

Multiplex amplicon sequencing dataset for genotyping the wheat blast fungus from the Minas Gerais state of Brazil

João P. Ascari¹, Ignacio Cazón¹, Emerson M. Del Ponte¹, Mark Farman², Kurt Lamour³,

¹ Departamento de Fitopatologia, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

² Department of Plant Pathology, University of Kentucky, Lexington, Kentucky, United States of America

³ Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, Tennessee, United States of America

We report genotyping data for 81 *Magnaporthe oryzae* (syn. *Pyricularia oryzae*) *Triticum* pathotype (MoT) isolates obtained from wheat (*Triticum aestivum* L.) in Minas Gerais state, Brazil. The dataset consists of 84 single nucleotide polymorphisms (SNPs) selected from transcript sequences to specifically distinguish between the clonal lineage of the wheat blast fungus that emerged in South Asia in 2016 and other *M. oryzae* genotypes. The sample coverage includes isolates collected from leaves or spikes of wheat during the epidemic year of 2019 across 10 municipalities. The dataset is provided as an open science resource as part of the OpenWheatBlast initiative. We welcome analyses and feedback from the community.

Introduction

Wheat blast poses a serious threat for wheat production in the tropics of Brazil. There, epidemics are much more frequent and severe than in the subtropics of the south, where most of the wheat production is concentrated and from where most of the MoT isolates studied so far have been collected. Little is known about the diversity of the populations from Minas Gerais state, the third largest producer of wheat and where wheat blast is a current constraint to the expansion of wheat production. Here, we applied multiplex amplicon sequencing to genotype *M. oryzae* strains. We used a set of 84 single nucleotide polymorphisms (SNPs) as described in Tembo et al. (2021). The dataset is provided as an open science resource as part of the [OpenWheatBlast](#) initiative. We welcome analyses and feedback from the community.

Results and Conclusions

MoT isolates. A sample of 81 isolates, obtained from symptomatic leaves of heads, were selected from a larger collection of strains obtained from a survey in commercial wheat fields from Minas Gerais State, Brazil, during 2019 at 10 municipalities (Figure 1,

Supplemental Dataset 1). All isolates were purified, grown on PDA for further DNA extraction and PCR runs with MoT3 and C17 primers to confirm that they belong to *Triticum* pathotype. The DNA samples were sent to Kurt Lab for genotyping.

Selection of SNP panel. We selected 84 SNPs based on the transcriptomics data of Islam et al. (2016) and other blast fungus genome sequences available at the time (May 2017) (See Table 1, Supplemental Dataset 1 in Tembo et al. 2021). The panel of 84 SNPs was primarily designed to discriminate between the wheat blast clonal lineage of *MoT* in Bangladesh (BD lineage) from other genotypes. The targeted sequences and SNPs are provided in Supplemental Dataset 1 with the SNP indicated between brackets.

Multiplex amplification, sequencing and genotyping. We used 'MonsterPlex' amplification, a well-established method of multiplex amplicon sequencing (Nguyen-Dumont et al. 2013), as performed by Floodlight Genomics LLC as part of their no-cost Educational and Research Outreach Program (<http://floodlightgenomics.com/>, EROP, Knoxville, TN). We applied 'MonsterPlex' amplifications and Illumina DNA sequencing to the 81 isolates. This includes combining all 84 primer pairs into a single well for multiplex amplification of each sample. The genotypes were scored by aligning the raw sequences to reference sequences and requiring at least 20X coverage and an allele frequency >90% using CLC Genomics Workbench version 9.5.3 at the default settings (Qiagen, USA). The resulting genotype table with 81 isolates x 84 SNPs is provided for further analysis (Supplemental Dataset 2). Raw data is available on request. Missing data reflects partial coverage, possibly because the isolates may have failed to amplify with a subset of the amplicons due to low DNA template concentrations, genetic divergence or other reasons. The two transcript samples also yielded amplicons with only a subset of the primers. The MonsterPlex method resulted in a matrix of 84 SNPs for the collection of 81 samples (Supplemental Dataset 2). The merged sample-specific raw data was also delivered as FASTQ files for further analyses (Supplemental Dataset 3).

Data availability statement. We provide the dataset as an open science resource as part of the [OpenWheatBlast](#) initiative (Kamoun et al. 2019) and through the [Zenodo OpenWheatBlast Community](#).

Statement on continuing work. The datasets are being shared prior to formal publication and should be considered preliminary. We encourage and welcome analyses and feedback from the community. Please use the [Zenodo OpenWheatBlast Community](#) to share any analysis reports. We will post our own analyses of this dataset as soon as completed.

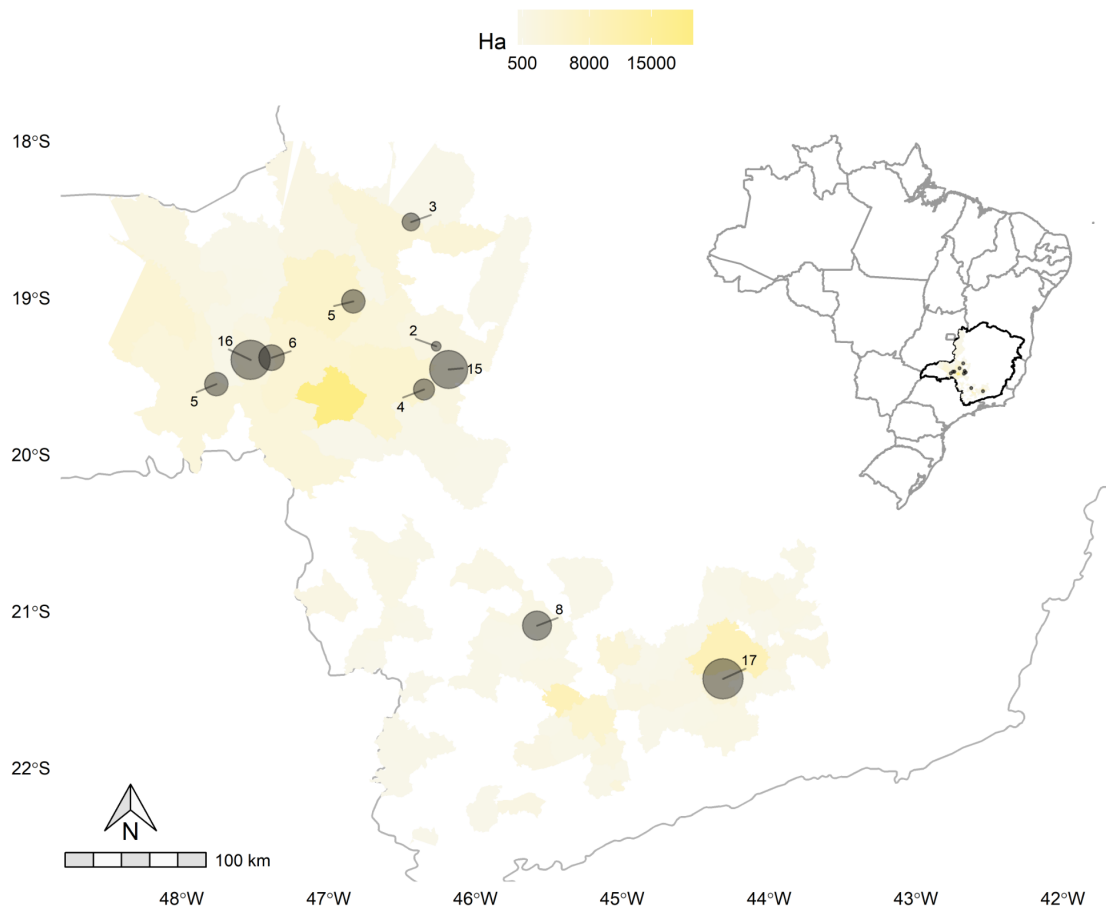


Figure 1. Map of the state of Minas Gerais, Brazil, depicting estimated wheat planting area (ha) per municipality (color gradient) and location (dots) of *Magnaporthe oryzae* wheat-infecting isolates. Dots size refers to the number of isolates collected.

Supplemental Datasets.

Supplemental Dataset 1. Information for 81 MoT isolates from Brazil analysed in this study

Supplemental Dataset 2. SNP data for 81 samples.

Supplemental Dataset 3. Targeted sequences for 85 SNPs and flanking

References

Islam, T., Croll, D., Gladieux, P., Soanes, D., Persoons, A., Bhattacharjee, P., Hossain, S., Gupta, D., Rahman, Md.M., Mahboob, M.G., Cook, N., Salam, M., Surovy, M.Z., Bueno Sancho, V., Maciel, J.N., Nani, A., Castroagudin, V., de Assis Reges, J.T., Ceresini, P., Ravel, S., Kellner, R., Fournier, E., Tharreau, D., Lebrun, M.-H., McDonald, B., Stitt, T., Swan, D., Talbot, N., Saunders, D., Win, J., and Kamoun, S. 2016. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biology*, 14:84.

Kamoun, S., Talbot, N.J., and Islam, M.T. 2019. Plant health emergencies demand open science: Tackling a cereal killer on the run. *PLOS Biology*, 17:e3000302.

Nguyen-Dumont, Tú, Pope, B.J., Hammet, F., Southey, M.C., and Park, D.J. 2013. A High-Plex PCR Approach for Massively Parallel Sequencing. *BioTechniques* 55, 2.

Tembo, B, Nur Uddin Mahmud, Sanjoy Kumar Paul, Soichiro Asuke, Adeline Harant, Thorsten Langner, C. Sarai Reyes-Avila, Emilie Chanclud, Vincent Were, Suwilanji Sichilima, Rabson M. Mulenga, Dipali Rani Gupta, Md. Shabab Meheub, Abu Naim Md. Muzahid, . M. Fajle Rabby, Pawan K. Singh, Alison Bentley, Yukio Tosa, Daniel Croll, Kurt Lamour, Tofazzal Islam, Nicholas J. Talbot, Sophien Kamoun, Joe Win. 2021. Multiplex amplicon sequencing dataset for genotyping pandemic populations of the wheat blast fungus. <https://doi.org/10.5281/zenodo.4605959>