The emergence of wheat blast in Zambia and Bangladesh was caused by the same genetic lineage of *Magnaporthe oryzae*

Sergio M. Latorre[1], Hernán A. Burbano[1]

[1]Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, London, United Kingdom.

Abstract

We analyzed genotyping data (84 SNPs) derived from blast fungus (*Magnaporthe oryzae*) isolates (N=186) from 13 different grasses in order to establish the genetic makeup and the phylogenetic placement of the wheat-infecting lineage that caused a wheat blast outbreak in Zambia (2018-2020). The wheat-infecting isolates from Zambia (2018-2020) clustered together with the wheat-infecting isolates that caused the initial outbreak in Bangladesh (2016) and with most of the other wheat-infecting isolates from Bangladesh (2016-2020). Therefore, we concluded that the same lineage responsible for the initial wheat blast outbreak in Bangladesh (2016) caused the wheat blast outbreak in Zambia (2018-2020). We identified one wheat-infecting isolate from Bangladesh (2018) - BTM-1831-5 - that did not group together with the main Bangladesh-Zambian lineage, but instead clustered with rice-infecting isolates. Manual curation of the samples in Bangladesh revealed that BTM-1831-5 is in fact a rice-infecting isolate mislabelled as wheat-infecting.

The genotyping of 84 SNPs in multiple grasses permitted the identification of the epidemic lineage responsible for the outbreaks in both Bangladesh and Zambia. The reconstruction of the detailed population history of this important epidemic lineage will require genome-wide genotyping, which will be distributed through the OpenWheatBlast initiative.

Introduction

The fungus *Magnaporthe* (Syn. *Pyricularia*) *oryzae*, the causal agent of the blast disease of cereals, includes a group of genetically differentiated lineages that are frequently associated with specific host genera [1, 2]. The wheat-infecting lineage of the blast fungus, known as pathotype *Triticum* (MoT) originated in South America [3], where it jumped host to wheat [4]. MoT spread for the first time from South America to Bangladesh in 2016 [5] and, more recently, a wheat blast outbreak has been reported in Zambia (2017-2018) [6]. While it has been ascertained that the wheat outbreak in Bangladesh was caused by a nearly-clonal lineage from South-America [5], the phylogenetic placement of the wheat-infecting lineage identified in Zambia [6] and the genetic relationships between the outbreaks in Bangladesh and Zambia are not known. The dissemination of wheat blast to multiple continents increases the possibility that these outbreaks would act as sources of MoT to breadbasket regions, making MoT a threat for global wheat production and food security. Describing the genetic relationships among plant pathogen populations is fundamental to understanding the mechanism and timescales of emerging and re-emerging outbreaks. Therefore, we used genotyping data (84 SNPs) for a set of 186 *M. oryzae* isolates from wheat and 12 additional grasses [7] (made available through the OpenWheatBlast initiative) to determine the relationships between the epidemic lineages from Bangladesh and Zambia.

Results and Discussion

Newly-genotyped blast isolates from Zambia (wheat-infecting) and Bangladesh (wheat- and rice-infecting) cluster with other wheat- and rice-infecting lineages

To ascertain the genetic makeup and the phylogenetic placement of the wheat-infecting isolates from Zambia in an unbiased way, we initially analyzed all 13 species of grasses together. We used UPGMA, a distance-based agglomerative clustering approach, to generate a dendrogram with all samples. The published set of 84 SNPs was sufficient to identify multiple clades comprising host-specific infecting blast fungus isolates. However, host-specific clades did not necessarily form monophyletic groups (Supp. Fig 1). We found that all of the newly-genotyped isolates from Zambia (wheat-infecting) and Bangladesh (wheat- and rice-infecting) clustered with other wheat- and rice-infecting lineages. Since none of the other 11 genotyped grass species clustered with isolates from Bangladesh and Zambia, for subsequent analysis we only used wheat- and rice-infecting isolates.

The same lineage responsible for the initial wheat blast outbreak in Bangladesh (2016) caused the wheat blast outbreak in Zambia (2018-2020)

We established the relationships between all wheat-infecting (N=106) and rice-infecting (N=39) isolates using three different approaches. First, we used the same clustering approach described in the previous section (UPGMA). This analysis revealed that all of the 13 wheat-infecting isolates from Zambia clustered together with the wheat-infecting Bangladesh isolates from the initial outbreak of 2016 and with all but one of the subsequently collected wheat-infecting isolates in Bangladesh (2016-2020) (Fig. 1A). Second, we carried out a phylogenetic reconstruction using maximum parsimony, which separated rice-infecting isolates from wheat-infecting ones from Zambia and Bangladesh. The obtained tree also shows the greater diversity of wheat-infecting isolates from Brazil in contrast with the more homogeneous wheat-infecting samples from Zambia and Bangladesh (Fig. 1B). Third, we performed a Principal Component Analysis (PCA)-based ancestry deconvolution approach. To this purpose, we carried out a PCA using a set of isolates with known phylogenetic placement. This set includes a worldwide group of 31 rice-infecting isolates and 23 isolates originally characterized during the first wheat blast outbreak in Bangladesh (2016) [5]. Subsequently, we projected into this PCA space (PC1 and PC2) the test samples that comprise 13 isolates from Zambia and 53 newly-genotyped wheat-infecting isolates from Bangladesh. This analysis showed that all but one of the test samples clustered together with the lineage that caused the initial outbreak in Bangladesh (2016) (Fig 1C). Our clustering and phylogenetic analysis identifies the Brazilian isolate PY0925 as the closest isolate to the Bangladesh-Zambian lineage. This goes in line with the initial report of the emergence of wheat blast in Bangladesh (2016), which suggested a South American origin for the wheat blast outbreak in Bangladesh [5].

We concluded based on the analysis of the 84 SNPs that the same lineage responsible for the initial wheat blast outbreak in Bangladesh (2016) caused the wheat blast outbreak in Zambia (2018-2020).
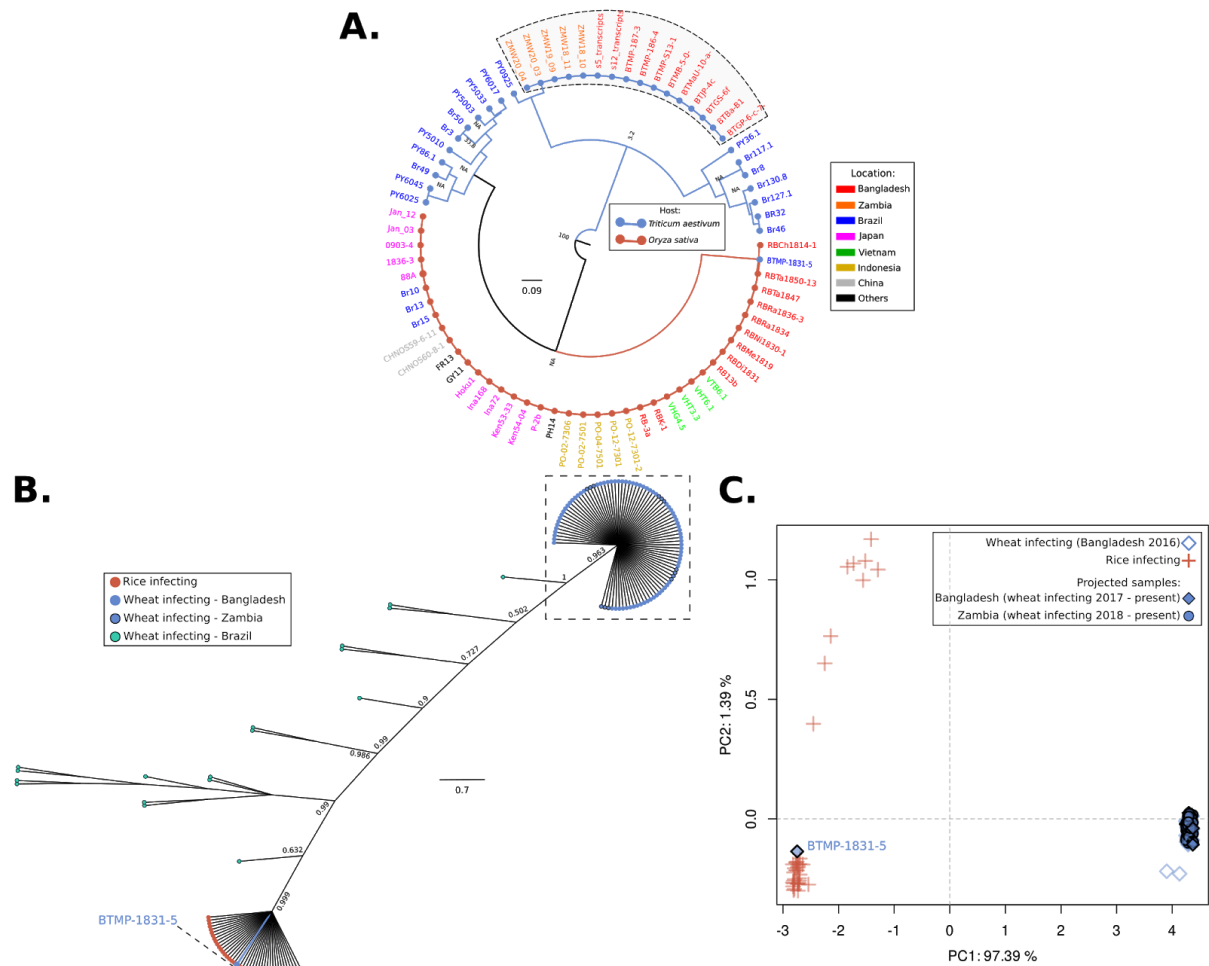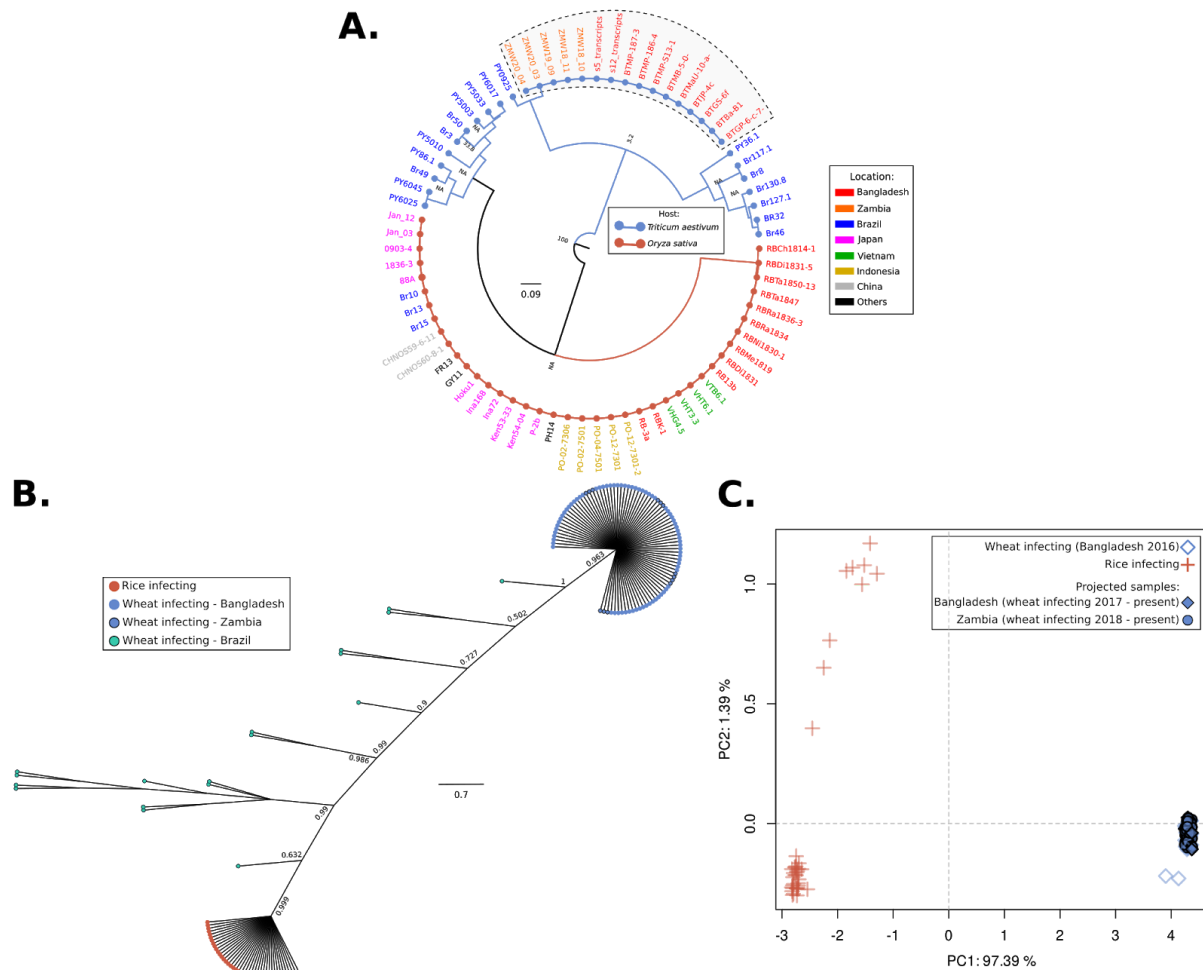
Figure 1. The emergence of wheat blast in Zambia and Bangladesh was caused by the same genetic lineage of *Magnaporthe oryzae*. All panels include only rice-infecting and wheat-infecting isolates. **A.** UPGMA-based dendrogram including 39 rice-infecting (blue tips) and 35 wheat-infecting (red tips) isolates. The support of the nodes is based on 1000 bootstrap replicates. The isolate labels are color coded according to country of origin (see inset). For clarity only a subset of the wheat-infecting isolates from Bangladesh were used in the analysis (all isolates are displayed in Supplementary Figure 2). The genetic lineage that caused the wheat blast outbreaks in Bangladesh and Zambia is enclosed within the dotted box. **B.** Maximum parsimony tree with node support based on 1000 bootstrap replicates. The analysis included the same isolates as in A. The circles at the tips are color coded according to host and country of origin (see inset). The scale bar represents the distance measured as the number of substitutions. **C.** PCA space of a worldwide set of rice-infecting isolates and wheat-infecting isolates from the outbreak in Bangladesh (2016), with wheat-infecting isolates from Zambia (2018-2020) and Bangladesh (2017-2020) projected into the PC space. The inset indicates the host and whether the isolate was used to build the PC space or was projected into it. The filled blue diamond corresponding to isolate BTMP-1831-5 is named in the PC space.

A wheat-infecting isolate from Bangladesh (2018) is more related to rice-infecting isolates than to wheat-infecting ones

The clustering, phylogenetic and PCA-based ancestry deconvolution analyses presented in the previous section consistently cluster BTMP-1831-5, a wheat-infecting isolate collected in Bangladesh in 2018, with rice-infecting isolates (Fig 1A-C). We noticed that BTMP-1831-5 is an outlier in the distribution of missing data per isolate (Supp. Fig 3A). Therefore, we repeat the clustering analysis allowing 5% of missing data or using only sites with full information. Even though the use of full information reduced the number of sites to 26, we found that the

BTMP-1831-5 isolate consistently clustered with rice-infecting isolates in the two analyses (Supp. Fig. 3B-C). Manual curation of the samples in Bangladesh revealed that BTM-1831-5 is in fact a rice-infecting isolate mislabelled as wheat-infecting (Nur Uddin Mahmud, personal communication). The identification of a mislabelled isolate shows the robustness of the set of 84 SNPs to identify host-specific blast isolates. Accordingly, we have re-labelled BTM-1835-5 as rice infecting RBDi1831-5 (Fig. 2).

Figure 2. The emergence of wheat blast in Zambia and Bangladesh was caused by the same genetic lineage of _Magnaporthe oryzae_. All panels include only rice-infecting and wheat-infecting isolates. The isolate BTMP-1831-5 has been relabelled as RBDi1831-5 (see main text). **A.** UPGMA-based dendrogram including 39 rice-infecting (blue tips) and 35 wheat-infecting (red tips) isolates. The support of the nodes is based on 1000 bootstrap replicates. The isolate labels are color coded according to country of origin (see inset). For clarity only a subset of the wheat-infecting isolates from Bangladesh were used in the analysis (all isolates are displayed in Supplementary Figure 2). The genetic lineage that caused the wheat blast outbreaks in Bangladesh and Zambia is enclosed within the dotted box. **B.** Maximum parsimony tree with node support based on 1000 bootstrap replicates. The analysis included the same isolates as in A. The circles at the tips are color coded according to host and country of origin (see inset). The scale bar represents the distance measured as the number of substitutions. **C.** PCA space of a worldwide set of rice-infecting isolates and wheat-infecting isolates from the outbreak in Bangladesh (2016), with wheat-infecting isolates from Zambia (2018-2020) and Bangladesh (2017-2020) projected into the PC space. The inset indicates the host and whether the isolate was used to build the PC space or was projected into it.

Conclusions

Surveillance of wheat blast is fundamental to design control strategies and prevent its further spread. Our analyses revealed that the emergence of wheat blast in Zambia and Bangladesh was caused by the same genetic lineage of *Magnaporthe oryzae*, a lineage closely related to South American isolates (Fig 2). Genome-wide Illumina sequencing of the Zambian isolates that will be available soon through the OpenWheatBlast initiative will help in the reconstruction of the population history of this important wheat-blast epidemic lineage.

Acknowledgements

Materials and Methods

Clustering and Phylogenetic Analysis

For each isolate we concatenated the genotypes calls for the 84 SNPs made available through the OpenWheatBlast Initiative [7] and generated *fasta*-formatted records, which we loaded into *R* using the package *ape* [8]. Subsequently, we generated UPGMA dendrograms with support values for the nodes based on 1000 replicates. Additionally, we computed maximum parsimony trees with support values for the nodes based on 1000 replicates. The UPGMA and maximum parsimony analyses were carried out using the *R* package *phangorn* [9]. All visualizations were generated using *FigTree* [10].

PCA-based ancestry deconvolution

As all SNPs are biallelic, we transformed the concatenated files in a binary matrix suitable as input for the program *SMARTPCA* [11] and its *R* wrapper *smartsnp.* In order to generate a PCA space and compute eigenvalues, we used a subset of isolates composed by all rice-infecting isolates and by wheat-infecting isolates from the initial wheat blast outbreak in Bangladesh (2016) [5, 7]. Subsequently, we used as a test set 13 isolates from Zambia and 53 newly-genotyped wheat-infecting isolates from Bangladesh, which we projected into the PCA space (PC1 and PC2).

Data availability

The missing information table is available as Supplementary Table 1. Custom scripts and trees used in this study can be found at https://github.com/Burbano-Lab/wheat_blast

References

1. Chiapello H, Mallet L, Guérin C, Aguileta G, Amselem J, Kroj T, et al. Deciphering Genome Content and Evolutionary Relationships of Isolates from the Fungus Magnaporthe oryzae Attacking Different Host Plants. Genome Biol Evol. 2015;7:2896–912. doi:10.1093/gbe/evv187.

2. Gladieux P, Condon B, Ravel S, Soanes D, Maciel JLN, Nhani A Jr, et al. Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus Magnaporthe oryzae. MBio. 2018;9. doi:10.1128/mBio.01219-17.

3. Urashima AS, Igarashi S, Kato H, Others. Host range, mating type, and fertility of Pyricularia grisea from wheat in Brazil. Plant Dis. 1993;77:1211–6. https://www.cabdirect.org/cabdirect/abstract/19942302466.

4. Inoue Y, Vy TTP, Yoshida K, Asano H, Mitsuoka C, Asuke S, et al. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. Science. 2017;357:80–3. doi:10.1126/science.aam9654.

5. Islam MT, Croll D, Gladieux P, Soanes DM, Persoons A, Bhattacharjee P, et al. Emergence of wheat blast in Bangladesh was caused by a South American lineage of Magnaporthe oryzae. BMC Biol. 2016;14:84. doi:10.1186/s12915-016-0309-7.

6. Tembo B, Mulenga RM, Sichilima S, M'siska KK, Mwale M, Chikoti PC, et al. 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. 2020;15:e0238724. doi:10.1371/journal.pone.0238724.

7. Tembo B, Mahmud NU, Paul SK, Asuke S, Harant A, Langner T, et al. Multiplex amplicon sequencing dataset for genotyping pandemic populations of the wheat blast fungus. 2021. doi:10.5281/zenodo.4605959.

8. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35:526–8. doi:10.1093/bioinformatics/bty633.

9. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27:592–3. doi:10.1093/bioinformatics/btq706.

10. Rambaut A. figtree. Github. https://github.com/rambaut/figtree. Accessed 18 Mar 2021.

11. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38:904–9. doi:10.1038/ng1847.

Supplementary Material
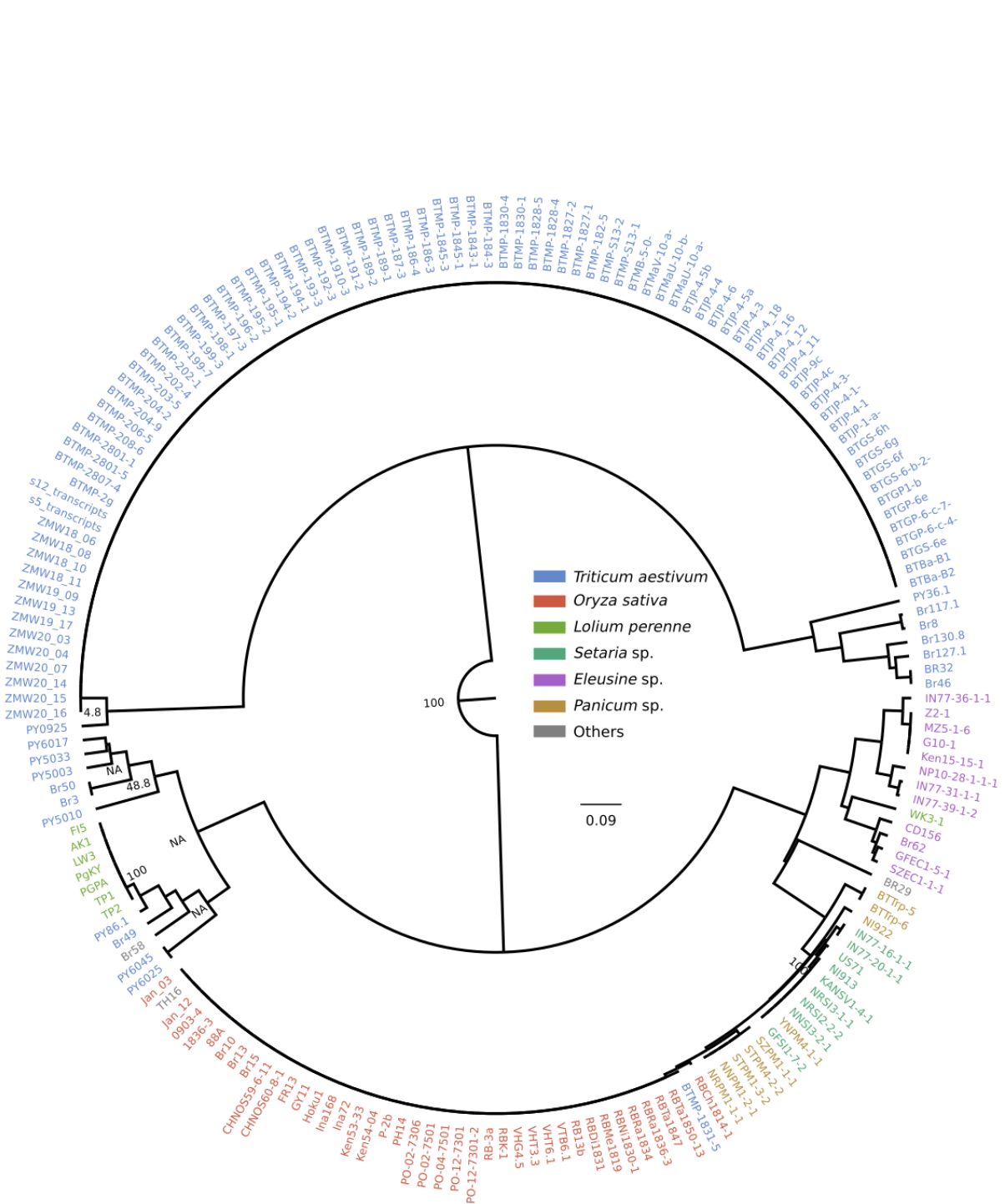
Supplementary Figure 1. UPGMA clustering of *Magnaporthe oryzae* isolates from a wide range of hosts. The dendrogram is based on the Genotype data from Tembo *et al* [7]. The labels are color coded according to host specificity (see inset). Values in the nodes represent the bootstrap support for 1000 replicates.
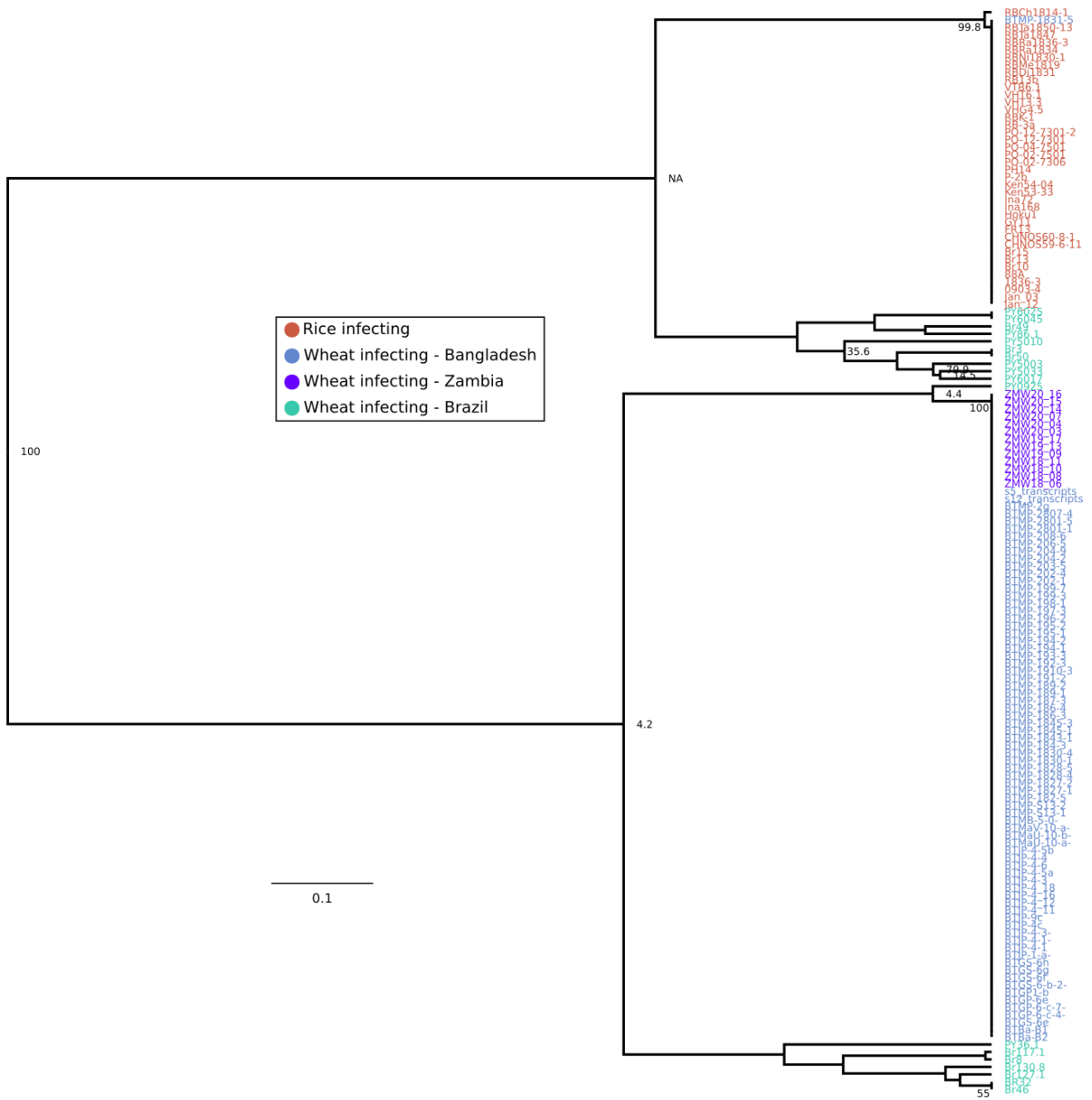
Supplementary. Figure 2. **UPGMA clustering of rice- and wheat blast infecting isolates.** The dendrogram is based on the Genotype data from Tembo *et al* [7]. The labels are color coded according to host specificity (see inset). Values in the nodes represent the bootstrap support for 1000 replicates.
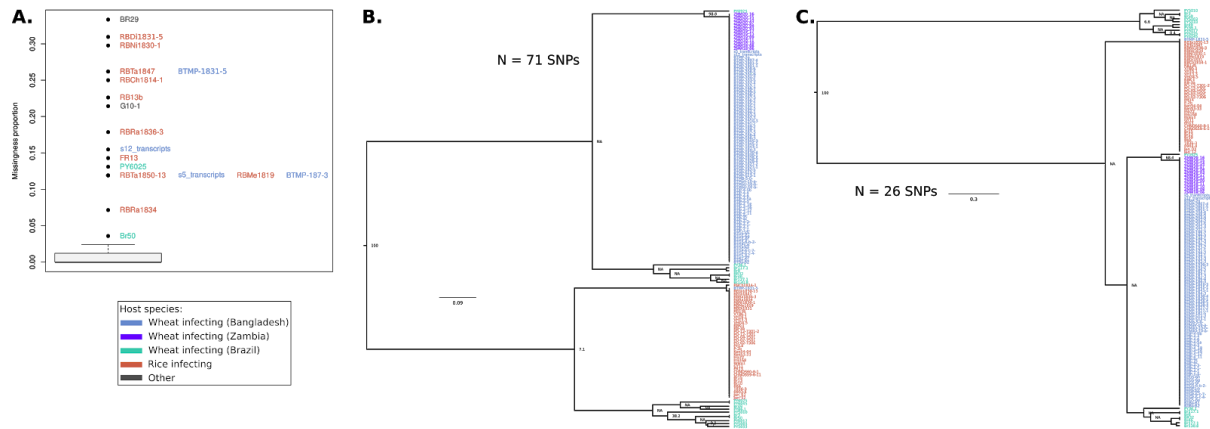
Supplementary Figure 3. Effect of missing data on the clustering of rice- and wheat- infecting isolates. A. The boxplot represents the distribution of the proportion of missing genotype positions per isolate. Isolates with higher missingess values are shown as outliers. **B.** UPGMA dendrogram based on a 5% of missing data allowance. **C.** UPGMA dendrogram based on full information SNPs. The labels are color coded according to host specificity (see inset).

Supplementary Table 1. Missing SNPs per isolate.

Isolate	Host_species	N_Missing_SNPs	Isolate	Host_species	N_Missing_SNPs
BR29	Digitaria_sanguinalis	28	PgKY	Lolium_perenne	0
BTMP-1831-5	Triticum_aestivum	26	P-2b	Oryza_sativa	0
RBNi1830-1	Oryzae_sativa	25	NRSI3-1-1	Setaria_italica	0
RBTa1847	Oryzae_sativa	22	NRSI2-2-2	Setaria_italica	0
RBDi1831	Oryzae_sativa	22	NRPM1-1-1	Panicum_miliaceum	0
RBCh1814-1	Oryzae_sativa	21	NP10-28-1-1-1	Eleusine_coracana	0
RB13b	Oryzae_sativa	19	NNSI3-2-1	Setaria_italica	0
G10-1	Eleusine_coracana	18	NNPM1-2-1	Panicum_miliaceum	0
RBRa1836-3	Oryzae_sativa	15	NI913	Setaria_viridis	0
s12_transcripts	Triticum_aestivum	13	MZ5-1-6	Eleusine_coracana	0
FR13	Oryza_sativa	12	LW3	Lolium_perenne	0
PY6025	Triticum_aestivum	11	Ken54-04	Oryza_sativa	0
s5_transcripts	Triticum_aestivum	10	Ken15-15-1	Eleusine_coracana	0
RBTa1850-13	Oryzae_sativa	10	KANSV1-4-1	Setaria_viridis	0
RBMe1819	Oryzae_sativa	10	Jan_03	Oryza_sativa	0
BTMP-187-3	Triticum_aestivum	10	IN77-39-1-2	Eleusine_coracana	0
RBRa1834	Oryzae_sativa	6	IN77-36-1-1	Eleusine_indica	0
Br50	Triticum_aestivum	3	IN77-31-1-1	Eleusine_coracana	0
BTGP-6e	Triticum_aestivum	2	IN77-20-1-1	Setaria_italica	0
Br130.8	Triticum_aestivum	2	IN77-16-1-1	Setaria_italica	0
PH14	Oryza_sativa	1	Hoku1	Oryza_sativa	0
NI922	Panicum_bisulcatum	1	Guy11	Oryza_sativa	0
Ken53-33	Oryza_sativa	1	GFSI1-7-2	Setaria_italica	0
Jan_12	Oryza_sativa	1	GFEC1-5-1	Eleusine_coracana	0
Ina72	Oryza_sativa	1	FI5	Lolium_perenne	0
Ina168	Oryza_sativa	1	CHNOS60-8-1	Oryza_sativa	0
BTTrp_6	torpedo_weed	1	CHNOS59-6-1-1	Oryza_sativa	0
BTTrp_5	torpedo_weed	1	CD156	Eleusine_indica	0
BTMP_S13_1	Triticum_aestivum	1	BTMP_S13-2	Triticum_aestivum	0
BTMP-202-4	Triticum_aestivum	1	BTMP-2g	Triticum_aestivum	0
BTJP4_6	Triticum_aestivum	1	BTMP-2807-4	Triticum_aestivum	0
BTJP4_5	Triticum_aestivum	1	BTMP-2801-5	Triticum_aestivum	0
BTJP4_3	Triticum_aestivum	1	BTMP-2801-1	Triticum_aestivum	0
BTJP4_18	Triticum_aestivum	1	BTMP-208-6	Triticum_aestivum	0
BTJP4_16	Triticum_aestivum	1	BTMP-206-5	Triticum_aestivum	0
BTJP4_12	Triticum_aestivum	1	BTMP-204-9	Triticum_aestivum	0
BTJP4_11	Triticum_aestivum	1	BTMP-204-2	Triticum_aestivum	0
BTGS6h	Triticum_aestivum	1	BTMP-203-5	Triticum_aestivum	0
BTGS6g	Triticum_aestivum	1	BTMP-202-1	Triticum_aestivum	0

BTGS6f	Triticum_aestivum	1	BTMP-199-7	Triticum_aestivum	0
BTG_S6e	Triticum_aestivum	1	BTMP-199-3	Triticum_aestivum	0
BTGP1_b	Triticum_aestivum	1	BTMP-198-1	Triticum_aestivum	0
BTBa_B2	Triticum_aestivum	1	BTMP-197-3	Triticum_aestivum	0
BTBa_B1	Triticum_aestivum	1	BTMP-196-2	Triticum_aestivum	0
Br8	Triticum_aestivum	1	BTMP-195-2	Triticum_aestivum	0
Br49	Triticum_aestivum	1	BTMP-195-1	Triticum_aestivum	0
Br3	Triticum_aestivum	1	BTMP-194-2	Triticum_aestivum	0
Br32	Triticum_aestivum	1	BTMP-194-1	Triticum_aestivum	0
Br117.1	Triticum_aestivum	1	BTMP-193-3	Triticum_aestivum	0
ZMW20_16	Triticum_aestivum	0	BTMP-192-3	Triticum_aestivum	0
ZMW20_15	Triticum_aestivum	0	BTMP-191-2	Triticum_aestivum	0
ZMW20_14	Triticum_aestivum	0	BTMP-1910-3	Triticum_aestivum	0
ZMW20_07	Triticum_aestivum	0	BTMP-189-2	Triticum_aestivum	0
ZMW20_04	Triticum_aestivum	0	BTMP-189-1	Triticum_aestivum	0
ZMW20_03	Triticum_aestivum	0	BTMP-186-4	Triticum_aestivum	0
ZMW19_17	Triticum_aestivum	0	BTMP-186-3	Triticum_aestivum	0
ZMW19_13	Triticum_aestivum	0	BTMP-1845-3	Triticum_aestivum	0
ZMW19_09	Triticum_aestivum	0	BTMP-1845-1	Triticum_aestivum	0
ZMW18_11	Triticum_aestivum	0	BTMP-184-3	Triticum_aestivum	0
ZMW18_10	Triticum_aestivum	0	BTMP-1843-1	Triticum_aestivum	0
ZMW18_08	Triticum_aestivum	0	BTMP-1830-4	Triticum_aestivum	0
ZMW18_06	Triticum_aestivum	0	BTMP-1830-1	Triticum_aestivum	0
Z2-1	Eleusine_coracana	0	BTMP-1828-5	Triticum_aestivum	0
YNPM4-1-1	Panicum_miliaceum	0	BTMP-1828-4	Triticum_aestivum	0
WK3-1	Lolium_perenne	0	BTMP-1827-2	Triticum_aestivum	0
VTB6.1	Oryza_sativa	0	BTMP-1827-1	Triticum_aestivum	0
VHT6.1	Oryza_sativa	0	BTMP-182-5	Triticum_aestivum	0
VHT3.3	Oryza_sativa	0	BTMB-5(0)	Triticum_aestivum	0
VHG4.5	Oryza_sativa	0	BTMaV-10(a)	Triticum_aestivum	0
US71	Setaria_sp.	0	BTMaU-10(b)	Triticum_aestivum	0
TP2	Lolium_perenne	0	BTMaU-10(a)	Triticum_aestivum	0
TP1	Lolium_perenne	0	BTJP-9c	Triticum_aestivum	0
TH16	Hordeum_vulgare	0	BTJP-4c	Triticum_aestivum	0
SZPM1-1-1	Panicum_miliaceum	0	BTJP4-5	Triticum_aestivum	0
SZEC1-1-1	Eleusine_coracana	0	BTJP4-4	Triticum_aestivum	0
STPM4-2-2	Panicum_miliaceum	0	BTJP-4(3)	Triticum_aestivum	0
STPM1-3-2	Panicum_miliaceum	0	BTJP-4(1)	Triticum_aestivum	0
RBK-1	Oryza_sativa	0	BTJP-4-1	Triticum_aestivum	0
RB-3a	Oryza_sativa	0	BTJP-1(a)	Triticum_aestivum	0
PY86.1	Triticum_aestivum	0	BTGS-6(b-2)	Triticum_aestivum	0
PY6045	Triticum_aestivum	0	BTGP-6(c-7)	Triticum_aestivum	0

PY6017	Triticum_aestivum	0	BTGP-6(C-4)	Triticum_aestivum	0
PY5033	Triticum_aestivum	0	Br62	Eleusine_indica	0
PY5010	Triticum_aestivum	0	Br58	Avena_sativa	0
PY5003	Triticum_aestivum	0	Br46	Triticum_aestivum	0
PY36.1	Triticum_aestivum	0	Br15	Oryza_sativa	0
PY0925	Triticum_aestivum	0	Br13	Oryza_sativa	0
PO-12-7301	Oryza_sativa	0	Br127.1	Triticum_aestivum	0
PO-12-7301-2	Oryza_sativa	0	Br10	Oryza_sativa	0
PO-04-7501	Oryza_sativa	0	AK1	Lolium_perenne	0
PO-02-7501	Oryza_sativa	0	88A	Oryza_sativa	0
PO-02-7306	Oryza_sativa	0	1836-3	Oryza_sativa	0
PGPA	Lolium_perenne	0	0903-4	Oryza_sativa	0