



# IMA Genome-F 11

## Draft genome sequences of *Fusarium xylarioides*, *Teratosphaeria gauchensis* and *T. zuluensis* and genome annotation for *Ceratocystis fimbriata*

Brenda D. Wingfield<sup>1\*</sup>, Arista Fourie<sup>1†</sup>, Melissa C. Simpson<sup>1†</sup>, Vuyiswa S. Bushula-Njah<sup>1</sup>, Janneke Aylward<sup>1</sup>, Irene Barnes<sup>1</sup>, Martin P. A. Coetzee<sup>1</sup>, Léanne L. Dreyer<sup>2</sup>, Tuan A. Duong<sup>1</sup>, David M. Geiser<sup>3</sup>, Francois Roets<sup>4</sup>, E. T. Steenkamp<sup>1</sup>, Magriet A. van der Nest<sup>1,6</sup>, Carel J. van Heerden<sup>5</sup> and Michael J. Wingfield<sup>1</sup>

### Abstract

Draft genomes of the fungal species *Fusarium xylarioides*, *Teratosphaeria gauchensis* and *T. zuluensis* are presented. In addition an annotation of the genome of *Ceratocystis fimbriata* is presented. Overall these genomes provide a valuable resource for understanding the molecular processes underlying pathogenicity and potential management strategies of these economically important fungi.

**Keywords:** Genome announcement, Genome annotation, Fungal pathogen, Coffee pathogen, Eucalyptus canker pathogen, *Fusarium xylarioides*, *Teratosphaeria gauchensis*, *T. zuluensis*

### IMA GENOME-F 11A

#### Genome annotation for *Ceratocystis fimbriata*: an aggressive fungal pathogen of root crops

##### Introduction

The genus *Ceratocystis* includes 41 species of mainly plant pathogenic fungi (Marin-Felix et al. 2017; Holland et al. 2019; Barnes et al. 2018; Liu et al. 2018). The type species, *Ceratocystis fimbriata*, was first described in the USA in 1890 as the causal agent of black rot of *Ipomoea batatas* (Halsted 1890). It shows strong host specificity and does not infect tree hosts, in contrast to many other species in this genus (Baker et al. 2003; Fourie et al. 2018). The pathogen is known in various parts of the USA as well as in Asian countries including Papua New Guinea, China, Japan, Hawaii and New Zealand (Halsted and Fairchild 1891; Lewthwaite et al. 2011; Li et al. 2016). Population genetic analyses of isolates collected in these countries have revealed a very low genetic diversity in all the populations

and *C. fimbriata* thus appears to be a near clonal species (Li et al. 2016; Scruggs et al. 2017).

The sexual fruiting structures in *Ceratocystis* spp., including *C. fimbriata*, are flask-shaped ascomata with long necks exuding sticky ascospore masses at their apices (Wingfield et al. 2017b). *Ceratocystis fimbriata*, along with all other species of *Ceratocystis*, is self-fertile (Halsted 1890; De Beer et al. 2014). The fungus includes isolates that are self-sterile, arising from a phenomenon known as unidirectional mating-type switching, where an isolate of the opposite mating-type is required for sexual reproduction to occur (Harrington and McNew 1997; Witthuhn et al. 2000; Wilken et al. 2014).

Genome sequences represent valuable scientific resources that provide an important source of information needed to understand the biology of organisms. A draft genome sequence of *C. fimbriata* (isolate CBS114723) was published in 2013 (Wilken et al. 2013) and these data were specifically used to characterise the mating-type locus of the isolate (Wilken et al. 2014). Genomes of other species of *Ceratocystis*, including *C. albifundus* (Van der Nest et al. 2014a), *C. eucalypticola* (Wingfield et al. 2015b) and *C. manginecans* (Van der Nest et al. 2014b), have also been published, and form part of a

\* Correspondence: [Brenda.Wingfield@fabi.up.ac.za](mailto:Brenda.Wingfield@fabi.up.ac.za)

†Arista Fourie and Melissa C. Simpson contributed equally to this work.

<sup>1</sup>Department of Biochemistry, Genetics and Microbiology (BGM), Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

Full list of author information is available at the end of the article



larger genome sequencing project. The aim of this study was firstly to improve the genome assembly of *C. fimbriata*, the type species of this genus, and secondly to provide a curated annotation of this genome that can serve as a resource for other species in the genus. Here, we present an improved assembly of *C. fimbriata* with significantly fewer contigs than the previous assembly, and the first annotation of this genome <http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?lvl=0&id=5158>, with RNA data incorporated to improve gene prediction. Genomes of Ceratocystidaceae provide an opportunity to investigate various taxonomic and evolutionary questions of an important group of plant pathogens.

#### Sequenced strains

USA: North Carolina, isolated from *Ipomoea batatas*, Dec, 1998, D. McNew (CMW14799, C1421, CBS114723).

#### Nucleotide sequence accession number

The genomic sequence of *Ceratocystis fimbriata* (CMW14799, CBS114723) has been deposited at DDBJ/EMBL/GenBank under accession no. APWK00000000. The version described in this paper is version APWK03000000. RNA sequencing data has been deposited in the NCBI Short Read Archive under accession number PRJNA67151.

#### Materials and methods

*Ceratocystis fimbriata* was grown on 2% (w/v) MEA (Biolab, Merck, South Africa), supplemented with 150 mg/L streptomycin and 100 µg/L thymine (MEA-ST medium), at room temperature (22–24 °C) for two weeks. DNA extraction was performed using the protocol described by Goodwin et al. (1992) with minor modifications. An RNase treatment step was included after the first phenol/chloroform step and incubated for 1 h at 37 °C. The RNase was then removed with a subsequent phenol/chloroform step. DNA purity was determined using a ND\_1000 Spectrophotometer (Nanodrop, Wilmington, DE). The integrity of the genomic DNA was determined with agarose gel electrophoresis on a 1% gel. DNA quantity was determined with a Qubit® 2.0 Fluorometer (ThermoFisher Scientific, Waltham, USA) following the manufacturer's protocol.

To verify the identity of the isolate and the absence of bacterial contamination, the fungal ITS regions 1 and 2, including the 5.8S rRNA gene (referred collectively as the ITS region) of the ribosomal operon and a portion of the bacterial 16S rRNA were amplified. Amplification and amplicon purification of the ITS region were performed, as described by Fourie et al. (2015), using primers ITS1 and ITS4 (White et al. 1990). The sequence of the PCR product was determined with Sanger

sequencing using an ABI Big DYE Terminator Cycle Sequencing Ready Reaction Kit (Applied BioSystems, Thermo Fisher, California, USA). The ITS sequence was then aligned against that of *C. fimbriata*. The 16S rRNA region was amplified using primers 27F and 1492R (DeLong 1992) and PCR cycler conditions from Beukes et al. (2013), including a positive control bacterial DNA sample. The absence of a PCR band confirmed the absence of bacterial DNA in the sample.

The taxonomic placement of *C. fimbriata* among the Microascales was investigated by means of a phylogenetic analysis of three combined gene regions, the 28S and 60S ribosomal RNA and the Mcm7 (DNA replication licensing factor) gene regions. The sequences of representative isolates of the different genera in this order were obtained from GenBank, as reported by De Beer et al. (2014), and aligned using Muscle alignment in MEGA v.7 (Kumar et al. 2016). A maximum likelihood analysis was performed with the sequence data, using RaxML v.8.2.11 (Stamatakis 2014), model parameters were estimated by the software and 1000 bootstrap replicates were performed to obtain branch support values. The *Graphium* genus was selected as outgroup.

Genome sequencing was performed by FASTERIS SA, Switzerland. All libraries were combined in one lane of an Illumina HiSeq 2500 instrument that produces reads of 125 bp in length. Three libraries were constructed; consisting of a 3000 bp insert mate-pair library and both a 300 bp and 500 bp paired-end library. Illumina adapter sequences were removed from the reads by FASTERIS. Initial quality control (QC) of the sequence reads of each library was performed using FastQC (Andrews 2010). Low quality reads (leading and trailing Phred score < 20) and reads shorter than 20 bp (mate-pair) and 100 bp (paired-end), respectively, were removed from the data using Trimmomatic (Bolger et al. 2014). Analysis with FastQC was repeated, after read trimming, to ensure all reads had a Phred quality score above 20 and that all adapters were removed.

Genome assemblies were performed using Velvet Optimiser v.2.2.5 (Zerbino and Birney 2008) optimising for k-mer values between 61 and 99. Velvet Optimiser indicated a k-mer value of 99 to be optimal. All contigs below 500 bp were discarded. The assembled contigs were subsequently joined into scaffolds by incorporating the mate-pair library data, using the program SSPACE v.2 (Boetzer et al. 2011). Parameters were set to extend contigs, using unmapped reads and five read pairs were required to support the joining of two contigs for the creation of scaffolds. Raw reads were mapped back to the genome to fill in gaps within scaffolds using the software GapFiller v.1.10 (Boetzer and Pirovano 2012). Standard parameters were retained except that a minimum overlap of 100 bp was selected for reads mapped

back to the scaffolds and read trimming was switched off. The pipeline BUSCO (Benchmarking Universal Single-Copy Orthologs; Simão et al. 2015) was used to determine the percentage of conserved Ascomycete (1315 genes) and Sordariomycete (3375 genes) single-copy orthologs present in the genome.

RNA extraction was performed on mycelial and ascomatal tissue harvested from three agar plates after ten days of growth. Liquid nitrogen was used to flash freeze the harvested tissue after which a mortar and pestle was used to grind it to a fine powder. RNA extractions were then performed using the RNeasy® Plant Mini Kit (Qiagen, Limburg, The Netherlands) following the manufacturer's instructions but the RLC buffer was replaced with RLT buffer, and the optional DNase-1 digestion step was included. Quality of the extracted total RNA was evaluated using agarose gel electrophoresis with 2% (w/v) agarose (Seakem). The concentration was then measured using a ND\_1000 Spectrophotometer (Nanodrop, Wilmington, DE). Further quality assessment was performed using the Experion™ automated electrophoresis system (Bio-Rad Laboratories, California, USA). The extracted RNA was enriched for mRNA using Dynabead® mRNA purification kit (Thermo-Fisher Scientific, Waltham, USA) and then subjected to cDNA synthesis, library preparation and sequencing at the Central Analytical Facilities, Stellenbosch University, South Africa, using the Ion Proton Platform and PI™ Chip system (Life Technologies, Carlsbad, CA). Quality of the raw reads was checked using FastQC as described above. Sequences were trimmed using Trimmomatic by removing low quality reads (leading and trailing Phred score < 20), reads below 20 bp in length, the first 20 bp of each read (due to low quality), and any bases over 300 bp in length.

The MAKER genome annotation pipeline v2.31.8 (Cantarel et al. 2008; Holt and Yandell 2011) was used for structural annotation of the genome. RNA-Seq data was incorporated into training of gene predictors and used as gene evidence during all MAKER iterations. The RNA-Seq reads were aligned to the genome with STAR (Dobin et al. 2013) and transcripts were assembled with Cufflinks (Trapnell et al. 2012). The aligned RNA-Seq reads in BAM format were incorporated into the Braker1 pipeline (Hoff et al. 2016) for training of AUGUSTUS v.3.2.1 (Stanke et al. 2004) and GeneMark-ET (Lomsadze et al. 2014). SNAP (Korf 2004) was trained with transcripts assembled from Cufflinks using the est2genome function in MAKER. A species specific repeat library was created using RepeatScout (Price et al. 2005) and RepeatMasker (Smit et al. 1996–2010). This repeat library was used to mask the genome against repetitive elements before performing gene prediction in MAKER. A final MAKER run was conducted using trained parameters from SNAP, AUGUSTUS and GeneMark.

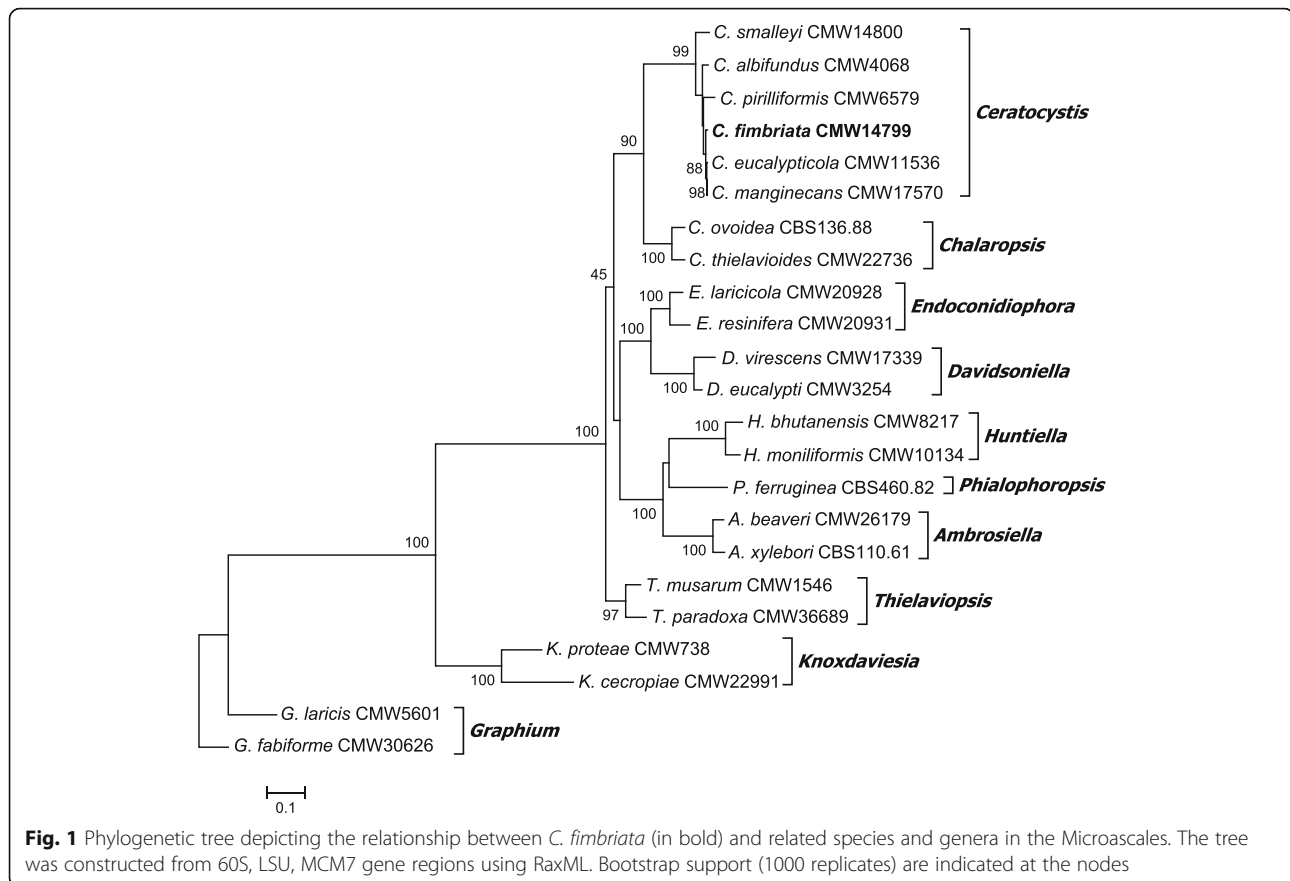
FgeneSH (Solovyev et al. 2006) was run based on parameters pre-optimised for *Neurospora crassa* and the gene models obtained were passed over to MAKER in the final run. Gene models obtained from MAKER were visualised using WebApollo (Lee et al. 2013), along with RNA-seq evidence aligned to the genome, and all gene models were examined and curated, where necessary, based on the RNA-Seq evidence.

Functional annotation of the genes was based on comparison to various databases. A BLASTp analysis was performed for all proteins against the Swissprot database. Proteins were also compared to the Protein family (Pfam) database using InterProScan v.5.24 (Jones et al. 2014). SignalP v.4.0 (Petersen et al. 2011) was used to predict secretion signals and Phobius v.1 (Käll et al. 2004) was used to predict the presence of transmembrane domains. To predict the number of genes present in internal clusters, the genes involved in secondary metabolite production were determined using the anti-SMASH v.3 software available online (Weber et al. 2015). The functional annotations predicted for each protein were added to the gff file using the software ANNIE (Ooi et al. 2009) and GAG (Hall et al. 2014).

### Results and discussion

The paired-end and mate-pair sequencing generated approximately 42.5 million raw reads with an average length of 166 bp. Trimmed reads were assembled into 399 scaffolds, ranging in size from 500 bp to 516 595 bp, with an average read coverage of 630 and N50 value of 173 733 bp. The *C. fimbriata* genome was 30 159 98 bp in size with a GC content of 45.6%. There were 7728 predicted genes, of which 7266 (94%) were protein-coding genes, 105 (1.36%) were rRNA genes and 348 (4.5%) were tRNA genes. In total, 62% of the genes could be annotated with a known function. The *C. fimbriata* genome contained 98% Ascomycete and 90% Sordariomycete completed BUSCO gene models. The taxonomic placement of *C. fimbriata* among the Microascales is illustrated in Fig. 1. The sequence alignments were submitted to Treebase (24031).

This study showed that *C. fimbriata* has far fewer genes than other fungal plant pathogens with similar genome sizes. For example, the head blight pathogen *Fusarium graminearum* (genome size: 36.5 Mb) has approximately 14 164 genes (King et al. 2015), the pine needle blight pathogen *Dothistroma septosporum* (genome size: 31.2 Mb) has 12 580 genes (De Wit et al. 2012) and the wilt pathogen *Verticillium dahliae* (genome size: 33.8 Mb) has 10 535 genes (Klosterman et al. 2011). The number of genes in *C. fimbriata* does, however, correspond with those of other *Ceratocystis* species (Van der Nest et al. 2014a; Van der Nest et al. 2014b; Wingfield et al. 2015b; Wingfield et al. 2016b). There are currently eight genomes



publicly available for species of *Ceratocystis* and 18 in the Ceratocystidaceae (Wilken et al. 2013; Van der Nest et al. 2014a; Van der Nest et al. 2014b; Van der Nest et al. 2015; Wingfield et al. 2015a; Wingfield et al. 2015b; Wingfield et al. 2016a; Wingfield et al. 2017; Molano et al. 2018; Vanderpool et al. 2018; Wingfield et al. 2018a). The annotation of the *C. fimbriata* genome sequence presented in this study is the first manually curated genome including additional RNA evidence, for any species in the Ceratocystidaceae. This annotation will provide the means to improve annotations for other species of *Ceratocystis* and provides an improved assembly that can be used for future comparative studies.

**Authors:** A. Fourie\*, M.C. Simpson\*, T.A. Duong, I. Barnes, M.P.A. Coetzee, M.A. van der Nest, M.J. Wingfield and B.D. Wingfield.

\*Contact: [melissa.simpson@fabu.up.ac.za](mailto:melissa.simpson@fabu.up.ac.za) or [arista.fourie@fabu.up.ac.za](mailto:arista.fourie@fabu.up.ac.za) (authors contributed equally)

## IMA GENOME-F 11B

### Draft genome sequence of *Fusarium xylarioides*

#### Introduction

*Fusarium xylarioides* Steyaert (1948) is a soilborne fungal pathogen that causes coffee wilt disease (CWD) in many coffee growing regions in Africa (Rutherford 2006). The

sexual stage, previously referred to as *Gibberella xylarioides* (Heim 1950), is readily observed in CWD infested trees (Ploetz 2006). This heterothallic fungus is a member of the *F. fujikuroi* species complex (FFSC), a group of phylogenetic species that infect a number of important crops (O'Donnell et al. 1998; O'Donnell et al. 2000). Previous studies have reported the presence of two genetically and biologically distinct forms of this pathogen (Geiser et al. 2005; Lepoint et al. 2005), but their taxonomy has not yet been conclusively resolved. The availability of a complete genome sequence will, therefore, serve as the starting point for addressing the taxonomic confusion about this species in the literature. Also, more genetic data on this species will enable studies on its biology and evolution.

#### Sequenced strain

Uganda: Iganga District: Isolated from *Coffea canephora*, Dec. 2000, D.M. Geiser (KSU 18978 = FRC L-0394 = CMW 53787 – living culture).

#### Nucleotide sequence accession number

The whole genome shotgun sequencing project of *Fusarium xylarioides* KSU 18978 (FRC-L0394 = CMW 53787) has been deposited at DDBJ/ENA/GenBank under the

accession no. SRZU00000000. The version described in this paper is version SRZU01000000.

### Materials and methods

*Fusarium xylarioides* KSU 18978 (CMW 53787) was obtained from the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria. The isolate was grown in 40 ml potato dextrose broth (20% potato dextrose broth w/v) and incubated on an orbital shaker (135 rpm) for 2 d at room temperature (22–25 °C). Genomic DNA was extracted following the method of Duong et al. (2013). One pair-end library of 250 bp read length and 550 bp insert size was prepared and sequenced using the Illumina HiSeq 2500 platform. Quality control of pair-end reads received and adapter trimming was performed in the program Trimmomatic v. 0.36 (Bolger et al. 2014). Genome assembly was performed from trimmed reads using SPAdes v. 3.10 (Bankevich et al. 2012) and further scaffolding was performed using SSPACE Standard v. 3.0 (Boetzer et al. 2011). The genome quality and completeness were evaluated with BUSCO v. 2.0 (Simão et al. 2015) using the dataset for *Sordariomycetes*. The number of protein-coding genes encoded by the genome was evaluated with the program AUGUSTUS v. 3.2.2 (Stanke et al. 2006) using *Fusarium graminearum* as a species model. The taxonomic identity of the sequenced genome was confirmed by Maximum likelihood (ML) analysis of authenticated sequences using MEGA version X (Kumar et al. 2018).

### Results and discussion

The assembled draft genome of *Fusarium xylarioides* was estimated to be 55.24 Mb with a coverage of 61x, corresponding to 424 scaffolds larger than 500 bp with an N50 value of 250 204 bp, and an average GC content of 43.4%. Among the FFSC species for which genomes sequences are available, *F. xylarioides* thus has the largest predicted genome size, i.e., the genomes of *F. circinatum* is 43.43 Mb, *F. temperatum* 45.46 Mb, *F. fracticaudum* 46.29 Mb, *F. pininemorale* 47.83 Mb, and *F. nygamai* is 51.61 Mb (Wingfield et al. 2015a, 2015b, 2017, 2018a). The predicted size of our sequenced genome was similar to that of another *F. xylarioides* strain K1 (55.11 Mb), whose genome sequence has recently been made available on NCBI under nucleotide accession number GCA\_004329255.

Based on BUSCO analysis, genome completeness was 99.0% (C:3687 [S:3680, D:7], F:26, M:12, n:3725), suggesting that the assembly covers the majority of the organism's gene content. AUGUSTUS predicted that the assembly encodes 14 588 open reading frames (ORFs) and this was in agreement with data from other members of the FFSC (Wingfield et al. 2015a, 2015b, 2017, 2018a).

Phylogenetic analysis based on partial gene sequences of  $\beta$ -tubulin and TEF-1 $\alpha$  confirmed the sequenced genome to be of *F. xylarioides* (Fig. 2). *F. xylarioides* and *F. udum* are the only two members of the FFSC that cause true vascular wilt diseases (Geiser et al. 2005). Therefore, the increase in available whole genome data will not only allow for taxonomic re-evaluation of *F. xylarioides*, but will also enable comparative genomics studies to better understand the biology and evolution of members of the FFSC.

**Authors:** V.S. Bushula-Njah\*, T.A. Duong, D.M. Geiser, E.T. Steenkamp, and B.D. Wingfield.

\*Contact: [Vuyiswa.bushula@fabi.up.ac.za](mailto:Vuyiswa.bushula@fabi.up.ac.za)

## IMA GENOME-F 11C

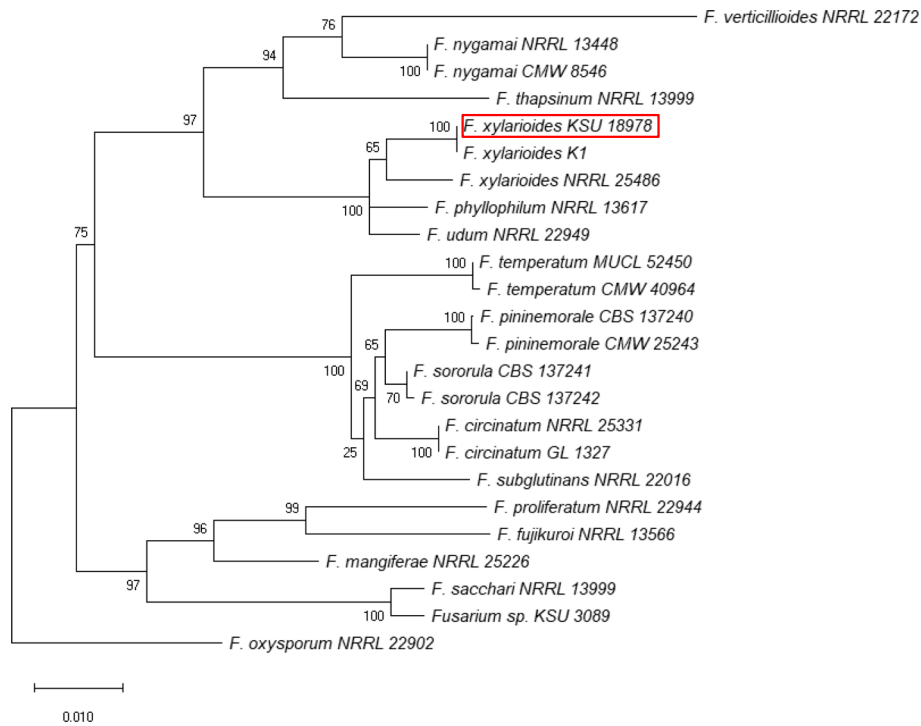
### Draft genome sequences of *Teratosphaeria gauchensis* and *T. zuluensis*, causal agents of *Teratosphaeria* stem canker

#### Introduction

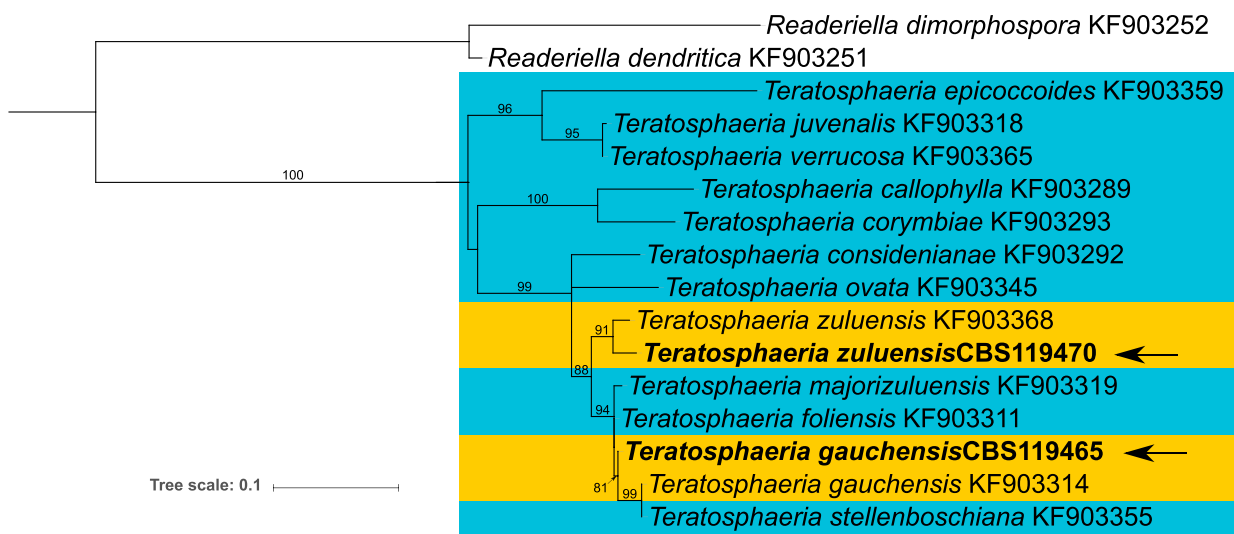
*Teratosphaeria* stem canker (previously known as Coniothyrium canker) is a fungal disease of *Eucalyptus* trees planted outside of their native range for the production of wood and non-wood products. Since its discovery in a South African plantation in the late 1980's (Wingfield et al. 1996), this disease has emerged in *Eucalyptus* plantations in 14 additional countries where tropical or subtropical climates predominate (Aylward et al. 2019). Infection is characterized by necrotic lesions that grow and eventually merge to form large gum-filled, bleeding cankers (Wingfield et al. 1996). Gum-stained wood is unsuitable for timber production and lesions hinder debarking so that pulping is affected (Old et al. 2003).

*Teratosphaeria gauchensis* and *T. zuluensis*, Dothideomycete fungi (Capnodiales, Teratosphaeriaceae), cause *Teratosphaeria* stem canker independently (Cortinas et al. 2006). These two species are unique in that they are the only known stem canker pathogens in a genus predominantly associated with the leaves of *Eucalyptus* species (Fig. 3; Quaadvlieg et al. 2014). Until recently, they were thought to have distinct geographic distributions, but in 2014 both were identified in a Ugandan plantation (Jimu et al. 2014). Together with the 2015 discovery of *T. gauchensis* in southern Africa (Jimu et al. 2015), this suggests that concurrent infections are bound to occur in plantations of central and southern Africa (Aylward et al. 2019).

We present the genome sequences of the ex-holotype of *T. gauchensis* (CBS 119465) and the ex-type of *T. zuluensis* (CBS 119470). These genomes will enable exploration of the intriguing case of two different species causing identical disease symptoms. Future studies will focus on the eucalypt stem specificity of these pathogens in contrast to the widespread eucalypt leaf association of other *Teratosphaeria* species.



**Fig. 2** Maximum likelihood (ML) tree based on partial gene sequences of  $\beta$ -tubulin and translation elongation factor 1- $\alpha$ . Sequence alignments were assembled with MAFFT version 7 (Katoh and Standley 2013). The program MEGA X version 7 (Kumar et al. 2018) was used to estimate the best-fit substitution model (TN93) with a discrete gamma distribution (+G). The same program was used for ML phylogenetic analysis and percentage bootstrap support (1000 replications) values are indicated at branch nodes



**Fig. 3** Maximum Likelihood phylogeny of the *Teratosphaeria* stem canker pathogens and related species based on the Elongation Factor (EF1- $\alpha$ ) gene. Leaf-associated species are highlighted in teal, whereas the stem canker pathogens are yellow. The genome isolates described in this study are indicated with arrows. GeneBank accessions are shown after the species name

### Sequenced strain

*Teratosphaeria gauchensis*: Uruguay: *La Juanita*: isol. Stem cankers on *Eucalyptus grandis*, Feb. 2005, M.J. Wingfield (CBS 119465 = CMW 50181- culture, PREM 62331 – dried culture).

*Teratosphaeria zuluensis*: South Africa: *KwaZulu-Natal Province* (Kwambonambi plantation): isol. Stem cankers on *Eucalyptus grandis*, Feb. 2005, M.J. Wingfield (CBS 119470 = CMW 50183 – culture, PREM 62332 – dried culture).

### Nucleotide accession number

The genomic sequence data of *T. gauchensis* and *T. zuluensis* have been deposited at DDJ/EMBL/GenBank under the accessions VCMR00000000 (*T. gauchensis*) and VCMQ00000000 (*T. zuluensis*). This paper describes the first versions of these genomes.

### Material and methods

Fungi were cultured on 2% Malt Extract Agar (Merck, Wadeville, South Africa) at 25 °C for approximately 2 weeks. DNA was extracted as described for *T. destructans* (Wingfield et al. 2018b). The quality of the extracted DNA was estimated from the absorbance curve and the 260/280 and 260/230 absorbance ratios determined by a NanoDrop ND-1000 spectrophotometer (ThermoFisher Scientific, Wilmington, USA). DNA concentrations were determined with a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, California).

*Teratosphaeria gauchensis* (CBS 119465) was sequenced at Macrogen (Seoul, Korea) using a single PacBio Sequel SMRT Cell as well as a portion of an Illumina HiSeq 2500 lane at Inqaba Biotec (Pretoria, South Africa). The single paired-end Illumina library had an insert size of 550 bp and a target read length of 250 bp. A hybrid assembly was computed by trimming the raw Illumina sequence reads with Trimmomatic 0.38 (Bolger et al. 2014) and using these in LorDEC 0.6 (Salmela and Rivals 2014) to correct the PacBio reads. The corrected PacBio reads were assembled with Canu 1.7.1 (Koren et al. 2017). An Illumina assembly was subsequently constructed in SPAdes 3.10.1 (Bankevich et al. 2012), using k-mer values of 21, 33, 55, and 77 and applying the Canu assembly as “trusted reads”. A final error correction of the hybrid assembly was performed with Pilon 1.22 (Walker et al. 2014).

*Teratosphaeria zuluensis* (CBS 119470) was sequenced with the Ion 520™ & Ion 530™ ExT Kit and an Ion 530™ Chip (ThermoFisher Scientific, MA) at the Central Analytical Facility (CAF, Stellenbosch University, South Africa). Assembly of reads was done with SPAdes 3.10.1 (Bankevich et al. 2012) using k-mer values 21, 33, 55, 77, 99, and 127. For both *T. gauchensis* and *T. zuluensis*, genome completeness was estimated with Benchmarking

Universal Single-Copy Orthologs (BUSCO) 2.0.1 (Simão et al. 2015) using the “Ascomycota odb9” dataset.

Repeats within each genome were identified with RepeatScout and masked with RepeatMasker Open-4.0.7 (<http://www.repeatmasker.org>). Annotation was performed with the MAKER 2.31.10 pipeline (Campbell et al. 2014) using pre-trained de novo gene predictors AUGUSTUS 3.3 (Stanke et al. 2006), GeneMark-ES Suite 4.35 (<http://exon.gatech.edu/GeneMark/>) and SNAP 2006-07-28 (Korf 2004). External EST and protein evidence from 15 other Capnodiales species were included in the annotation pipeline.

The Elongation Factor genes (EF1- $\alpha$ ) of the two stem canker pathogens were extracted from their genomes and used along with the EF1- $\alpha$  genes of related species (Quaedvlieg et al. 2014) for phylogenetic analysis. The Maximum Likelihood tree was computed with the PhyML+SMS “one-click” method on NGPhylogeny.fr (Lemoine et al. 2019). This protocol employs MAFFT (Katoh et al. 2017) for multiple alignment, BMGE (Criscuolo and Gribaldo 2010) for alignment curation and the aLRT SH-like method (Anisimova and Gascuel 2006) for calculating bootstrap support.

### Results and discussion

*Teratosphaeria gauchensis* was assembled into 53 contigs larger than 1 kb, with an N50 = 1.44 Mb and L50 = 8. This equated to a total assembly size of 30.27 Mb with a GC content of 45.6%. The PacBio reads provided an estimated genome coverage of 124x with an additional 160x coverage obtained from the Illumina data. The *T. zuluensis* assembly yielded 86 contigs above 1 kb (N50 = 1.00 Mb, L50 = 12), an estimated genome size of 28.71 Mb, 44.5% GC and estimated 155x genome coverage. The search for 1315 Ascomycota ortholog proteins identified 97.6% (1285) single-copy orthologs in *T. gauchensis* and 96.9% (1274) in *T. zuluensis*, indicating that these genomes are ca. 97% complete. This presents a higher level of completeness than the 84.5% of *T. destructans*, the only other published *Teratosphaeria* genome (Wingfield et al. 2018b).

A similar number of genes were predicted in *T. gauchensis* and *T. zuluensis* at 11699 and 11520, respectively. Of these, 9304 predictions in *T. gauchensis* and 9457 in *T. zuluensis* were supported by external evidence. Both genomes had a low repeat content of 2.2%, much lower than the ca.17% estimated for *T. destructans* (GenBank RIBY01000000). Since the genomes of these stem canker pathogen species, as well as *T. destructans*, were sequenced with long-read technologies, we are confident that this is not an underestimate, but reflects the true repeat content in these genomes.

The EF1- $\alpha$  gene tree (Fig. 3) illustrates the relationship of *T. zuluensis* and *T. gauchensis* to leaf-associated *Teratosphaeria* species. This phylogeny, as well as others

(Aylward et al. 2019; Quaedvlieg et al. 2014), suggests that the association with *Eucalyptus* stem cankers evolved more than once. The *T. zuluensis* genome isolate CBS119470 groups with the ex-epitype strain CBS120301 (KF903368), however, the *T. gauchensis* strain sequenced in this study (CBS119465) groups sister to a clade that contains an ex-type strain (CBS120304; KF903314) of *T. gauchensis* and *T. stellenboschiana*. A similar phylogenetic relationship is apparent with both the ITS and  $\beta$ -tubulin gene regions as well as with concatenated gene trees (data not shown). *Teratosphaeria gauchensis* is known to represent a species complex displaying statistical support for subgroups within the species (Aylward et al. 2019; Silva et al. 2015). Future studies should take the taxonomic position of this genome isolate into consideration when interpreting genomic data.

Several other fungal groups affect *Eucalyptus* stems, including species in the genera *Chrysosporthe* (Gryzenhout et al. 2004), *Cytospora* (Adams et al. 2006), *Holocryphia* (Van der Westhuizen et al. 1993) and *Neofusicoccum* (Slippers et al. 2009). Of these, genomes of *Eucalyptus*-specific isolates are available only for *Chrysosporthe austroafricana*, *C. cubensis* and *C. deuterocubensis*, the causal agents of Cryphonectria canker (Wingfield et al. 2015a, 2015b). *Teratosphaeria gauchensis* and *T. zuluensis*, therefore, represent the second group of *Eucalyptus* stem canker pathogens to be sequenced. In future, comparative genomics projects that include various *Eucalyptus* stem pathogens as well as closely related species associated with other plant organs (e.g. *T. destructans*), may reveal fungal characteristics that enable stem pathogenicity.

**Authors:** J. Aylward, B.D. Wingfield, L.L. Dreyer, F. Roets, C.J. van Heerden, M.J. Wingfield.

**Contact:** [Janneke.Aylward@fabi.up.ac.za](mailto:Janneke.Aylward@fabi.up.ac.za)

#### Abbreviations

aLRT SH: Approximate Likelihood Ratio Test for branches; ANNIE: ANnotation Information Extractor; AUGUSTUS: Software for gene prediction; BLAST: Basic Local Alignment Search Tool; BLASTp: Basic Local Alignment Search Tool for protein sequences; BMGE: Block Mapping and Gathering with Entropy; BUSCO: Benchmarking Universal Single-Copy Orthologs; Canu: A hierarchical assembly pipeline; cDNA: Complementary Deoxyribose Nucleic Acid; CWD: Coffee wilt disease; DDBJ: DNA Data Bank of Japan; DNA: Deoxyribose Nucleic Acid; DST: Department of Science and Technology of South Africa; EF1- $\alpha$ : Elongation Factor 1 alpha gene; EMBL: Nucleotide Sequence Data Library; ENA: European Nucleotide Archive; FABI: Forestry and Agricultural Biotechnology Institute; FFSC: *F. fujikuroi* species complex; GenBank: NIH genetic sequence database; GeneMark: Gene prediction software; ITS Internal: Transcribed Spacer; LoRDEC: Hybrid error correction program; MAFFT: Multiple Alignment using Fast Fourier Transform; MEGA: Molecular Evolutionary Genetics Analysis; NCBI: National Center for Biotechnology Information; NRF: National Research Foundation (NRF); ORF(s): Open reading frame(s); PhyML: Phylogeny software based on maximum-likelihood principle; PREM: South African National Collection of Fungi; RNA: Ribose Nucleic Acid; rRNA: Ribosomal Ribo Nucleic Acid; SARCHI: South African Research Chairs Initiative; SMS: Smart Model Selection in PhyML; SNAP: Web-based tool for identification and annotation of proxy Single Nucleotide Polymorphisms; SPAdes: St. Petersburg genome assembler; SSPACE: SSAKE-based Scaffolding of Pre-Assembled Contigs after Extension program; tRNA: Ribo Nucleic Acid; USA: United States of America

#### Acknowledgements

For the *Ceratocystis* annotation project Alisa Postma-Smidt is acknowledged for providing assistance with the genome assembly and read mapping. Dr. Fourie Joubert and Alisa Postma-Smidt also provided assistance with genome annotation and curation. We acknowledge a team of students and staff members in FABI that assisted in the genome jamboree for manual curation of the gene models: Tayo Adenigba, Dr. Dave Berger, Wilma Janine, Claire Jordaan, Kayleigh Maier, Mandy Messal, Ramesh Moolam, Lizel Potgieter, Catherine Tatham, Conrad Trollip, Salome van Hoepen, Bradley Vincent, Dr. Markus Wilken and Andi Wilson.

#### Adherence to national and international regulations

Not applicable to the specific isolates used in this manuscript. All isolates are maintained in culture collections as per government regulations and quarantine specifications.

#### Authors' contributions

The manuscript is comprised of three sections, 11A, 11B, 11C. Each are authored separately and there is a specific contact person. In each case the contact person has taken the lead in writing and compiling the manuscript. The other authors have been involved in the idea behind the manuscript, editing and analysis of the data. All authors read and approved the final manuscript.

11A Genome annotation for *Ceratocystis fimbriata*: an aggressive fungal pathogen of root crops

Authors: A. Fourie\*, M.C. Simpson\*, T.A. Duong, I. Barnes, M.P.A. Coetzee, M.A. van der Nest, M.J. Wingfield and B.D. Wingfield

Contact: [melissa.simpson@fabi.up.ac.za](mailto:melissa.simpson@fabi.up.ac.za) or [arista.fourie@fabi.up.ac.za](mailto:arista.fourie@fabi.up.ac.za)

(\*authors contributed equally)

11B Draft genome sequence of *Fusarium xyloarioides*

Authors: V.S. Bushula-Njah, T.A. Duong, D.M. Geiser, E.T. Steenkamp, and B.D. Wingfield

Contact: [Vuyiswa.bushula@fabi.up.ac.za](mailto:Vuyiswa.bushula@fabi.up.ac.za)

11C Draft genome sequences of *Teratosphaeria gauchensis* and *T. zuluensis*, causal agents of Teratosphaeria stem canker

Authors: J. Aylward, B.D. Wingfield, L.L. Dreyer, F. Roets, C.J. van Heerden, M.J. Wingfield

Contact: [Janneke.Aylward@fabi.up.ac.za](mailto:Janneke.Aylward@fabi.up.ac.za)

#### Authors' information

No additional information provided.

#### Funding

We acknowledge the National Research Foundation (NRF) and Department of Science and Technology (DST) of South Africa for providing scholarship funding for students as well as members of the Tree Protection Cooperative Programme at the Forestry and Agricultural Biotechnology Institute for providing financial support for this project. These projects from part of the DST/NRF Centre of Excellence in Tree Health Biotechnology and the SARCHI in Fungal Genomics.

#### Availability of data and materials

All data and material is available the relevant details (data banks, culture collections and herbaria) are given in the manuscript.

#### Ethics approval and consent to participate

Not applicable, no humans, human subjects nor data were used in this manuscript.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Biochemistry, Genetics and Microbiology (BGM), Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Hatfield 0028, South Africa. <sup>2</sup>Department of Botany and Zoology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. <sup>3</sup>Fusarium Research Center, Department of Plant Pathology and



Environmental Microbiology, 121 Buckhout Lab, University Park, State College, PA 16802, USA. <sup>4</sup>Department of Conservation Ecology and Entomology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. <sup>5</sup>Central Analytical Facilities, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. <sup>6</sup>Biotechnology Platform, Agricultural Research Council, Private Bag X05, Onderstepoort 0002, South Africa.

Received: 8 June 2019 Accepted: 1 July 2019

Published online: 13 September 2019

## References

- Adams G, Roux J, Wingfield M (2006) *Cytospora* species (Ascomycota, Diaporthales, Valsaceae): introduced and native pathogens of trees in South Africa. *Australas Plant Pathol* 35:521–548
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* 55:539–552
- Aylward J, Roets F, Dreyer LL, Wingfield MJ (2019) *Teratosphaeria* stem canker of *Eucalyptus*: two pathogens, one devastating disease. *Mol Plant Pathol* 20:8–19
- Baker CJ, Harrington TC, Krauss U, Alfenas AC (2003) Genetic variability and host specialization in the Latin American clade of *Ceratocystis fimbriata*. *Phytopathology* 93:1274–1284
- Bankveich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477
- Barnes I, Fourie A, Wingfield M, Harrington T, McNew D et al (2018) New *Ceratocystis* species associated with rapid death of *Metrosideros polymorpha* in Hawai'i. *Persoonia Mol Phylogeny Evol Fungi* 40:154–181
- Beukes CW, Venter SN, Law IJ, Phalane FL, Steenkamp ET (2013) South African papilionoid legumes are nodulated by diverse *Burkholderia* with unique nodulation and nitrogen-fixation loci. *PLoS One* 8:e68406
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579
- Boetzer M, Pirovano W (2012) Toward almost closed genomes with GapFiller. *Genome Biol* 13:R56
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Campbell MS, Holt C, Moore B, Yandell M (2014) Genome annotation and curation using MAKER and MAKER-P. *Curr Protoc Bioinformatics* 48:4.11.11–4.11.39
- Cantarel BL, Korf I, Robb SMC, Parra G, Ross E et al (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 18:188–196
- Cortinas MN, Crous PW, Wingfield BD, Wingfield MJ (2006) Multi-gene phylogenies and phenotypic characters distinguish two species within the *Colletogloeopsis zuluensis* complex associated with *Eucalyptus* stem cankers. *Stud Mycol* 55:133–146
- Crisuolo A, Gribaldo S (2010) BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10:210
- De Beer ZW, Duong TA, Barnes I, Wingfield BD, Wingfield MJ (2014) Redefining *Ceratocystis* and allied genera. *Stud Mycol* 79:187–219
- De Wit PJGM, Van Der Burgt A, Ökmen B, Stergiopoulos I, Abd-Elsalam KA et al (2012) The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. *PLoS Genet* 8:e1003088
- DeLong EF (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci* 89:5685–5689
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21
- Duong TA, De Beer ZW, Wingfield BD, Wingfield MJ (2013) Characterization of the mating-type genes in *Leptographium procerum* and *Leptographium profanum*. *Fungal Biology* 117:411–421
- Fourie A, Wingfield MJ, Wingfield BD, Barnes I (2015) Molecular markers delimit cryptic species in *Ceratocystis sensu stricto*. *Mycol Prog* 14:1020
- Fourie A, Wingfield MJ, Wingfield BD, van der Nest MA, Loots MT et al (2018) Inheritance of phenotypic traits in the progeny of a *Ceratocystis* interspecific cross. *Fungal Biology* 122:717–729
- Geiser DM, Lewis-Ivey ML, Hakiza G, Juba JH, Miller SA (2005) *Gibberella xylarioides* (anamorph: *Fusarium xylarioides*), a causative agent of coffee wilt disease in Africa, is a previously unrecognized member of the *G. fujikuroi* species complex. *Mycologia* 97:191–201
- Goodwin SB, Drenth A, Fry WE (1992) Cloning and genetic analyses of two highly polymorphic, moderately repetitive nuclear DNAs from *Phytophthora infestans*. *Curr Genet* 22:107–115
- Gryzenhout M, Myburg H, Van der Merwe NA, Wingfield BD, Wingfield MJ (2004) *Chrysosporthe*, a new genus to accommodate *Cryphonectria cubensis*. *Stud Mycol* 50:119–142
- Hall B, DeRego T, Geib S (2014) GAG: the genome annotation generator. (Version 1.0) [Software]. Available from <http://genomeannotation.github.io/GAG>
- Halsted BD (1890) Some fungal disease of the sweet potato. *Agric College Exp Station Bull* 76:1–32
- Halsted BD, Fairchild DG (1891) Sweet-potato black rot. *J Mycol* 7:1–11
- Harrington TC, McNew DL (1997) Self-fertility and uni-directional mating-type switching in *Ceratocystis coerulescens*, a filamentous ascomycete. *Curr Genet* 32:52–59
- Heim R (1950) La carbunculariose du Caféier. *Suppl Colon a la Revue de Mycol* 15:89–98
- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M (2016) BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32:767–769
- Holland LA, DP Lawrence, MT Nouri, R Travadon, TC Harrington, FP Trouillas, (2019) Taxonomic Revision and Multi-locus Phylogeny of the North American Clade of *Ceratocystis*. *Fungal Systematics and Evolution* 3 (1):135–156.
- Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491
- Jimu L, Wingfield MJ, Mwenje E, Roux J (2014) First report of *Teratosphaeria zuluensis* causing stem canker of *Eucalyptus grandis* in Uganda. *For Pathol* 44:242–245
- Jimu L, Wingfield MJ, Mwenje E, Roux J (2015) Diseases on *Eucalyptus* species in Zimbabwean plantations and woodlots. *South Forests* 77:221–230
- Jones P, Binns D, Chang H-Y, Fraser M, Li W et al (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240
- Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* 338:1027–1036
- Katoh K, Rozewicki J, Yamada KD (2017) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. <https://doi.org/10.1093/bib/bbx108>.
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780
- King R, Urban M, Hammond-Kosack MCU, Hassani-Pak K, Hammond-Kosack KE (2015) The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* 16:544
- Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE et al (2011) Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog* 7:e1002137
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736
- Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* 5:59
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874
- Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP et al (2013) Web Apollo: a web-based genomic annotation editing platform. *Genome Biol* 14:R93
- Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S et al (2019) NGPhylogeny. Fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res* 47(W1):W260–W265
- Lepoint PCE, Munaut FTJ, Maraitte HMM (2005) *Gibberella xylarioides* Sensu Lato from *Coffea canephora*: a new mating population in the *Gibberella fujikuroi* species complex. *Appl Environ Microbiol* 71:8466–8471
- Lewthwaite SL, Wright PJ, Triggs CM (2011) Sweetpotato cultivar susceptibility to infection by *Ceratocystis fimbriata*. *New Zealand Plant Prot* 64:1–6
- Li Q, Harrington TC, McNew D, Li J, Huang Q et al (2016) Genetic bottlenecks for two populations of *Ceratocystis fimbriata* on sweet potato and pomegranate in China. *Plant Dis* 100:2266–2274

- Liu F, Barnes I, Roux J, Wingfield MJ, Chen S (2018) Molecular phylogenetics and microsatellite analysis reveal a new pathogenic *Ceratocystis* species in the Asian-Australian clade. *Plant Pathol* 67:1097–1113
- Lomsadze A, Burns PD, Borodovsky M (2014) Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res* 42:e119
- Marin-Felix Y, Groenewald JZ, Cai L, Chen Q, Marincowitz S et al (2017) Genera of phytopathogenic fungi: GOPHY 1. *Stud Mycol* 86:99–216
- Molano EPL, Cabrera OG, Jose J, do Nascimento LC, Carazzolle MF et al (2018) *Ceratocystis cacaofunesta* genome analysis reveals a large expansion of extracellular phosphatidylinositol-specific phospholipase-C genes (PI-PLC). *BMC Genomics* 19:58–58
- O'Donnell K, Cigelnik E, Nirenberg HI (1998) Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia* 90:465–493
- O'Donnell K, Nirenberg HI, Aoki T, Cigelnik E (2000) A multigene phylogeny of the *Gibberella fujikuroi* species complex: detection of additional phylogenetically distinct species. *Mycoscience* 41:61–78
- Old KM, Wingfield MJ, Yuan ZQ (2003) A manual of diseases of eucalypts in South-East Asia. CIFOR, Jakarta
- Ooi HS, Kwo CY, Wildpaner M, Sirota FL, Eisenhaber B et al (2009) ANNIE: integrated de novo protein sequence annotation. *Nucleic Acids Res* 37:W435–W440
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785
- Ploetz RC (2006) *Fusarium*-induced diseases of tropical, perennial crops. *Phytopathology* 96:648–652
- Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21:i351–i358
- Quaedvlieg W, Binder M, Groenewald JZ, Summerell BA, Carnegie AJ, Burgess TI et al (2014) Introducing the consolidated species concept to resolve species in the Teratosphaeriaceae. *Persoonia* 33:1–40
- Rutherford MA (2006) Current knowledge of coffee wilt disease, a major constraint to coffee production in Africa. *Phytopathology* 96:663–666
- Salmela L, Rivals E (2014) LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30:3506–3514
- Scruggs AC, Basaiah T, Adams ML, Quesada-Ocampo LM (2017) Genetic diversity, fungicide sensitivity, and host resistance to *Ceratocystis fimbriata* infecting Sweetpotato in North Carolina. *Plant Dis* 101:994–1001
- Silva MRC, Diogo E, Bragança H, Machado H, Phillips AJL (2015) Teratosphaeria gauchensis associated with trunk, stem and foliar lesions of Eucalyptus globulus in Portugal. *For Pathol* 45:224–234
- Simão 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
- Slippers B, Burgess T, Pavlic D, Ahumada R, Maleme H, Mohali S et al (2009) A diverse assemblage of Botryosphaeriaceae infect Eucalyptus in native and non-native environments. *South Forests* 71:101–110
- Smit AFA, Hubley R, Green P (1996–2010) RepeatMasker Open-3.0. <http://www.repeatmasker.org>
- Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol* 7:1–12
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* 34:W435–W439
- Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 32:W309–W312
- Steyaert RL (1948) Contribution à l'étude des parasites des végétaux du Congo Belge. *R Bot Soc Belgium Ser 2* 30:11–58
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc* 7:562–578
- Van der Nest MA, Beirn LA, Crouch JA, Demers JE, de Beer ZW et al (2014a) Draft genomes of *Amanita jacksonii*, *Ceratocystis albifundus*, *Fusarium circinatum*, *Huntliella omanensis*, *Leptographium procerum*, *Rutstroemia sydowiana*, and *Sclerotinia echinophila*. *IMA Fungus* 5:472–485
- Van der Nest MA, Bihon W, de Vos L, Naidoo K, Roodt D et al (2014b) Draft genome sequences of *Diplodia sapinea*, *Ceratocystis manginecans*, and *Ceratocystis moniliformis*. *IMA Fungus* 5:135–140
- Van der Nest MA, Steenkamp ET, McTaggart AR, Trollip C, Godlonton T et al (2015) Saprophytic and pathogenic fungi in the Ceratocystidaceae differ in their ability to metabolize plant-derived sucrose. *BMC Evol Biol* 15:273
- Van der Westhuizen I, Wingfield M, Kemp G, Swart W (1993) First report of the canker pathogen *Endothia gyrosa* on Eucalyptus in South Africa. *Plant Pathol* 42:661–663
- Vanderpool D, Bracewell RR, McCutcheon JP (2018) Know your farmer: ancient origins and multiple independent domestications of ambrosia beetle fungal cultivars. *Mol Ecol* 27:2077–2094
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S et al (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963
- Weber T, Blin K, Duddela S, Krug D, Kim HUK et al (2015) antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43:W237–W243
- White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a sequencing guide to methods and applications. Academic Press Inc., San Diego, pp 315–322
- Wilken PM, Steenkamp ET, Wingfield MJ, De Beer ZW, Wingfield BD (2013) Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. *IMA Fungus* 4:357–358
- Wilken PM, Steenkamp ET, Wingfield MJ, De Beer ZW, Wingfield BD (2014) DNA loss at the *Ceratocystis fimbriata* mating locus results in self-sterility. *PLoS One* 9:e92180
- Wingfield BD, Ades PK, Al-Naemi FA, Beirn LA, Bihon W et al (2015a) Draft genome sequences of *Chrysosporthe austroafricana*, *Diplodia scrobiculata*, *Fusarium nygamai*, *Leptographium lundbergii*, *Limonomyces culmigenus*, *Stagonosporopsis tanacetii*, and *Thielaviopsis punctulata*. *IMA Fungus* 6:233–248
- Wingfield BD, Amblar JM, Coetzee M, De Beer ZW, Duong TA et al (2016a) Draft genome sequences of *Armillaria fuscipes*, *Ceratocystiopsis minuta*, *Ceratocystis adiposa*, *Endoconidiophora laricicola*, *E. polonica* and *Penicillium frei* DAOMC 242723. *IMA Fungus* 7:217–227
- Wingfield BD, Barnes I, de Beer ZW, De Vos L, Duong TA et al (2015b) Draft genome sequences of *Ceratocystis eucalypticola*, *Chrysosporthe cubensis*, *C. deuterocubensis*, *Davidsoniella virescens*, *Fusarium temperatum*, *Graphilbum fragrans*, *Penicillium nordicum*, and *Thielaviopsis musarum*. *IMA Fungus* 6:493–506
- Wingfield BD, Berger DK, Steenkamp ET, Lim H-J, Duong TA et al (2017) Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntliella decipiens* and *Ophiostoma ips*. *IMA Fungus* 8:385–396
- Wingfield BD, Bills GF, Dong Y, Huang W, Nel WJ et al (2018a) Draft genome sequence of *Annulohyphoxylon stygium*, *Aspergillus mulundensis*, *Berkeleyomyces basicola* (syn. *Thielaviopsis basicola*), *Ceratocystis smalleyi*, two *Cercospora beticola* strains, *Coleophoma cylindrospora*, *Fusarium fracticaudum*, *Phialophora cf. hyalina*, and *Morchella septemlobata*. *IMA Fungus* 9:199–223
- Wingfield BD, Duong TA, Hammerbacher A, Van der Nest MA, Wilson A et al (2016b) Draft genome sequences for *Ceratocystis fagacearum*, *C. harringtonii*, *Grossmannia penicillata*, and *Huntliella bhutanensis*. *IMA Fungus* 7:317–323
- Wingfield BD, Liu M, Nguyen HDT, Lane FA, Morgan SW et al (2018b) Nine draft genome sequences of *Claviceps purpurea* s.l., including *C. arundinis*, *C. humidiphila*, and *C. cf. spartinae*, pseudomolecules for the pitch canker pathogen *Fusarium circinatum*, draft genome of *Davidsoniella eucalypti*, *Grossmannia galeiformis*, *Quambalaria eucalypti*, and *Teratosphaeria destructans*. *IMA Fungus* 9:401–418
- Wingfield MJ, Barnes I, De Beer ZW, Roux J, Wingfield BD et al (2017b) Novel associations between ophiostomatoid fungi, insects and tree hosts: current status - future prospects. *Biol Invasions* 19:3201–3214
- Wingfield MJ, Crous PW, Coutinho TA (1996) A serious canker disease of Eucalyptus in South Africa caused by a new species of *Coniothyrium*. *Mycopathologia* 136:139–145
- Witthuhn RC, Harrington TC, Wingfield BD, Steimel JP, Wingfield MJ (2000) Deletion of the *MAT-2* mating-type gene during uni-directional mating-type switching in *Ceratocystis*. *Curr Genet* 38:48–52
- Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.