

# The First Annotated Genome Assembly of *Macrophomina tecta* Associated with Charcoal Rot of Sorghum

Barsha Poudel<sup>1</sup>, Neeraj Purushotham<sup>1,2</sup>, Ashley Jones<sup>3</sup>, Jamila Nasim<sup>2,3</sup>, Dante L. Adorada<sup>1</sup>, Adam H. Sparks <sup>1,4</sup>, Benjamin Schwessinger <sup>3</sup>, and Niloofar Vaghefi <sup>1,5,\*</sup>

<sup>1</sup>Centre for Crop Health, University of Southern Queensland, Toowoomba, QLD, Australia

<sup>2</sup>Loam Bio, Orange, NSW, Australia

<sup>3</sup>Research School of Biology, Australian National University, Canberra, Australia

<sup>4</sup>Department of Primary Industries and Regional Development, Perth, WA, Australia

<sup>5</sup>School of Agriculture and Food, University of Melbourne, Melbourne, VIC, Australia

\*Corresponding author: E-mail: vaghefin@unimelb.edu.au.

Accepted: 25 May 2022

## Abstract

Charcoal rot is an important soilborne disease caused by a range of *Macrophomina* species, which affects a broad range of commercially important crops worldwide. Even though *Macrophomina* species are fungal pathogens of substantial economic importance, their mechanism of pathogenicity and host spectrum are poorly understood. There is an urgent need to better understand the biology, epidemiology, and evolution of *Macrophomina* species, which, in turn, will aid in improving charcoal rot management strategies. Here, we present the first high-quality genome assembly and annotation of *Macrophomina tecta* strain BRIP 70781 associated with charcoal rot symptoms on sorghum. Hybrid assembly integrating long reads generated by Oxford Nanopore Technology and short Illumina paired-end reads resulted in 43 contigs with a total assembly size of ~54 Mb, and an N50 of 3.4 Mb. In total, 12,926 protein-coding genes and 7,036 repeats were predicted. Genome comparisons detected accumulation of DNA transposons in *Macrophomina* species associated with sorghum. The first reference genome of *M. tecta* generated in this study will contribute to more comparative and population genomics studies of *Macrophomina* species.

**Key words:** stalk/dry root rot, de novo assembly, genome annotation, transposable elements, *Macrophomina phaseolina*.

## Significance

The high-quality annotated reference genome of *Macrophomina tecta* will provide a valuable resource for comparative and population genomics studies. Such studies will be important for shedding light on the biology and evolution of *Macrophomina* species, developing species-specific molecular markers and identifying genetic determinants of pathogenicity in these agriculturally important plant pathogens. The generated information can broaden the understanding of *Macrophomina*-sorghum interaction and can be translated to better management options for charcoal rot.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Introduction

*Macrophomina tecta* (family: *Botryosphaeriaceae*) is a recently discovered fungal plant pathogen (Poudel et al. 2021). This pathogen was isolated from stems of sorghum (*Sorghum bicolor*) and mungbean (*Vigna radiata*) plants with charcoal rot symptoms. Charcoal rot is an important soilborne disease that affects a broad range of broadacre, horticultural, and vegetable crops worldwide (Kaur et al. 2012; Marquez et al. 2021). To date, five species of *Macrophomina* have been identified based on multi-locus phylogenetic analyses, namely, *M. phaseolina* (Tassi 1901; Sarr et al. 2014), *M. pseudophaseolina* (Sarr et al. 2014), *M. euphorbiicola* (Machado et al. 2019), *M. vaccinii* (Zhao et al. 2019), and *M. tecta* (Poudel et al. 2021). *Macrophomina phaseolina* and *M. pseudophaseolina* have been reported to infect >400 host plants including cotton (*Gossypium hirsutum*), mungbean, peanut (*Arachis hypogaea*), sorghum, soybean (*Glycine max*), and sunflower (*Helianthus annuus*) (Sarr et al. 2014; Farr and Rossman 2021). Conversely, the other three *Macrophomina* species have been reported on a limited number of host plants. *Macrophomina euphorbiicola* has only been reported on *Ricinus communis* and *Jatropha gossypifolia* (Machado et al. 2019); *M. vaccinii* on blueberry (*Vaccinium corymbosum* × *V. darrowii*) (Zhao et al. 2019); and *M. tecta* on sorghum and mungbean (Poudel et al. 2021). The host range of a pathogen has a major impact on the emergence and spread of a disease, especially with a pathogen like *M. phaseolina* infecting multiple economically important crops. Therefore, understanding the underlying molecular mechanisms and the driving evolutionary forces facilitating host spectrum in *Macrophomina* species is crucial in controlling the spread of the disease.

One such component that has been considered to play an important role in the evolutionary mechanism and adaptive nature of fungi are transposable elements (TEs) (Möller and Stukenbrock 2017). Previous studies in phytopathogens have reported influence of TEs on genome plasticity (Moolhuijzen et al. 2018; Lorrain et al. 2021), pathogenicity (Singh et al. 2021), host range (Li et al. 2021), and evolution (Oggenfuss et al. 2021). In *M. phaseolina*, some groups of TEs were found to be specific to the strain obtained from strawberry (*Fragaria* × *ananassa*) when compared with a strain from alfalfa (*Medicago sativa*) (Burkhardt et al. 2019). This is suggestive of the potential role of TEs in host adaptation of *M. phaseolina*. The high-quality *M. tecta* genome produced here can provide further insights on role of TEs in host adaptation by facilitating comparative analysis of TEs in *Macrophomina* genomes (Islam et al. 2012; Burkhardt et al. 2019; Purushotham et al. 2020).

To date, *M. tecta* has only been detected in Australia and its geographical distribution, host range, pathogenicity, and center of origin are not known. Grain sorghum is

considered to have been domesticated around central Africa and later was introduced to Asia, America, and Australia (Spenceley et al. 2005). Grain sorghum was first grown in Queensland in 1938 and in New South Wales in 1940, using dwarf varieties imported from the USA. The pathogen may have been introduced to Australia through infested seed during international import or jumped on sorghum from other hosts. *Macrophomina tecta* could have emerged due to the selection of genotypes highly adapted to sorghum, leading to a selective sweep. Population genomics analyses may provide further insights on the role of selection sweeps in shaping *M. tecta* populations in Australia. Furthermore, comparative analyses among *Macrophomina* species will help identify genes underlying adaptation and their distribution in the genome. Here, we present a first genome assembly for the ex-holotype strain of *M. tecta* produced by incorporation of whole-genome DNA sequence data from both Oxford Nanopore Technology (ONT) and Illumina sequencing. In addition, an annotation of the genome was conducted using RNA-seq transcript evidence.

## Results and Discussion

### Genome Assembly and Annotation

*Macrophomina tecta* strain BRIP 70781 was assembled into 43 contigs with an N50 of 3.4 Mb and the largest contig of 5.1 Mb (table 1). The size of the draft genome was ~54 Mb with a GC content of 52.09% (table 1). Illumina sequencing accounted for ~110× and ONT sequencing accounted for ~48× coverage of the whole genome. Based on the identification of core Dothideomycetes Benchmarking Universal Single-Copy Orthologs (BUSCOs), the assembled genome is 94.7% complete [complete core genes = 3,586; single: 3,560 (94.0%); duplicated: 26 (0.7%); fragmented: 81 (2.1%); missing: 119 (3.2%) from a total of 3,786 BUSCOs].

Currently, four published genomes of *M. phaseolina* are publicly available, which include strain BRIP 70780 from sorghum (Purushotham et al. 2020), strain MS6 from Jute (*Corchorus olitorius*) (Islam et al. 2012), strain 11–12 from strawberry, and strain AI-1 from alfalfa (Burkhardt et al. 2019). The reported assembly sizes of these *M. phaseolina* genomes range from 48.9 to 51.3 Mb, which are slightly smaller than the *M. tecta* assembly size of ~54 Mb reported here (table 1).

Gene prediction with BRAKER2 (Hoff et al. 2019) annotation tool identified 12,926 protein-coding genes in the assembled *M. tecta* genome. The number of predicted genes estimated was comparable to the four publicly available *M. phaseolina* strains (table 1). The assembled genome of *M. tecta* strain BRIP 70781 consisted of 233 noncoding RNA consisting of 119 transferase RNA, and 43 ribosomal RNA.

**Table 1**Genome Statistics for *Macrophomina* Strains Sequenced to Date

Species	Strain	Host	Assembly size (Mb)	No. of contigs	N50 (Mb)	Largest contig (Mb)	Protein-coding genes	Genome completeness <sup>a</sup>	NCBI accession number	Reference
<i>M. phaseolina</i>	BRIP 70780	Sorghum	52.5	22	4.2	7.02	14,471	95%	PRJNA577531	Purushotham et al. (2020)
	11–12	Strawberry	51.3	60	4.3	6.8	14,103	94.9%	PRJNA428521	Burkhardt et al. (2019)
	Al-1	Alfalfa	49.8	18	5.0	6.8	13,443	95%	PRJNA432410	Burkhardt et al. (2019)
	MS6	Jute	48.9	1,506	0.15	1.1	14,249	94.8%	PRJNA78845	Islam et al. (2012)
<i>M. tecta</i>	BRIP 70781	Sorghum	54.1	43	3.4	5.1	12,926	94.7%		This study

<sup>a</sup>Genome completeness for all genomes was estimated based on Benchmarking Universal Single-Copy Orthologs (BUSCOs) against the *dothideomycetes\_odb10* database (Simao et al. 2015).

Carbohydrate-active enzymes (CAZymes) produced by pathogens are involved in breakdown of plant cell wall as well as in host–pathogen interaction (Ospina-Giraldo et al. 2010). Three hundred and thirty-five putative CAZymes were predicted by all three tools (hmmer, diamond, and Hotpep) in dbCRAN meta-server (Zhang et al. 2018), 151 of which contained a signal peptide. CAZymes consisted of 191 glycoside hydrolases, 57 glycosyl transferases, 10 polysaccharide lyases, 9 carbohydrate esterases (CEs), 4 noncatalytic carbohydrate-binding modules, and 64 auxiliary activities (AA) classes. A diverse array of CAZymes classes was also previously detected in *M. phaseolina* (Islam et al. 2012; Burkhardt et al. 2019).

Sixty-two putative secondary metabolites biosynthetic gene clusters (BGCs) comprising of 14 nonribosomal peptide synthetases (NRPS), 14 Type 1 polyketide synthases (T1PKS), 7 terpene, 2 betalactone, and 19 T1PKS/NRPS like metabolites were detected using antiSMASH fungal v.6.0.1 (Blin et al. 2021). Terpene-derived secondary metabolites have been shown to contribute to pathogenicity in *Fusarium* species (Proctor et al. 2009). Secondary metabolite BGCs have been shown to play a role in host range determination. In *Eutiarospora* species, the presence of a PKS–NRPS gene cluster in *E. darliae* and *E. pseudodarliae*, but absence in *E. tritici-australis* is likely to allow the former two species to infect woody plants (Thynne et al. 2019).

Secretory proteins and effector proteins play an important role in fungal pathogenicity and virulence (Selin et al. 2016). In total, 1,201 putative secretory proteins were detected using SignalP/TMHMM (Chen et al. 2003; Armenteros et al. 2019) and 332 cytoplasmic/apoplast putative candidate effectors were detected in *M. tecta* strain BRIP 70781 using EffectorP v.3.0 (Sperschneider and Dodds 2021).

### Repeat Annotation

A total repeat content of 11 Mb, consisting of ~20.34% of the genome, was identified. Among the classified TEs,

1,562 were DNA transposons (3.35% of the genome), 1,015 were long interspersed nuclear elements (LINE) (1.80%), and 4,459 were long terminal repeat retrotransposons elements (11.63%). The total number of classified TEs detected was similar in the genome of *M. tecta* (7,036) and *M. phaseolina* (7,313) from sorghum, whereas fewer TEs have been detected in the genome of *M. phaseolina* from jute (4,885), strawberry (5,294), and alfalfa (4,717), which were annotated using RepeatModeller/RepeatMasker (repeatmasker.org). The genome of *M. tecta* and *M. phaseolina* strains obtained from sorghum contained a greater number of DNA transposons that is, 1417–1562, including hobo-Activator and Tc1-IS630-Pogo compared with the genome of *M. phaseolina* strain obtained from strawberry (785 DNA transposons), jute (507 DNA transposons), and alfalfa (522 DNA transposons) (fig. 1A). The accumulation of DNA transposons in *Macrophomina* species associated with sorghum is suggestive of the potential role of TEs in the adaptation and evolution of *Macrophomina* on this crop. In the wheat fungal pathogen *Zymoseptoria tritici* and its four wild-grass infecting sister species, the TE proportions varied in the genome of the *Zymoseptoria* species (Lorrain et al. 2021). The TE proportions of four wild-grass infecting *Zymoseptoria* species were higher than the TE proportions of *Z. tritici* strains. The TE variation observed *Zymoseptoria* species was shown to be associated with host adaptation and genome evolution (Lorrain et al. 2021).

### Comparison of Orthologous Genes

Groups of orthologous genes (orthogroups) were inferred from *M. tecta* and four *M. phaseolina* strains using OrthoVenn2 (Xu et al. 2019). We found 9,618 orthogroups with 49,463 genes common among *Macrophomina* strains (fig. 1B). Enriched gene ontology (GO) terms associated with these orthogroups are oxidoreductase activity (GO ID: GO:0016705 and GO:0016491) and torpedoed biosynthetic process (GO ID: GO:0016114). Oxidoreductase

activity and terpene-derived secondary metabolites have been correlated with virulence in *M. phaseolina* (Khan et al. 2017) and *Fusarium* (Proctor et al. 2009). In *M. tecta* strain BRIP 70781, sixty-nine unique orthogroups (fig. 1B) containing a total of 151 proteins and 870 singletons were detected. These strain-specific orthogroups warrant further investigations as they have been associated with host specialization in other fungal pathogens such as in host-specific *Ustilago* (Benevenuto et al. 2018) and *Botrytis* species (Valero-Jiménez et al. 2019).

In conclusion, this work provides a first high-quality annotated genome assembly of *M. tecta*. Comparative genomics of *Macrophomina* species has detected accumulation of DNA transposons in *Macrophomina* species associated with charcoal rot in sorghum. We identified strain-specific orthogroups, which could have a potential role in host specialization, thus, require further investigations. This genome assembly will make possible future comparative and population genomics studies to understand the biology and evolution process of *M. tecta* and other *Macrophomina* species.

## Materials and Methods

### Fungal Strains, DNA and RNA Extraction

The ex-holotype of *M. tecta* strain BRIP 70781 (Poudel et al. 2021) was grown for 7 days in potato dextrose broth (Amyl Media, Australia), at room temperature on a shaker at 220 RPM. For Illumina sequencing, DNA was extracted using a DNeasy Plant Mini Kit (Qiagen, Australia) according to manufacturer's instructions. For ONT sequencing, high molecular weight DNA was obtained by grinding ~2 g of fresh mycelia with a mortar and pestle, followed by a customized cetyltrimethylammonium bromide extraction method (Jones et al. 2019). The DNA was purified with chloroform and size-selected for fragments larger than 25 kb, using a Short-Read Eliminator Kit (Circulomics, US) (Jones et al. 2019). RNA was isolated using a RNeasy Plant Mini kit (Qiagen, Australia) following the manufacturer's instructions. Genomic DNA was removed via on-column DNase treatment. The final DNA and RNA samples were quantified using a Qubit v.3.0 fluorometer (Thermo Fisher Scientific, Australia) and absorbance at 260/280 and 260/230 nm was assessed using a Nanodrop 1000 (Thermo Fisher Scientific, Australia). Gel electrophoresis on a 1% agarose gel was used to assess DNA integrity.

### Sequencing

Short-read DNA sequencing was performed using an Illumina MiSeq platform with a 600-cycle paired-end V3 reagents kit. Illumina Nextera DNA Flex pair-end libraries were prepared according to the manufacturer's instructions, resulting in 550 bp fragments. Long reads were

obtained using the ONT MinION sequencer on FLO-MIN106 R9.4.1 revD flow cell. The library was prepared using the ligation kit SQK-LSK109 according to the manufacturer's instructions. Fast5 files were base-called to fastq with Guppy v.3.1.5 (Wick et al. 2019). Short-read whole-transcriptome (RNA) shotgun sequencing was conducted by the Australian Genome Research Facility (Melbourne, Australia) on a NextSeq platform to acquire 150-bp paired-end reads.

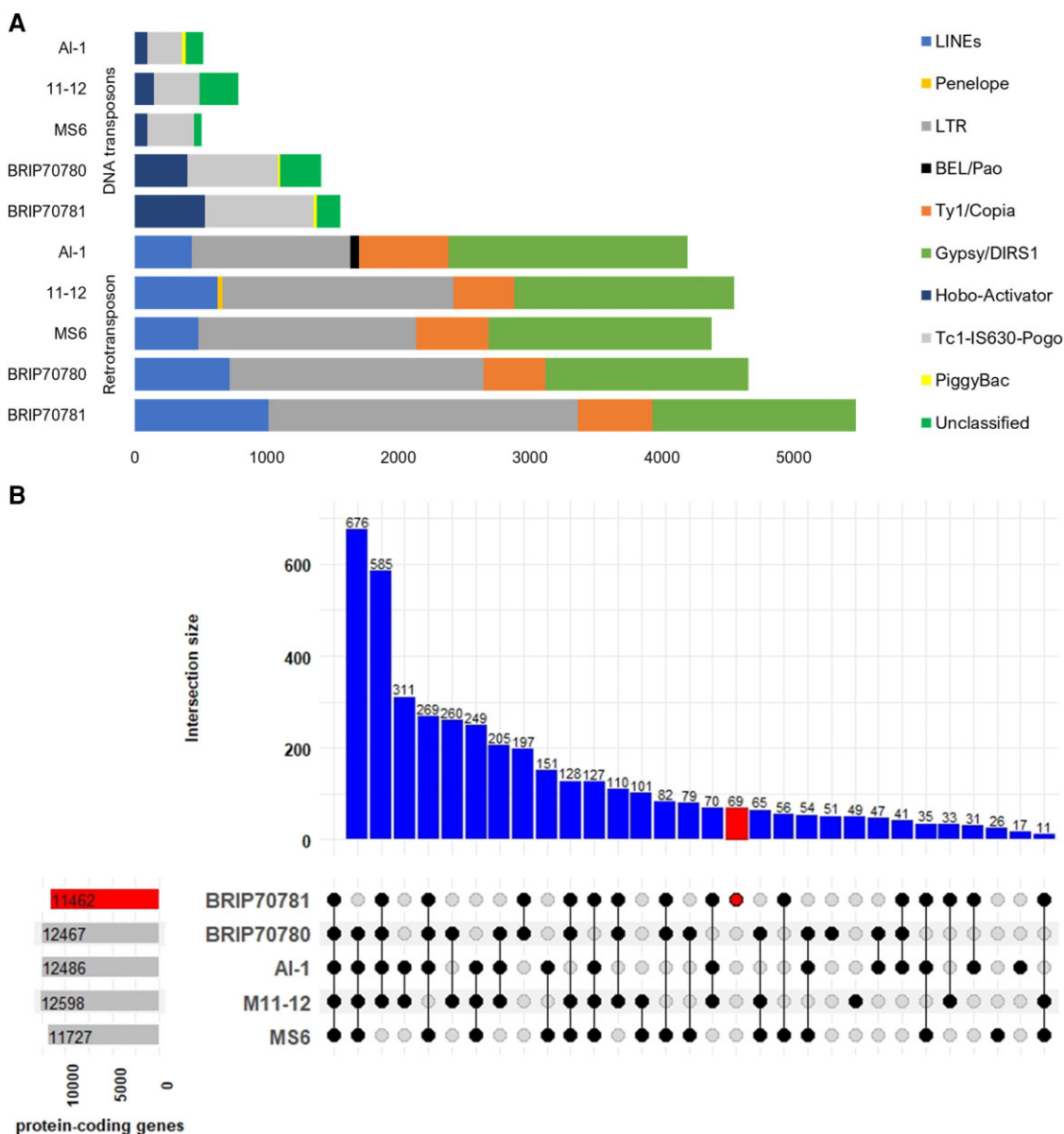
### Genome Assembly

Potential bacterial contaminants in the raw reads were removed using Kraken2 v.2.1.1 (Wood et al. 2019) with a confidence threshold of 0.05 for Nanopore and 0.1 for Illumina reads. The de novo assembly was conducted using a hybrid approach integrated in the Maryland Super Read Cabog Assembler v.3.4.2 (MaSuRCA) (Zimin et al. 2013). For assembly, both the long Nanopore reads, and short Illumina reads were used without quality filtering as recommended by the developer. The resulting intermediate assembly sequences were corrected using Illumina reads via POLCA available in MaSuRCA v.3.4.2 (Zimin and Salzberg 2020). Sequencing statistics for the final corrected assembly including overall percent GC content, were estimated using QUAST v.5.0.2 (Gurevich et al. 2013). We assessed the completeness of the genome assembly with BUSCO v.5.1.2 (Simao et al. 2015) against the dothideomycetes\_odb10 database.

Raw RNA reads were adapter-trimmed and bacterial contamination was removed using BBDuk v.38.90 (sourceforge.net/projects/bbmap/). Quality trimming was conducted using Trimmomatic (Bolger et al. 2014) v.0.36 with following settings: ILLUMINACLIP:TruSeq2-PE.fa:5:30:10 SLIDINGWINDOW:3:18 LEADING:6 TRAILING:6 MINLEN:90. The quality of the RNA-seq data was assessed via FastQC v.0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The reads were mapped to the reference assembly of *M. tecta* using HISAT2 v.2.2.1 (Kim et al. 2019) and the output file was parsed through samtools v.1.7 (Li et al. 2009) to obtain sorted BAM file.

### Genome and Functional Annotation

For ab initio gene prediction, BRAKER2 v.2.1.6 (Hoff et al. 2019) was used with RNA-seq data of *M. tecta* as an input. Functional annotation of the predicted protein was conducted using Interproscan v.5.44-79.0 (Jones et al. 2014). For the noncoding RNA families, the Rfam database v.14.6 (Kalvari et al. 2021) was scanned with cmscan of Infernal v.1.1.3 (Nawrocki and Eddy 2013). The dbCAN2 meta-server was used for predicting CAZymes (Zhang et al. 2018). Secondary Metabolites BGCs were detected using antiSMASH v.6.0.1 web-based tool (Blin et al.



**Fig. 1.**—Comparison of repeat elements and orthologous groups of *Macrophomina* genomes. (A) A total number of repeat elements for DNA transposons and retrotransposons detected in *M. phaseolina* (strains Al-1, 11–12, MS6, and BRIP 70780) and *M. tecta* (strain BRIP 70781). DNA transposons including hobo-Activator and Tc1-IS630-Pogo are more abundant in *M. phaseolina* strain BRIP 70780 and *M. tecta* strain BRIP 70781, both of which are associated with charcoal rot of sorghum. (B) UpSet plot showing the comparison of orthologous groups of *M. tecta* strain BRIP 70781 and four genomes of *M. phaseolina* strains. The red bar and dot represent the unique orthologous detected in *M. tecta*. The black dots/lines represent overlaps between orthologous clusters. The orthologous groups shared between five genomes was 9,618, which is excluded in the intersection size plot.

2021). Secreted proteins were predicted based on the presence of signal peptide and zero/one transmembrane domain as identified by SignalP v.5.0 (Armenteros et al. 2019) and TMHMM v.2.0 (Chen et al. 2003), respectively. Further, candidate effectors potentially involved in virulence were detected using EffectorP v.3.0 (Sperschneider and Dodds 2021) in the predicted secreted proteins.

### Repeat Annotation

Repeats for *M. tecta* and all other published *M. phaseolina* genomes were identified using RepeatModeler v.2.0.1 (repeatmasker.org/RepeatModeler/) with Repbase v.20.4 library (Bao et al. 2015) and RepeatMasker v.4.0.9 (repeatmasker.org/RMDownload.html) to perform comparison.

## Orthologous Comparison

Predicted proteins of *M. tecta* and other publicly available *M. phaseolina* genome were clustered into orthologous groups using OrthoVenn2 online toolkit (Xu et al. 2019) and visualized using a ComplexUpset library in Rstudio to create UpSet plot (Lex et al. 2014). In addition, GO term enrichment analysis was also performed. All enriched GO term groups had an *E*-value of 0.01 and an inflation value of 1.5 as specified with the OrthoVenn2.

## Acknowledgments

This project was supported by the University of Southern Queensland, Australia, and the Australian Grains Research and Development Corporation (GRDC) project 9176684. The authors thank Ms. Lauren Huth and Ms. Katelynn Hadzi for their assistance in Illumina MiSeq work, and Ms. Encarnacion Adorada and Mr. Peter Buyoyu for excellent technical support.

## Data Availability

Raw Illumina and ONT genome sequencing data for *M. tecta* BRIP 70781 described here, and the reference genome assembly have been deposited in the NCBI database, under the BioProject accession number PRJNA808193 and BioSample number SAMN26035986. The GenBank assembly accession number is AKSGN000000000. Raw data and files with information about the genome annotation, functional prediction, repeats, effectors, and orthologous genes have been deposited on Zenodo with the DOI number (<https://doi.org/10.5281/zenodo.6496539>).

## Literature Cited

- Armenteros JJA, et al. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 37:420–423.
- Bao WD, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA.* 6:11.
- Benevenuto J, Teixeira-Silva NS, Kuramae EE, Croll D, Monteiro-Vitorello CB. 2018. Comparative genomics of *smut* pathogens: insights from orphans and positively selected genes into host specialization. *Front Microbiol.* 9:660.
- Blin K, et al. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49: W29–W35.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30:2114–2120.
- Burkhardt AK, Childs KL, Wang J, Ramon ML, Martin FN. 2019. Assembly, annotation, and comparison of *Macrophomina phaseolina* isolates from strawberry and other hosts. *BMC Genomics.* 20: 802.
- Chen YJ, Yu P, Luo JC, Jiang Y. 2003. Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mamm Genome.* 14:859–865.
- Farr D, Rossman A. 2021. Fungal databases, US national fungus collections. ARS, USDA. Available from <https://nt.ars-grin.gov/fungaldatabases/>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QCAST: quality assessment tool for genome assemblies. *Bioinformatics.* 29: 1072–1075.
- Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. *Methods Mol Biol.* 1962:65–95.
- Islam M, et al. 2012. Tools to kill: genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina*. *BMC Genomics.* 13:493.
- Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 30:1236–1240.
- Jones A, Nagar R, Sharp A, Schwessinger B. 2019. High-molecular weight DNA extraction from challenging fungi using CTAB and gel purification. *Protocols.io* online.
- Kalvari I, et al. 2021. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* 49: D192–D200.
- Kaur S, et al. 2012. Emerging phytopathogen *Macrophomina phaseolina*: biology, economic importance and current diagnostic trends. *Crit Rev Microbiol.* 38:136–151.
- Khan AN, et al. 2017. Molecular identification and genetic characterization of *Macrophomina phaseolina* strains causing pathogenicity on sunflower and chickpea. *Front Microbiol.* 8:1309.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 37:907–915.
- Lex A, Gehlenborg N, Strobel H, Vuillemot R, Pfister H. 2014. UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph.* 20: 1983–1992.
- Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics.* 25:2078–2079.
- Li J, Fokkens L, Rep M. 2021. A single gene in *Fusarium oxysporum* limits its host range. *Mol Plant Pathol.* 22:108–116.
- Lorrain C, Feurtey A, Möller M, Hauelsen J, Stukenbrock E. 2021. Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defenses. *G3 Genes|Genomes|Genetics.* 11:jkab068.
- Machado AR, Pinho DB, Soares DJ, Gomes AAM, Pereira OL. 2019. Bayesian analyses of five gene regions reveal a new phylogenetic species of *Macrophomina* associated with charcoal rot on oilseed crops in Brazil. *Eur J Plant Pathol.* 153:89–100.
- Marquez N, Giachero ML, Declerck S, Ducasse DA. 2021. *Macrophomina phaseolina*: general characteristics of pathogenicity and methods of control. *Front Plant Sci.* 12:634397–634397.
- Möller M, Stukenbrock EH. 2017. Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol.* 15:756–771.
- Moolhuijzen P, et al. 2018. Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. *BMC Genomics* 19:279.
- Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29:2933–2935.
- Oggenfuss U, et al. 2021. A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *eLife* 10:e69249.
- Ospina-Giraldo MD, Griffith JG, Laird EW, Mingora C. 2010. The CAZyme of phytophthora spp.: a comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus *Phytophthora*. *BMC Genomics* 11:525.
- Poudel B, et al. 2021. Hidden diversity of *Macrophomina* associated with broadacre and horticultural crops in Australia. *Eur J Plant Pathol.* 161:1–23.
- Proctor RH, McCormick SP, Alexander NJ, Desjardins AE. 2009. Evidence that a secondary metabolic biosynthetic gene cluster has grown by gene relocation during evolution of the filamentous fungus *Fusarium*. *Mol Microbiol.* 74:1128–1142.

- Purushotham N, et al. 2020. Draft genome resource for *Macrophomina phaseolina* associated with charcoal rot in sorghum. *Mol Plant Microb Interact.* 33:724–726.
- Sarr MP, Ndiaye MB, Groenewald JZ, Crous PW. 2014. Genetic diversity in *Macrophomina phaseolina*, the causal agent of charcoal rot. *Phytopathol Mediterr.* 53:250–268.
- Selin C, de Kievit TR, Belmonte MF, Fernando WGD. 2016. Elucidating the role of effectors in plant-fungal interactions: progress and challenges. *Front Microbiol.* 7:600.
- Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.
- Singh NK, Badet T, Abraham L, Croll D. 2021. Rapid sequence evolution driven by transposable elements at a virulence locus in a fungal wheat pathogen. *BMC Genomics* 22:393.
- Spenceley J, et al. 2005. Grain sorghum. *Dubbo: Communication Branch.* p. 1–27.
- Sperschneider J, Dodds PN. 2021. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. Cold Spring Harbor Laboratory.
- Tassi A. 1901. Novae micromycetum species. *Ser III Bull Lab Orto Bot Univ Siena.* 4: 7–12.
- Thynne E, Mead OL, Chooi Y-H, McDonald MC, Solomon PS. 2019. Acquisition and loss of secondary metabolites shaped the evolutionary path of three emerging phytopathogens of wheat. *Genome Biol Evol.* 11:890–905.
- Valero-Jiménez CA, Veloso J, Staats M, Van Kan JAL. 2019. Comparative genomics of plant pathogenic *Botrytis* species with distinct host specificity. *BMC Genomics.* 20:203.
- Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 20: 129.
- Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20:257.
- Xu L, et al. 2019. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 47:W52–W58.
- Zhang H, et al. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46: W95–W101.
- Zhao L, Cai J, He W, Zhang Y. 2019. *Macrophomina vaccinii* sp. nov. causing blueberry stem blight in China. *MycKeys* 55:1.
- Zimin AV, et al. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677.
- Zimin AV, Salzberg SL. 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Comput Biol.* 16:e1007981.

**Associate editor:** Li-Jun Ma