. The basic algorithmic approaches, as implemented by iSNPcaller, involved: 1) masking repeats in all genomes; 2) performing pairwise alignments using BLASTn (-evalue 1e-20 -max_target_seqs 20000); 3) identifying SNPs that occur in uniquely aligned segments of both the reference and the query genome; 4) normalizing divergence by determining the total number of uniquely aligned nucleotide positions. Accuracy of the SNP-calling pipeline averaged >99.99% (Suppl. Fig. S2), as determined by calling variants between two assemblies of the same genome - one generated using the forward reads and another with reverse reads.

Phylogenetic analyses. Neighbor joining trees were built by importing pairwise distance data from iSNPcaller into MEGA X (Kumar et al. 2018) and using the default parameters for tree construction. Bootstrapping was performed by resampling the alignment data. For maximum likelihood analysis, fasta files were generated from the variant call data using a custom perl script. The data were filtered so that only those nucleotide positions called in every isolate were retained (i.e. no missing data). Tree building was performed using RAxML-NG-MPI (Kozlov et al. 2019)

with the GTR + Gamma substitution model to generate 10 starting random trees, 10 starting parsimony trees, and then 100 bootstrap replications. The best tree, as determined by RAxML, was plotted using the R package *ggtree* (Yu et al. 2017).

Population clustering

Isolates were clustered into discrete populations using the adegenet package (Jombart 2008) implemented in Poppr (Kamvar et al. 2014). First, we determined the most probable number of discrete populations by assessing Bayesian Information Criteria obtained after performing 50 iterations of k-means clustering, with values of k (# clusters) between 1 and 50. The most probable number of discrete populations was taken as the k value with the lowest mean BIC score. Discriminant analysis of principal components (DAPC) was then used to determine population memberships for each isolate using a very conservative k value of 10 and retaining 100 principal components.

Data and Code Availability. DNA sequences have been deposited at NCBI under BioProject PRJNA320483. Custom code used for data analyses is available at https://github.com/drdna/WheatBlast.

RESULTS

Phylogenetic analysis of *P. oryzae* from Brazilian grasses in the context of the global population. To confirm our suspicion that isolates purported to be members of endemic grass-infecting populations are actually members of the PoT and PoL1 lineages, we compared them with 185 isolates from a large, global collection of *P. oryzae* sampled from 25 different monocot genera. Using whole genome divergence data to build a neighbor-joining tree revealed that all but one of the previously-characterized, “suspect” isolates from Brazilian non-wheat hosts grouped in the PoT and PoL1 clades along with large numbers of bona-fide wheat blast and gray leaf spot pathogens. For clarity, these isolates are highlighted with bold typeface in Figure 1. The isolates from *Cenchrus*, *Eleusine*, *Elionurus* and *Melinis* grouped with PoT members, along with three of the four isolates from *Urochloa*. The three isolates from *Avena* grouped with PoL1, which also included individuals from *Echinochloa*, *Digitaria* and *Urochloa*. The one exceptional isolate, Ds363, had previously been assigned as a separate species, *P. grisea*, and yet it clearly grouped with the *P. oryzae Echinochloa* lineage (PoEc), which itself was firmly positioned within the *P. oryzae* tree (Fig. 1B).

Interestingly, many of the “suspect” isolates from non-wheat hosts can be considered clones of wheat blast isolates because the genome-wide nucleotide divergence relative to the latter were well within the range of SNP calling error rates (Suppl. Fig. S2), and because they shared identical chromosomal haplotypes (see Rahnama et al. 2021). Isolates Ub007i, Ub009i, and Ub012i from *Urochloa* are clones of wheat blast isolates Py221, PY6025 and PY6045; Ce642i and Ce535 from *Cenchrus*, Ei534i and EiJA22 (*Eleusine*), and Mr051i (*Melinis*) are clones of PY0925, WB37 and WB035i (all from wheat); and As073 (*Avena*) and Ds555i (*Digitaria*) are clones of Br58 - an isolate from *Avena* that belongs to the PoL1 lineage (Gladieux et al. 2018). Interestingly, isolates As321, As345 and As347 defined a second clonal lineage of *Avena* pathogens, yet neither oat-associated lineage contained isolates from *Lolium*, despite falling under the PoL1 umbrella.

Bootstrapping of the distance tree was uninformative because resampling of such high density data yielded 100% support for nearly all branches (data not shown). Nevertheless, we can be confident that the other suspects are truly PoT/PoL1 members because: i) the longest branch length between a suspect isolate and its nearest ingroup neighbor (1,635 for As321 vs. GG11) is more than three times smaller than the nearest distance to any outgroup isolate (5,180 for Elcan194 vs. U169 from *P. oryzae* lineage E1), and with such high SNP calling accuracy and coverage (>99.99% confidence across ~85-90% of each genome, Suppl. Fig. S2A), none of the suspects can possibly belong to other groups; ii) all suspects grouped comfortably within the population membership frameworks for PoT/PoL1 previously established using clustering methods (Gladieux et al. 2018); and iii) a fully comprehensive and highly detailed phylogenomic analysis revealed that all isolates within the PoT/PoL1 clades - including the suspect isolates - recently descended from a single fungal individual that belongs in the PoT1 clade (Rahnama et al 2021).

Nevertheless, we also conducted maximum likelihood analysis of a whole genome dataset based on 364,573 SNPs which produced a very similar tree topology with strong bootstrap support for the relevant clades (Fig. S3) and, thereby, confirmed complete phylogenetic resolution of the PoT/PoL1 clades. However, it should be noted that the nucleotide substitution models underlying maximum likelihood analyses are invalidated for the PoT/PoL1 population because most of the phylogenetic signal comes from the variable partitioning of standing variation and not ongoing nucleotide substitution (Rahnama et al, 2021). This, along with the presence of significant presence/absence polymorphism in some *P. oryzae* lineages, means that a distance-based measure of nucleotide divergence is a more appropriate metric to use for phylogenetic analyses of wheat blast/gray leaf spot.

Lastly, we analyzed population differentiation using discriminant analysis of principal components. Distributions of the Bayesian information criterion for k values from 1 to 50 suggested that the 199 isolates belonged to 42 discrete populations (Fig. S4). However, even when we selected a low value of k (10) to encourage population merging, the suspect isolates continued to group with the wheat blast/gray leaf spot pathogens, which in turn were fully resolved from all other outgroups (Fig. S5B). In fact, 19 of the 20 *P. oryzae* isolates that came from grasses growing in wheat fields - including one sampled in Minas Gerais and three from Paraguay - were PoT/PoL1 members. In striking contrast, none of the 21 isolates we sampled away from wheat fields grouped in the PoT/Pol1 clades.

Analysis of S. American *P. oryzae* from non-wheat hosts in Uruguay and Minas Gerais State. Because all but one of the Brazilian *P. oryzae* grouped phylogenetically with PoT/PoL1, while isolates from *Digitaria*, *Eleusine*, and *Urochloa* collected elsewhere around the world all grouped according to host-of-origin, this suggested that Ceresini and coworkers had failed to sample the true, endemic grass-infecting populations in that country. However, we also considered the possibility that South American *Pyricularia* populations are very different to those found in the rest of the world, with the landscape being completely dominated by PoT/PoL1. To distinguish between these scenarios, we examined a small number of isolates collected in S. America but from locations at varying distances from wheat fields. First, we used whole genome sequencing to analyze isolates from a number of non-wheat hosts in Uruguay where there have been no reports of wheat blast, despite being located just 750 km from Rio Grande do Sul where wheat blast occasionally occurs. Then we sampled isolates from locations close to and at distance from wheat growing areas in Minas Gerias. In total, twenty-six isolates were analyzed, including ones from Cynodon sp. (n=1), *Digitaria* sp. (n=1), *Echinochloa* sp. (n=1), *Eleusine indica* (n=6), *Lolium multiflorum* (n=4), *Luziola* sp. (n=2), *Oryza sativa* (n=3), *Setaria italica* (n=1), *Stenotaphrum secundatum* (n=2), and Urochloa (n=3) (Table S1).

Sixteen of the 18 isolates from Uruguay either grouped according to their host-of-origin or, in the case of the *Luziola* pathogens, they defined a new lineage (PoLu). The only exceptions were U237 from *Lolium multiflorum* that was recovered from a highly unusual fall-flowering plant and was placed firmly in the *P. oryzae Echinochloa* (PoEc) lineage (Figs. 1A and S3), and U167 from *Echinochloa* that was *P. grisea* (Fig. 1B). Critically, none of the Uruguayan isolates grouped with PoT/PoL1 (Figs. 1, S3 & S5C).

In Minas Gerais, fungal strains were obtained from *Cenchrus echinatus*, *Cynodon plechtostachyus*, *Digitaria* spp., *Eleusine indica*, *Hordeum vulgare, Melinis (Rhynchelytrum)*

repens, *Panicum maximum*, *Triticum aestivum* and *Urochloa* spp. (Table S1). Then, because our primary goal was to determine if the Brazilian grass-infecting *P. oryzae* are more in line with the global population than is suggested by prior research, we amplified and sequenced PCR-based markers (MPG1, CH7BAC7 and CH7BAC9) to pre-screen isolates from non-wheat hosts with the goal of identifying isolates that did not appear to be members of the PoT/PoL1 lineages and were, thus, likely to represent the endemic, grass-specialized lineages. Sequencing of CH7BAC7, CH7BAC9 and MPG1 PCR products for just 25 isolates revealed ten that had novel sequences and, thus, appeared to be from previously unsampled *P. oryzae* lineages, and four that were so divergent that they appeared to come from another *Pyricularia* species (data not shown). Genome sequences were obtained for nine isolates which variously came from *Cynodon* (n = 1), *Eleusine* (n=2), *Melinis* (n=2), *Pennisetum* (n = 2); and *Urochloa* (n = 2). For comparison purposes, we also performed genome sequencing on a single isolate collected from an *Eleusine* plant growing in a heavily diseased wheat plot in an experiment station.

Phylogenetic analyses of the whole genome SNP data showed that all but one of the isolates from Minas Gerais belonged to previously undefined lineages with strong bootstrap support (Figs. 1A & S3): the two *Eleusine* pathogens defined *P. oryzae Eleusine* lineage 3 (PoE3); CdJA159 from *Cynodon distachya* represented *P. oryzae Cynodon* lineage2 (PoC2), and MrJA49 and UbJA92 belonged to a *Melinis*-adapted lineage (PoM) (Figs. 1A and S3). Though the groups were phylogenetically distinct, the low k value used for DAPC analysis resulted in merging of PoU3 with PoSt into group 9; and PoC2, PoE3, and PoM into group 3, which also included the phylogenetic lineages PoU1, PoEc and PoEr (Fig. S5C).

The *Panicum maximum* pathogens, PmJA1 and PmJA115, showed far greater sequence divergence relative to the other isolates (~11% versus ~1%), consistent with their being members of a different species (Fig. 1B). Comparison with the NCBI database, revealed sequence identity to *Pyricularia urashimae* at a number of phylogenetic marker loci (data not shown).

The one exceptional isolate that we sampled from a wheat plot in Minas Gerais, EiJA22, grouped within the PoT lineage (Figs. 1, S3 and S5B) and was in the same haplotype group (PoT14) as - and therefore clonally related to - Ei534i, which also came from *Eleusine* growing near wheat. However, the latter isolate was collected two years earlier in Paraná state. Significantly, the only other PoT/PoL1 members that we found on grasses in S. America, were P28 and P29 from *Bromus*, and P25 from *Urochloa* (Figs. 1A, S3 and S5B), and these were all collected from grasses immediately adjacent to infected wheat.

DISCUSSION

Current understanding on the evolution, population biology and epidemiology of wheat blast disease is largely based on the studies of Ceresini and coworkers who made several key conclusions after examining the genetic relationships between *P. oryzae* found on Brazilian wheat and those found on neighboring grasses (Castroagudín et al. 2016; Ceresini et al. 2018, 2019). Throughout their work there was an implicit assumption that the isolates collected from non-wheat hosts represent the populations typically found infecting endemic grasses in Brazil. However, several inconsistencies with prior research, led us to suspect that this might not be true. Here, we report a re-analysis of their data in the context of a broader, community dataset which confirmed that all but one of the isolates used in their studies are members of the PoT or PoL1 lineages. Furthermore, by surveying grass-infecting isolates collected away from wheat production areas in Brazil, we confirmed a failure to sample the true, endemic, grass-infecting populations (see also Ascari et al. 2023). Finally, by analyzing genome sequences for select isolates, we identified a number of previously unknown *P. oryzae* lineages.

Clearly, limited sampling was a central factor in the previous oversight, as collecting diseased grasses away from wheat fields readily turned up isolates with the expected patterns of host specialization (Ascari et al. 2023). Another key mistake was to ignore key foundational research. Virtually all foregoing phylogenetic studies had shown that blast isolates from non-wheat hosts grouped according to host-of-origin (Borromeo et al. 1993; Couch and Kohn 2002; Farman 2002). So when isolates from multiple host species were found essentially to be clones of one another (e.g. Ce535i, Ei534i, Mr051i and WB037 from *Cenchrus*, *Eleusine*, *Melionis* and wheat, respectively) (Castroagudin et al. 2017; Ceresini et al. 2018, 2019) (Figs 1 and S2), this should automatically have thrown up warning flags because the sheer number of isolates that bucked the long-established trends of host specialization, by extension, would have implied that the blast populations infecting the grasses endemic to South America must behave completely differently to those found everywhere else.

Unfortunately, an immediate consequence of the aforementioned oversights is the invalidation of most findings arising from the work. The first major conclusion to be drawn was that the wheat blast population represents a separate species (Pygt) (Castroagudín et al. 2016). When the Ceresini data were integrated into the community dataset a very different picture emerged and revealed that the seminal studies didn't actually have the power to resolve species. This is nicely illustrated by the lone non-PoT/PoL isolate that was included in their study, Ds363 (a.k.a. 12.0.363), which was reported as having been collected from *Digitaria* spp. (Castroagudin

et al. 2017). This appears to have been assigned to *P. grisea,* i) based on precedent (isolates from *Digitaria* historically have been *P. grisea*); and ii) because Ds363 was phylogenetically resolved from Pygt and PoO, and on an adjacent branch to a known *P. grisea* isolate, Br29 (Castroagudin et al 2017). Here, we show that Ds363 clearly belongs to *P. oryzae* and is very distantly related to Br29 (*P. grisea*) (Fig. 1B). These conflicts illustrate the potential danger of trying to define species solely through empirical visualization of phylogenetic trees, and emphasizes the need to incorporate statistical approaches.

The fact that nearly all of the grass-infecting isolates included in prior studies are *bona fide* members of the PoT and PoL1 lineages has major implications for conclusions about wheat blast evolution, epidemiology and population biology. Previously, it was reported that "the closest relative of the wheat pathogen was found on the widely grown pasture grass *Urochloa*" (Stukenbrock and McDonald 2008) which led to the proposition that wheat blast evolved via a host jump from this grass; and that *Urochloa* probably plays a significant role in inoculum survival, production and epidemic spread. First, it is not clear that the similarity was accurately reported because the cited data (https://github.com/crolllab/wheat-blast) show that isolates from *Cenchrus* and *Eleusine* are just as closely related to wheat blast pathogens; and the present study shows that all of these isolates, as well as others from *Elionurus* and *Melonis*, are all clones of wheat blast isolates found in their collection. So, the reason that the wheat pathogen is most closely related to isolates from *Urochloa* (and *Cenchrus*, *Eleusine*, *Elionurus* and *Melinis*), is because all of these isolates are, in fact, wheat blast pathogens. Even if we take the reported similarity at face value, the fact that the endemic *Urochloa*-infecting population was never sampled should have precluded the drawing of any conclusions regarding its possible contribution to wheat blast evolution via host jumps.

Likewise, without any information on the structure of the endemic *P. oryzae* population(s) on *Urochloa* and, especially, the relative prevalence of PoT, it seems premature to make any inferences about *Urochloa*'s role in wheat blast epidemiology. However, a recent study, based on whole genome analyses of a comprehensive collection of PoT/Pol1 isolates, shows that both lineages can be sub-divided into discrete haplotypes, based on distinct chromosomal configurations that arose when wheat blast/gray leaf spot first evolved via a process that involved recombination of divergent *P. oryzae* genomes in a multi-hybrid swarm (Rahnama et al. 2021). To date, 44 discrete haplotypes have been identified for PoT (Rahnama et al, 2021; Ascari et al. 2023), with only one having been found thus far on *Urochloa* (PoT-14). Therefore, at present, it would appear that wheat blast is more a source of inoculum for infection of *Urochloa,* than vice versa.

Lastly, Ceresini and coworkers argued that wheat blast is not a host-specialized pathogen and, therefore, "the hypothesis of grass-specific populations for the overall *P. oryzae* species complex is falsified" (Ceresini et al. 2019), so that "*P. oryzae* may not provide a suitable model for understanding the biology of *Pygt*." Our data show that, to the contrary, wheat blast retains a high degree of host-specialization because, to date, only three out of 44 distinct PoT haplotypes have been found infecting more than one host. This is perfectly in line with what has been observed for *P. oryzae* as a whole, because while most *P. oryzae* lineages show strong host-specialization with very few exceptions, a small number of lineages show distinct non-specificity and are represented by isolates from three or more host genera - good examples are the PoEc lineage, whose members come from *Echinochloa*, *Leptochloa*, *Paspalum* and *Zea*; and PoSt comprising isolates from *Stenotaphrum*, *Hakonechloa*, *Triticum*, and *Urochloa* (Figure S3; Ascari et al. in preparation). Therefore, *P. oryzae* seems to be a perfect model for understanding wheat blast biology as it relates to host specialization.

ACKNOWLEDGMENTS

This work was supported by the United States Department of Agriculture, Agriculture and Food Research Initiative grant 2013-68004-20378, multistate project NE1602; Agricultural Research Service project 8044-22000-046-00D; Hatch project KY012037; the National Science Foundation, MCB-1716491; and the University of Kentucky College of Agriculture Food and the Environment. Emerson M. Del Ponte was supported by the National Council for Scientific and Technological Development (CNPq) through a Productivity Research Fellowship (PQ) project 310208/2019-0, and through research grants provided by FAPEMIG. João P. Ascari was supported by CNPq through a doctoral scholarship.

LITERATURE CITED

Borromeo, E. S., Nelson, R. J., Bonman, J. M., and Leung, H. 1993. Genetic differentiation among isolates of *Pyricularia* infecting rice and weed hosts. Phytopathology. 83:393–399.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30: 2114-2120.

Bruno, A. C., and Urashima, A. S. 2001. Inter-relação sexual de *Magnaporthe grisea* do trigo e de outros hospedeiros. Fitopatol. Bras. 26:21–26.

Castroagudin, V., Danelli, A., Intra Moreira, S., Teodora de Assis Reges, J., Carvalho, G., Maciel, J. L. N., Bonato, A. L. V., Forcelini, C. A., Alves, E., McDonald, B. A., Croll, D., Ceresini, P. C. 2017. The wheat blast pathogen *Pyricularia graminis-tritici* has complex

origins and a disease cycle spanning multiple grass hosts. bioRxiv. Available at: http://biorxiv.org/lookup/doi/10.1101/203455 [Accessed February 13, 2019].

Castroagudín, V. L., Moreira, S. I., Pereira, D. A. S., Moreira, S. S., Brunner, P. C., Maciel, J. L. N., Crous, P. W., McDonald, B. A., Alves, E., Ceresini, P. C. 2016. *Pyricularia graminis-tritici*, a new *Pyricularia* species causing wheat blast. Persoonia - Mol. Phylogeny Evol. Fungi. 37:199–216.

Ceresini, P. C., Castroagudín, V. L., Rodrigues, F. Á., Rios, J. A., Aucique-Pérez, C. E., Moreira, S. I., Croll, D., Alves, E., Carvalho, G., Maciel, J. L., McDonald, B. A. 2019. Wheat blast: from its origins in South America to its emergence as a global threat: Wheat Blast. Mol. Plant Pathol. 20:155–172.

Ceresini, P. C., Castroagudín, V. L., Rodrigues, F. Á., Rios, J. A., Eduardo Aucique-Pérez, C., Moreira, S. I., Alves, E., Croll, D., Maciel, J. L. 2018. Wheat blast: Past, present, and future. Annu. Rev. Phytopathol. 56:427–456.

Couch, B. C., and Kohn, L. M. 2002. A multilocus gene genealogy concordant with host preference indicates segregation of a new species, *Magnaporthe oryzae*, from *M. grisea*. Mycologia. 94:683–693.

Cruz, C. D., and Valent, B. 2017. Wheat blast disease: danger on the move. Trop. Plant Pathol. 42:210–222.

Farman, M. L. 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.

Galbieri, R., and Urashima, A. S. 2008. Caracterização, compatibilidade e ocorrência de reprodução sexual entre isolados de *Pyricularia grisea* de diferentes hospedeiros. Summa Phytopathol. 34:22–28.

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

Igarashi, S., Utiamada, C. M., Igarashi, L. C., Kazuma, A. H., Lopes, R. S. 1986. Occurrence of *Pyricularia* sp. in wheat (*Triticum aestivum* L.) in the State of Parana, Brazil. Fitopatol. Bras. 11:351–352.

Jombart, T. 2008 adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24:1403–1405.

Kamvar, Z. N., Tabima, J. F. and Grunwald, N. J. 2014 Poppr: an R package for genetic

analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ, 2*, e281. doi:10.7717/peerj.281.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453-4455.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evo. 35(6), p.1547.

Maciel, J. L. N., Ceresini, P. C., Castroagudin, V. L., Zala, M., Kema, G. H. J., and McDonald, B. A. 2014. Population structure and pathotype diversity of the wheat blast pathogen *Magnaporthe oryzae* 25 years after its emergence in Brazil. Phytopathology. 104:95–107.

Malaker, P. K., Barma, N. C. D., Tiwari, T. P., Collis, W. J., Duveiller, E., Singh, P. K., Joshi, A. K., Singh, R. P., Braun, H. J., Peterson, G. L., Pedley, K. F., Farman, M. L., Valent, B. 2016. First Report of wheat blast caused by *Magnaporthe oryzae* pathotype Triticum in Bangladesh. Plant Dis. 100:2330.

Milgroom, M. G. 2017. *Population Biology of Plant Pathogens: Genetics, Ecology, and Evolution*. The American Phytopathological Society.

Rahnama, M., Condon, B., Ascari, J. P., Dupuis, J. R., Del Ponte, E. M, Pedley, K. F., Martinez, S., Valent, B., Farman, M. L. 2021. Recombination of standing variation in a multi-hybrid swarm drove adaptive radiation in a fungal pathogen and gave rise to two pandemic plant diseases. BiorXiv :2021.11.24.DOI: 10.1101/2021.11.24.469688v1

Singh, P. K., Gahtyari, N. C., Roy, C., Roy, K. K., He, X., Tembo, B., Xu., K., Juliana, P., Sonder, K., Kabir, M. R., Chawade, A. 2021. Wheat blast: A disease spreading by intercontinental jumps and its management strategies. Front. Plant Sci. 12:1–21.

South, A. 2021. rnaturalearth: World Map Data from Natural Earth. Available at: https://github.com/ropensci/rnaturalearth.

South, A. 2022. rnaturalearthhires: High Resolution World Vector Map Data from Natural Earth used in rnaturalearth. Available at: https://github.com/ropensci/rnaturalearthhires.

Stukenbrock, E. H., and McDonald, B. A. 2008. The origins of plant pathogens in agro-ecosystems. Annu. Rev. Phytopathol. 46:75–100.

Tembo, B., Mulenga, R. M., Sichilima, S., M'siska, K. K., Mwale, M., Chikoti, P. C., Singh, P. K., he, X., Pedley, K. F., Peterson, G. L., Singh, R. P., Braun, H. J. 2020. Detection and characterization of fungus (*Magnaporthe oryzae* pathotype Triticum) causing wheat blast disease on rain-fed grown wheat (*Triticum aestivum* L.) in Zambia. PLOS ONE. 15:1–10.

Urashima, A. S., Igarashi, S., and Kato, H. 1993. Host range, mating type, and fertility of

Pyricularia grisea from wheat in Brazil. Plant Dis. 77:1211–1216.

Valent, B., Crawford, M. S., Weaver, C. G., and Chumley, F. G. 1984. Genetic studies of fertility and pathogenicity in *Magnaporthe grisea* (*Pyricularia oryzae*). United States: N. p.

Valent, B., Farman, M., Tosa, Y., Begerow, D., Fournier, E., Gladieux, P., Islam, M. F., Kamoun, S., Kemler, M. Kohn, L. M., Lebrun, M. H., Stajich, J. E., Talbot, N. J., Terauchi, R. Tharreau, D., Zhang, N. 2019. *Pyricularia graminis-tritici* is not the correct species name for the wheat blast fungus: response to Ceresini et al. (MPP 20:2). Mol. Plant Pathol. 20:173–179.

Yu, G. 2021. scatterpie: Scatter Pie Plot. R package version 0.1.7. Available at: https://CRAN.R-project.org/package=scatterpie.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8:28–36.

Zerbino, D.R. and Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-829.

FIGURE LEGENDS

Figure 1. Distance trees showing the phylogenetic grouping of grass-infecting isolates from Brazil and Uruguay. A) Tree includes only *P. oryzae* isolates; B) Tree includes the four species *P. oryzae*, *P. grisea*, *P. pennisetigena* and *P. urashimae*. Only the tips for isolates relevant to the present study are labeled. Names of "suspect" isolates previously sampled from grasses in Brazil are in bold typeface on a white background. Previously characterized PoT/PoL1 members that are clones of "suspect" isolates are highlighted with asterisks. Isolates collected in the present study are on a gray background and are identified by "U" followed by isolate number (Uruguay) or with a host species::collector (e.g. UbJA) abbreviation, followed by a number (Brazil). Black borders are drawn around isolates that did not group according to the host-of-origin. Groups connected to the tree with a single line are all clonally related based on limited, genome-wide divergence. The names of relevant phylogenetic clades are shown in bold (Hosts: PoC2 = *Cynodon*; PoE1 = *Eleusine*; PoE3 = *Eleusine*; Ec = *Echinochloa*; PoL1 = *Lolium*; PoM = *Melinis*; PoO = *Oryza*; PoS = *Setaria*; PoSt = *Stenotaphrum;* PoT = *Triticum*. Scales show phylogenetic distance measured in substitutions per site. All labeled clades exhibited 100% bootstrap support after resampling alignments and >90% bootstrap support with maximum likelihood analysis (Suppl. Fig. S3).

e-Xtras

TABLE

Supplemental Table 1. Fungal isolates used in this study

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Map showing the locations where the S. American grass-infecting isolates were collected. Each sampled location is represented by a pie chart depicting the proportion of isolates from each host genus. Numbers next to each chart show the total number of isolates sampled for that location and the border style indicates the distance to the nearest wheat field (<1 km = solid line; 1-10 km = long dash; > 10 km = dotted line). Note that the isolates collected by J. Ascari and S. Martinez mostly came from regions distant from wheat fields, although the Uruguayan isolates came from regions of high *Lolium* production (and gray leaf spot incidence).

Figure S2. SNP calling error rates compared with nucleotide divergence between clonal isolates. SNP divergence was measured by calling variants from whole genome alignments and scaling by total alignment length (after filtering out repeats). Plotted values represent: SNP calling error rates measured as nucleotide differences between separate assemblies of the same genome (n = 72); and pairwise nucleotide divergence between each suspected clone and the wheat blast/gray leaf spot isolate showing the closest similarity.

Figure S3. Phylogenetic relationships among *P. oryzae* isolates as revealed by maximum likelihood analysis of whole genome SNP data (362,258 variant sites). Isolate names are colored according to the host of origin and names of the main phylogenetic clades are shown on the right (PoC1/2, *Cynodon1/2*; PoEc, *Echinocloa*, PoE1/2/3, *Eleusine1/2/3*; PoL1/2/3, *Lolium1/2/3*; PoLe, *Leersia*; PoM, *Melinis*; PoP, *Panicum*; PoO, *Oryza*; PoS, *Setaria*; PoSt, *Stenotaphrum*; PoT, *Triticum*, PoU1/2/3, *Urochloa*1/2/3). Host genera for which there were no distinct phylogenetic clades include *Avena* (A) and *Hordeum* (Ho) - representative isolates grouped in PoL1 and PoO, respectively. Suspect isolates from previous studies are highlighted with blue circles, while those from the present study are marked with red triangles. Bootstrap values (100 repetitions) >90 are given for the nodes that defined main clades.

Figure S4. Population division using discriminant analysis of principal components. Distribution of Bayesian information criterion at each k value for 50 independent runs of k-means clustering (k = 1 to 50, 1E6 iterations/run).

Figure S5. Population memberships of isolates included in this study. Isolates were assigned to discrete populations using discriminant analysis of principal components using a k value of 10. Population memberships are shown for: A) all isolates; B) Isolates sampled from grasses in wheat fields; C) Isolates sampled away from wheat fields. In panels B and C, bars are given a “group” color if the isolate in question groups with a majority of the other isolates from the same host. Gray bars are used for isolates that group with non-canonical hosts (n.c.h.).

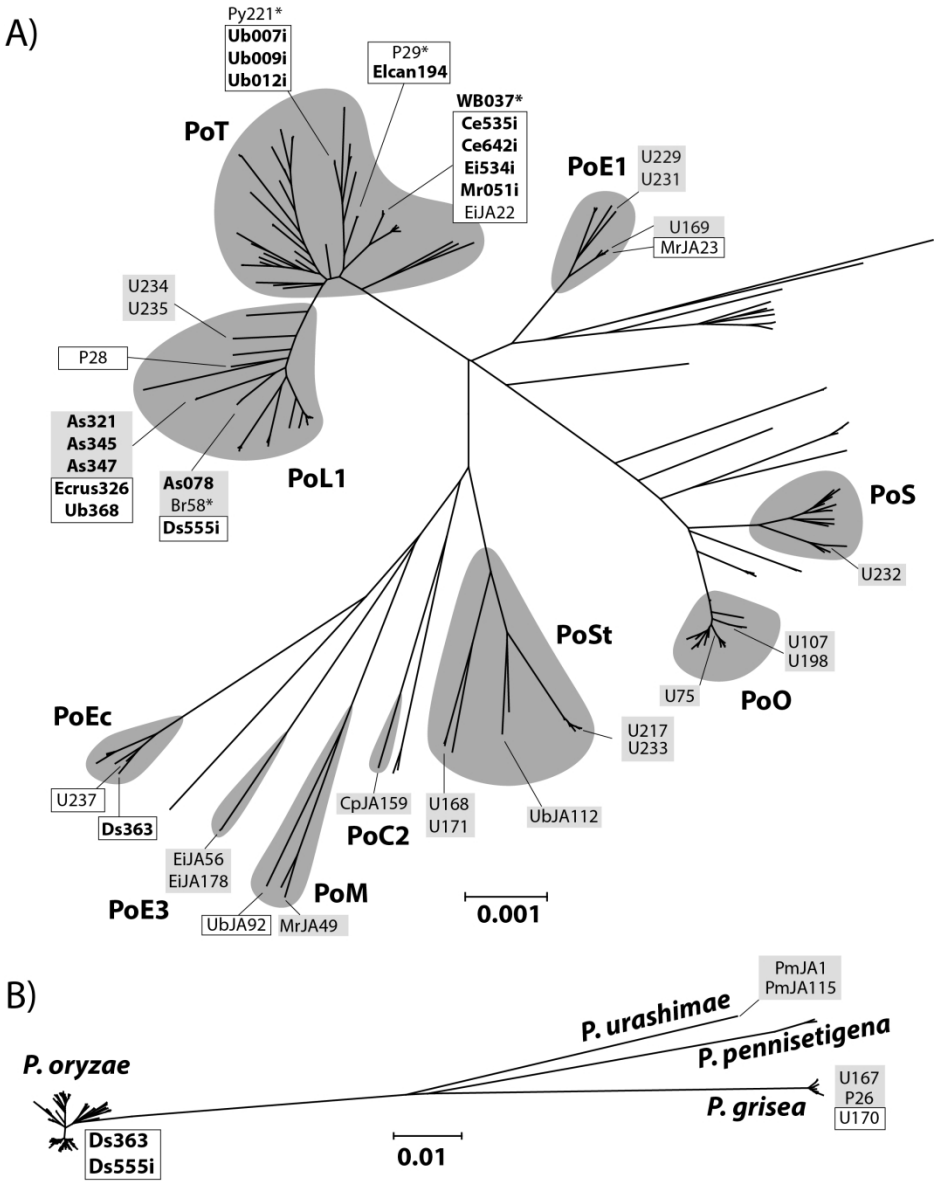

Figure 1

200x257mm (300 x 300 DPI)

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Map showing the locations where the S. American grass-infecting isolates were collected. Each sampled location is represented by a pie chart depicting the proportion of isolates from each host genus. Numbers next to each chart show the total number of isolates sampled for that location and the border style indicates the distance to the nearest wheat field (<1 km = solid line; 1-10 km = long dash; > 10 km = dotted line). Note that the isolates collected by J. Ascari and S. Martinez mostly came from regions distant from wheat fields, although the Uruguayan isolates came from regions of high *Lolium* production (and gray leaf spot incidence).

Figure S2. SNP calling error rates compared with nucleotide divergence between clonal isolates. SNP divergence was measured by calling variants from whole genome alignments and scaling by total alignment length (after filtering out repeats). Plotted values represent: SNP calling error rates measured as nucleotide differences between separate assemblies of the same genome (n = 72); and pairwise nucleotide divergence between each suspected clone and the wheat blast/gray leaf spot isolate showing the closest similarity.

Figure S3. Phylogenetic relationships among *P. oryzae* isolates as revealed by maximum likelihood analysis of whole genome SNP data (362,258 variant sites). Isolate names are colored according to the host of origin and names of the main phylogenetic clades are shown on the right (PoC1/2, *Cynodon1/2*; PoEc, *Echinocloa*, PoE1/2/3, *Eleusine1/2/3*; PoL1/2/3, *Lolium1/2/3*; PoLe, *Leersia*; PoM, *Melinis*; PoP, *Panicum*; PoO, *Oryza*; PoS, *Setaria*; PoSt, *Stenotaphrum*; PoT, *Triticum*, PoU1/2/3, *Urochloa*1/2/3). Host genera for which there were no distinct phylogenetic clades include *Avena* (A) and *Hordeum* (Ho) - representative isolates grouped in PoL1 and PoO, respectively. Suspect isolates from previous studies are highlighted with blue circles, while those from the present study are marked with red triangles. Bootstrap values (100 repetitions) >90 are given for the nodes that defined main clades.

Figure S4. Population division using discriminant analysis of principal components. Distribution of Bayesian information criterion at each k value for 50 independent runs of k-means clustering (k = 1 to 50, 1E6 iterations/run).

Figure S5. Population memberships of isolates included in this study. Isolates were assigned to discrete populations using discriminant analysis of principal components using a k value of 10. Population memberships are shown for: A) all isolates; B) Isolates sampled from grasses in wheat fields; C) Isolates sampled away from wheat fields. In panels B and C, bars are

given a “group” color if the isolate in question groups with a majority of the other isolates from the same host. Gray bars are used for isolates that group with non-canonical hosts (n.c.h.).

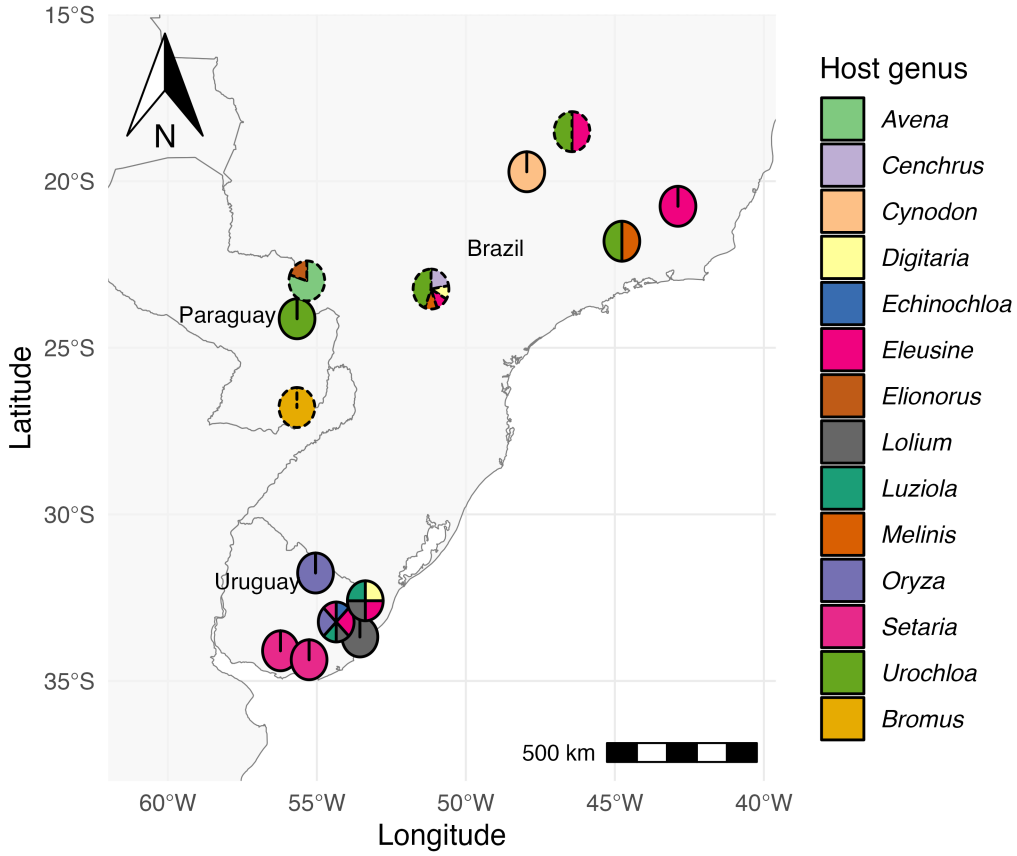
15°S
20°S
25°S
30°S
35°S
Latitude
60°W
55°W
50°W
45°W
40°W
Longitude
N
Brazil
Paraguay
Uruguay
500 km
Host genus
Avena
Cenchrus
Cynodon
Digitaria
Echinochloa
Eleusine
Elionorus
Lolium
Luziola
Melinis
Oryza
Setaria
Urochloa
Bromus

Figure S1

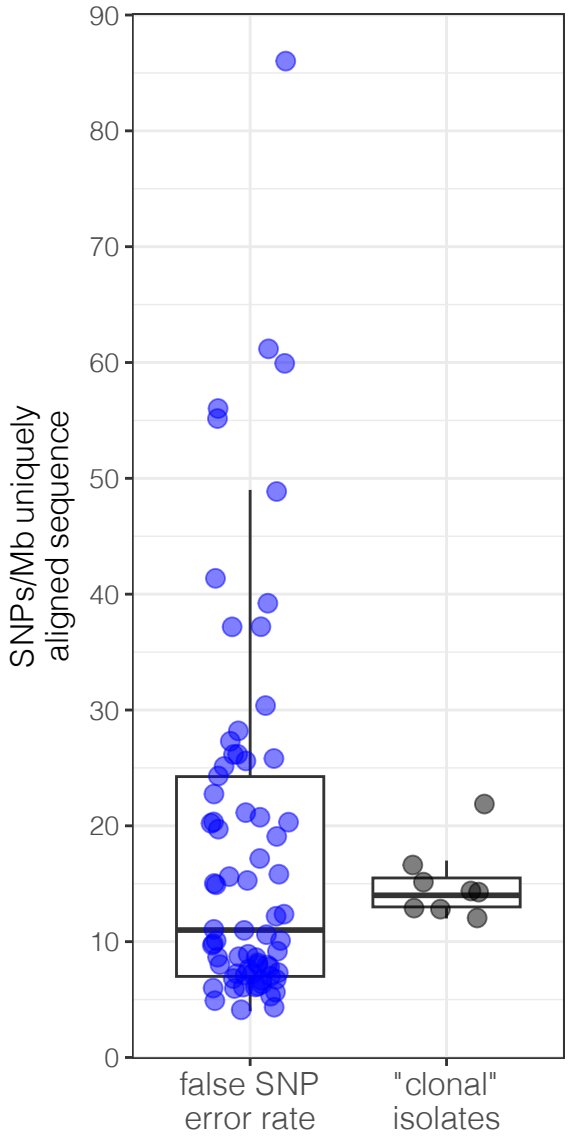
SNPs/Mb uniquely aligned sequence
90
80
70
60
50
40
30
20
10
0
false SNP error rate
"clonal" isolates

Figure S2

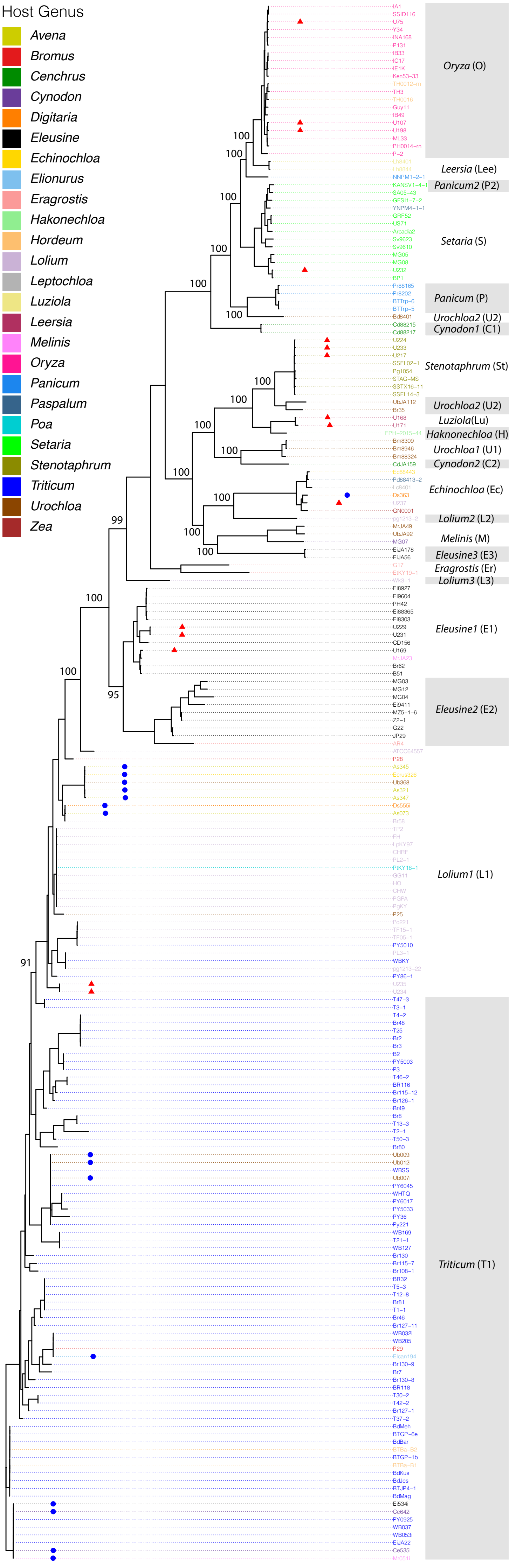
Host Genus
Avena
Bromus
Cenchrus
Cynodon
Digitaria
Eleusine
Echinochloa
Elionurus
Eragrostis
Hakonechloa
Hordeum
Lolium
Leptochloa
Luziola
Leersia
Melinis
Oryza
Panicum
Paspalum
Poa
Setaria
Stenotaphrum
Triticum
Urochloa
Zea
Oryza (O)
Leersia (Lee)
Panicum2 (P2)
Setaria (S)
Panicum (P)
Urochloa2 (U2)
Cynodon1 (C1)
Stenotaphrum (St)
Urochloa2 (U2)
Luziola(Lu)
Haknonechloa (H)
Urochloa1 (U1)
Cynodon2 (C2)
Echinochloa (Ec)
Lolium2 (L2)
Melinis (M)
Eleusine3 (E3)
Eragrostis (Er)
Lolium3 (L3)
Eleusine1 (E1)
Eleusine2 (E2)
Lolium1 (L1)
Triticum (T1)
0.7

Figure S3

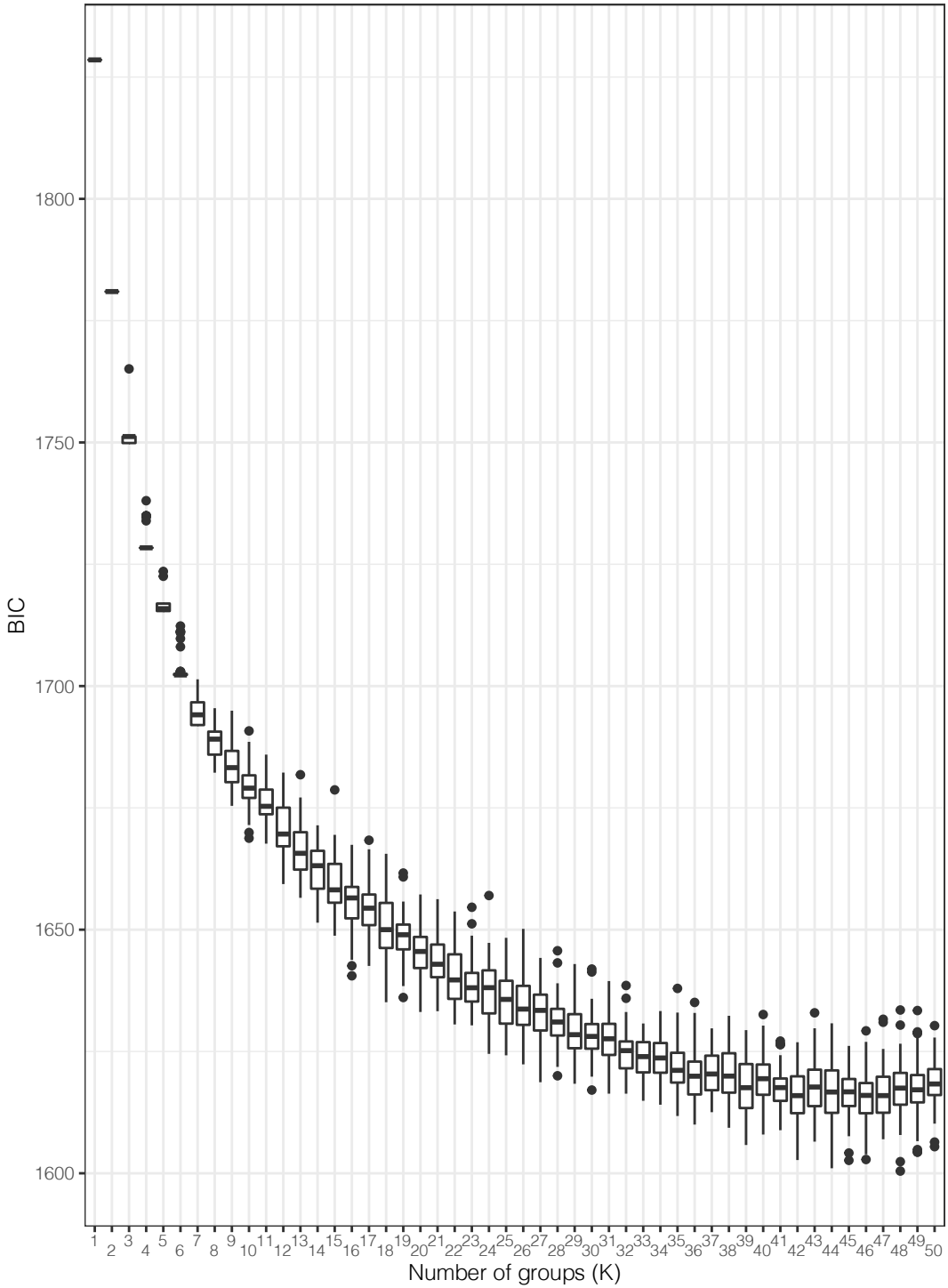
BIC
1800
1750
1700
1650
1600
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
Number of groups (K)

Figure S4

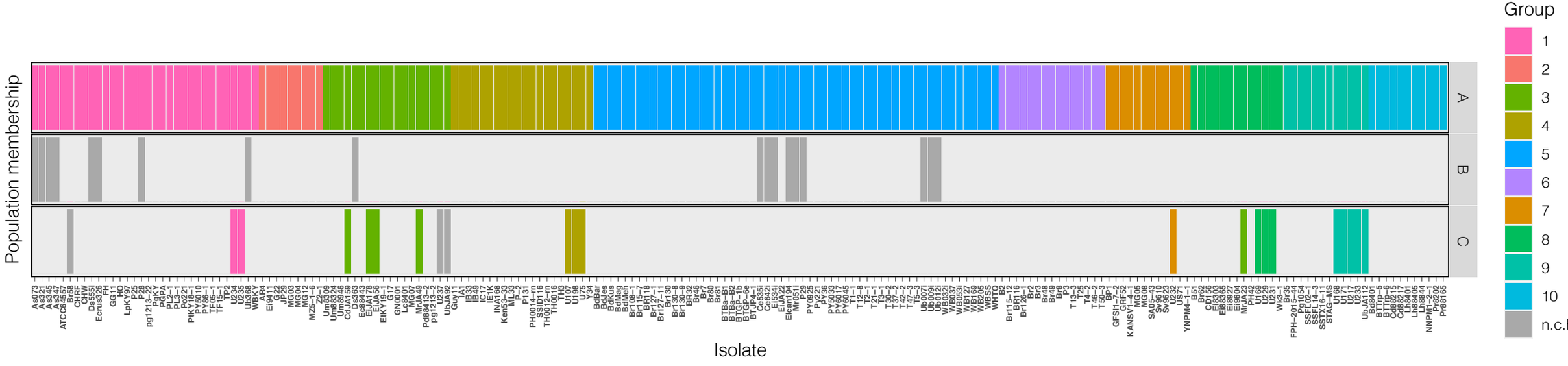

Figure S5