

## IMA Genome-F 9

**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 septimelata***

Brenda D. Wingfield<sup>1</sup>, Gerald F. Bills<sup>2</sup>, Yang Dong<sup>3,4,5</sup>, Wenli Huang<sup>6</sup>, Wilma J. Nel<sup>1</sup>, Benedicta S. Swalarsk-Parry<sup>1</sup>, Niloofar Vaghefi<sup>7</sup>, P. Markus Wilken<sup>1</sup>, Zhiqiang An<sup>2</sup>, Z. Wilhelm de Beer<sup>1</sup>, Lieschen De Vos<sup>1</sup>, Li Chen<sup>2</sup>, Tuan A. Duong<sup>1</sup>, Yun Gao<sup>8</sup>, Almuth Hammerbacher<sup>9</sup>, Julie R. Kikkert<sup>10</sup>, Yan Li<sup>2,12</sup>, Huiying Li<sup>11</sup>, Kuan Li<sup>13</sup>, Qiang Li<sup>6</sup>, Xingzhong Liu<sup>13</sup>, Xiao Ma<sup>14</sup>, Kershney Naidoo<sup>1</sup>, Sarah J. Pethybridge<sup>7</sup>, Jingzu Sun<sup>12,13</sup>, Emma T. Steenkamp<sup>1</sup>, Magriet A. van der Nest<sup>1</sup>, Stephanie van Wyk<sup>1</sup>, Michael J. Wingfield<sup>1</sup>, Chuan Xiong<sup>6</sup>, Qun Yue<sup>2,15</sup>, and Xiaoling Zhang<sup>13</sup>

<sup>1</sup>Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag x20, Hatfield, Pretoria, 0028, South Africa; corresponding author e-mail: [brenda.wingfield@up.ac.za](mailto:brenda.wingfield@up.ac.za)

<sup>2</sup>Texas Therapeutics Institute, The Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77054, USA

<sup>3</sup>State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming, 650201, Yunnan, China

<sup>4</sup>Key Laboratory for Agro-biodiversity and Pest Control of Ministry of Education, Yunnan Agricultural University, Kunming, 650201, Yunnan, China

<sup>5</sup>College of Biological Big Data, Yunnan Agriculture University, Kunming 650504, Yunnan, China

<sup>6</sup>Biotechnology and Nuclear Technology Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu 610065, Sichuan, China

<sup>7</sup>School of Integrative Plant Science, Plant Pathology & Plant-Microbe Biology Section, Cornell University, Geneva, NY 14456, USA

<sup>8</sup>Nowbio Biotechnology Company, Kunming, 650201, Yunnan, China

<sup>9</sup>Department of Zoology Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag x20, Hatfield, Pretoria, 0028, South Africa

<sup>10</sup>Cornell Cooperative Extension, Canandaigua, NY 14424, USA

<sup>11</sup>Kunming University of Science and Technology, Kunming 650500, Yunnan, China

<sup>12</sup>Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China

<sup>13</sup>State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

<sup>14</sup>Yunnan Plateau Characteristic Agricultural Industry Research Institute, Kunming 650201, Yunnan, China

<sup>15</sup>Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China

**Abstract:** Draft genomes of the species *Annulohyphoxylon stygium*, *Aspergillus mulundensis*, *Berkeleyomyces basicola* (syn. *Thielaviopsis basicola*), *Ceratocystis smalleyi*, two *Cercospora beticola* strains, *Coleophoma cylindrospora*, *Fusarium fracticaudum*, *Phialophora* cf. *hyalina* and *Morchella septimelata* are presented. Both mating types (MAT1-1 and MAT1-2) of *Cercospora beticola* are included. Two strains of *Coleophoma cylindrospora* that produce sulfated homotyrosine echinocandin variants, FR209602, FR220897 and FR220899 are presented. The sequencing of *Aspergillus mulundensis*, *Coleophoma cylindrospora* and *Phialophora* cf. *hyalina* has enabled mapping of the gene clusters encoding the chemical diversity from the echinocandin pathways, providing data that reveals the complexity of secondary metabolism in these different species. Overall these genomes provide a valuable resource for understanding the molecular processes underlying pathogenicity (in some cases), biology and toxin production of these economically important fungi.

**Key words:**

*Beta vulgaris*  
*Carya cordiformis*  
 echinocandin gene clusters  
 mulundocandins  
 Pitch canker  
 peumocandins

**Article info:** Submitted: 28 March 2018; Accepted: 28 May 2018; Published: 11 June 2018.

© 2018 International Mycological Association

You are free to share - to copy, distribute and transmit the work, under the following conditions:

**Attribution:** You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

**Non-commercial:** You may not use this work for commercial purposes.

**No derivative works:** You may not alter, transform, or build upon this work.

For any reuse or distribution, you must make clear to others the license terms of this work, which can be found at <http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode>. Any of the above conditions can be waived if you get permission from the copyright holder. Nothing in this license impairs or restricts the author's moral rights.

## IMA Genome-F 9A

Draft genome sequence of *Annulohyphoxylon stygium*

## INTRODUCTION

*Annulohyphoxylon stygium* (Xylariales, Ascomycota) is a white-rot fungus commonly found on dead wood (Hsieh *et al.* 2005). *Annulohyphoxylon stygium* displays an extremely high performance in lignin and carbohydrate degradation. Some species of *Annulohyphoxylon* may be used in the cultivation of *Tremella fuciformis*, one of the foremost medicinal and culinary fungi of China (Stamets 2000). *Tremella fuciformis*, the white jelly mushroom, is a symbiotic fungus that does not form an edible basidiome without the presence of a specific host fungus (Li *et al.* 2014). Its preferred host has traditionally been indicated as “Xianghui” in China. Recently, *A. stygium* was identified to be the main Xianghui species and this has been confirmed experimentally (Deng *et al.* 2016). Cultivators usually pair cultures of *T. fuciformis* with this species for industrial production and the formation of *T. fuciformis* basidiomes is highly dependent on the presence of the specific host fungus, both in nature and for industrial production.

To date, the symbiotic mechanism of *A. stygium* and *T. fuciformis* has not been understood yet. The genome sequence of *A. stygium* from this study may provide some useful information to reveal the symbiotic mechanism of *A. stygium* with *T. fuciformis*.

## SEQUENCED STRAIN

**China:** Sichuan: Tongjiang, N 31°42', E 120°17', alt. 1523 m, isolated from dead wood, 8 Aug. 2015, Qiang Li & Chuan Xiong (MG137 – dried culture).

## NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The Whole Genome Shotgun project isolate (culture collection number SAAS137) has been deposited at DDBJ/EMBL/GenBank under accession number PYLT00000000. The version described in this paper is version PYLT01000000.

## MATERIALS AND METHODS

*Annulohyphoxylon stygium* MG137 was isolated from dead wood in Tongjiang, Sichuan province, China, and was preserved in the Fungal Culture Collection, Center in Biotechnology and Nuclear Technology Research Institute, Chengdu, Sichuan, China. Genomic DNA was isolated from this isolate and subjected to sequencing on the Genome Analyzer IIx next-generation sequencing platform (Illumina) at the Beijing Genomics Institute (Shenzhen, China). Paired-end libraries with respective insert sizes of 425 bp and 725 bp

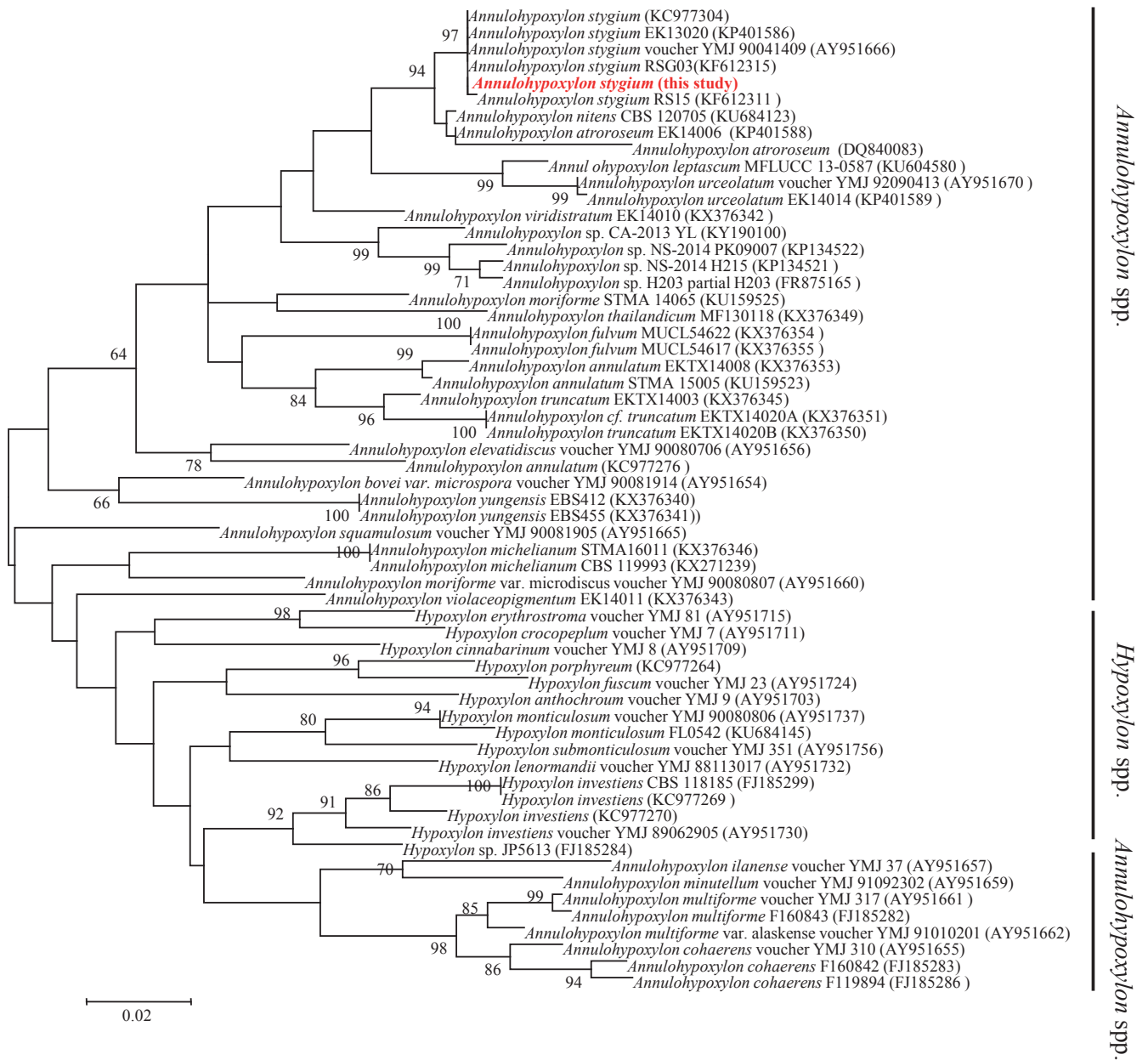
were used to generate read lengths of 150 bases. The CLC Genomics Workbench v. 6.0.1 (CLCBio, Aarhus, Denmark) was subsequently used to trim reads of poor quality (limit of 0.05) as well as terminal nucleotides. The remaining reads were assembled using the SPAdes 3.0.0 with an optimized k-mer value of 103 (Bankevich *et al.* 2012). Thereafter, scaffolding was completed using SSPACE v. 2.0 (Boetzer *et al.* 2011) and gaps reduced with the use of GapFiller v. 2.2.1 (Boetzer & Pirovano 2012). The completeness of the assembly was evaluated using the BUSCO v3 (Simão *et al.* 2015). Homology-based gene prediction and *ab initio* prediction were performed to search *A. stygium* gene models. Homologous protein from *Laccaria bicolor* was used for alignment to the repeat-masked *A. stygium* genome using Exonerate v 2.2.0 (Slater & Birney, 2005). The filtered alignment results (above 300 bp and 90 % coverage) were built as training models for *ab initio* gene prediction. The *ab initio* prediction was conducted using Augustus v. 3.2.3 (Stanke *et al.* 2008) and GeneMark-ES (Ter-Hovhannisyan *et al.* 2008) guided by training models from homology-based alignments. All gene prediction results were integrated into the final gene models by EVIDENCEModeler (Haas *et al.* 2008). Carbohydrate-active enzymes (CAZyme), including the repertoire of auxiliary enzymes, were predicted using dbCAN (Yin *et al.* 2012).

## RESULTS AND DISCUSSION

The genome of *A. stygium* had an estimated size of 47.5 Mb with an average coverage of 31.26× (Table 1). The N50 size was 598 310 bases, and the assembly had a mean GC content of 46 %. The total number of scaffold generated was 1854. MAKER predicted a total of 12 498 genes with an average length of 1662 bp. The average gene density of *A. stygium* was 263 genes/Mb. A phylogenetic analysis of the genus *Annulohyphoxylon* and the closely related genus *Hypoxylon* is presented to reflect the position of this genome (Fig. 1.).

The draft genome of *A. stygium* is larger than that of the allied species *Xylaria hypoxylon* OSC100004 and *Hypoxylon* sp. CI-4A (Wu *et al.* 2017), which are 42.9 Mb and 37.7 Mb, respectively. The genome is closer in size to that of *Hypoxylon* sp. CO27-5 and *Hypoxylon* sp. EC38 (Wu *et al.* 2017), which have genome sizes of 46.6 Mb and 47.7 Mb, respectively. *Annulohyphoxylon stygium* also has a similar number of putative genes when compared to *Hypoxylon* sp. EC38 (12261 predicted gene models) and *Hypoxylon* sp. CO27-5 (12 256 predicted gene models).

A total 757 CAZymes were identified in the genome of *A. stygium*, more than that in the closely related *Hypoxylon* sp. CO27-5 (599 CAZymes) and *Hypoxylon* sp. CI-4A (526 CAZymes). The number of CAZymes in *A. stygium* was much higher than that in *Tremella encephala* (265 CAZymes; Magnuson *et al.* 2017) and *T. mesenterica* (206 CAZymes; Floudas *et al.* 2012), indicating that *A. stygium* may assist *Tremella* species in the degradation of lignin and carbohydrates in nature or for industrial production. The genome sequence data of *A. stygium* in this study will provide useful information for understanding the mechanism of the symbiotic interaction between *A. stygium* and *T. fuciformis*.



**Fig. 1.** Maximum Likelihood (ML) phylogenetic analysis of the genus *Annulohypoxyton* and the closely related genus *Hypoxyton* using MEGA 6.06 based on partial gene sequences of  $\beta$ -tubulin. Bootstrap values were calculated using 1000 replicates to assess node support. *Annulohypoxyton stygium* isolates used for verification was extracted from the assembled genomes. Reference sequences are obtained from the NCBI database with accession number.

Authors: Q. Li, X. Ma, H. Li, C. Xiong, Y. Gao, Y. Dong\*, and W. Huang\*

\*Contact: Yang Dong: loyalyang@163.com; Wenli Huang: wenlih11@126.com

**Table 1.** Whole genome DNA sequence assemblies generated in *Annulohypoxyton stygium* MG137. The genomes of *A. stygium* MG137 were generated using next generation sequencing technology.

Genome	<i>Annulohypoxyton stygium</i> MG137
Coverage	31.26
BUSCO	96.6%
Total sequence length (Mb)	47.5
<b>Scaffolds</b>	1854
Scaffold N50	598 310
GC (%)	46
Predicted gene models	12 498
<b>Predicted CAZymes</b>	
Total CAZymes	757
Auxiliary activities	153
Pectate lyases	13
Glycosyltransferases	106
Glycoside hydrolases	297
Carbohydrate esterases	125
Carbohydrate binding motifs	63
<b>Predicted secondary Metabolite (sM) Clusters</b>	
Total SM clusters	90
Type I polyketide synthetases (PKSs)	36
Type III PKSs	1
Nonribosomal peptide synthetases (NRPSs)	21
Terpene clusters	10
HgIks	0

## IMA Genome-F 9B

# Draft genome sequence of *Aspergillus mulundensis*, a fungus that produces mulundocandins

## INTRODUCTION

A strain of *Aspergillus* (Y-30462 = DSMZ 5745) was isolated at Hoechst India, then located in the Mulund district of Mumbai, India, from a soil sample collected in Bangladesh (Mukhopadhyay *et al.* 1987, Roy *et al.* 1987). In the original publication, the fungus was described as an unusual variant of *A. sydowii* because of the presence of abundant Hülle cells and was published without a Latin description or type specimen as “*A. sydowii* var. *mulundensis*”. This strain was subsequently re-examined using multi-gene phylogenetic analysis, chemotaxonomic markers, and morphological data and was determined as representing a novel species within *Aspergillus* sect. *Nidulantes* (Bills *et al.* 2016, Chen *et al.* 2016).

The primary objective for sequencing the genome of *A. mulundensis* was the identification of the gene cluster-encoding the biosynthesis of the mulundocandins (Yue *et al.* 2015). Mulundocandin and deoxymulundocandin (Fig. 2) are lipohexapeptides and potent antifungal antibiotics of the echinocandin class (Mukhopadhyay *et al.* 1987, Roy *et al.* 1987, Mukhopadhyay *et al.* 1992). Biosynthetically, they are closely related to echinocandin B, but they differ in the

substitution of serine instead of threonine in the fifth position of the hexapeptide core and by a 12-methyl myristoyl side chain instead of a lineoyl side chain. Mulundocandin and its deoxymulundocandin have been investigated extensively as potential lead structures for the development of echinocandin-type antifungal drugs (Mukhopadhyay *et al.* 1992, Hawser *et al.* 1999, Lal *et al.* 2003, 2004). This draft genome will expand genomic data sets for comparative genomics of species in *Aspergillus* sect. *Nidulantes*.

## SEQUENCED STRAIN

**Bangladesh:** unknown location, isolated from soil at Hoechst (Mumbai, India) (Hoechst Y-30462 = DSMZ 5745 = CBS 140610 = IBT 33104).

## NUCLEOTIDE SEQUENCE ACCESSION NUMBER

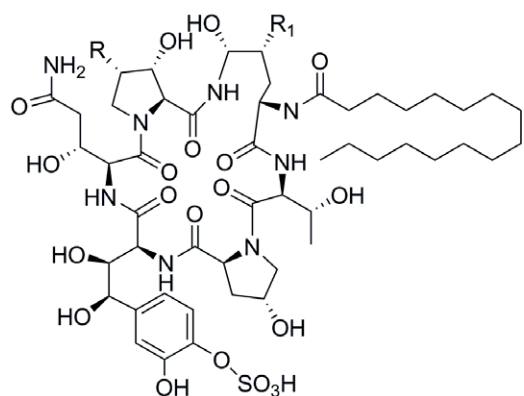
The *Aspergillus mulundensis* isolate DSMZ 5745 Whole Genome Shotgun project has been deposited in GenBank under the accession number PVWQ00000000.

## MATERIALS AND METHODS

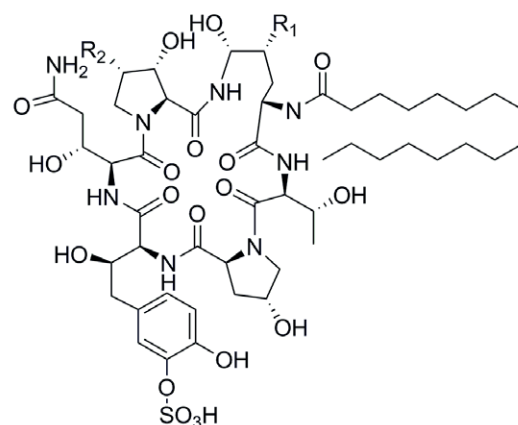
For methods for DNA extraction, sequencing, and genome assembly and annotation, see Bills *et al.* (2016).

**Table 2.** General features of the genomes of *Coleophoma cylindrospora* BP6252 and BP5796, *Phialophora cf. hyalina* BP5553, and *Aspergillus mulundensis* DSMZ 5745.

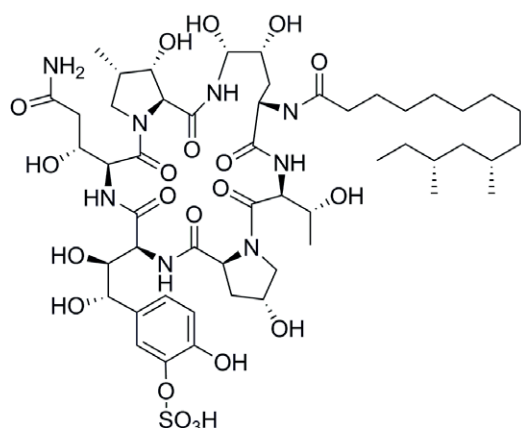
Features	BP6252	BP5796	BP5553	DSMZ 5745
Assembly size (Mb)	42.4	40.4	33.6	45
Scaffolds	77	45	32	160
Scaffold N50 (Mb)	2.3	2	3.8	2.8
Coverage (fold)	100	100	102	100
G+C content (%)	48.7	48.5	48.2	43.2
Protein-coding genes	14177	13257	10707	11603
Gene density (per Mb)	337.55	331.42	324.45	257.84
Exons per gene	3.15	3.13	3.12	3.02
PKSs	15	15	19	25
NRPSs	8	6	13	19
PKS-NRPS hybrids	0	0	6	1
DMATs	2	2	2	4
Terpene synthases	1	1	4	4
Chalcone or stilbene synthase gene	0	0	1	0
Secondary metabolite gene clusters	30	28	48	59



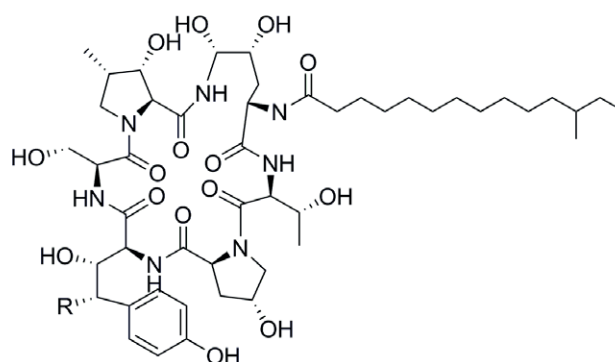
FR220897 R=CH<sub>3</sub>  
FR220899 R=H  
from *Coleophoma cylindrospora* BP6252



FR209602 R1=OH, R2=CH<sub>3</sub>  
FR209603 R1=OH, R2=H  
FR209604 R1=H, R2=OH  
from *Coleophoma cylindrospora* BP5796



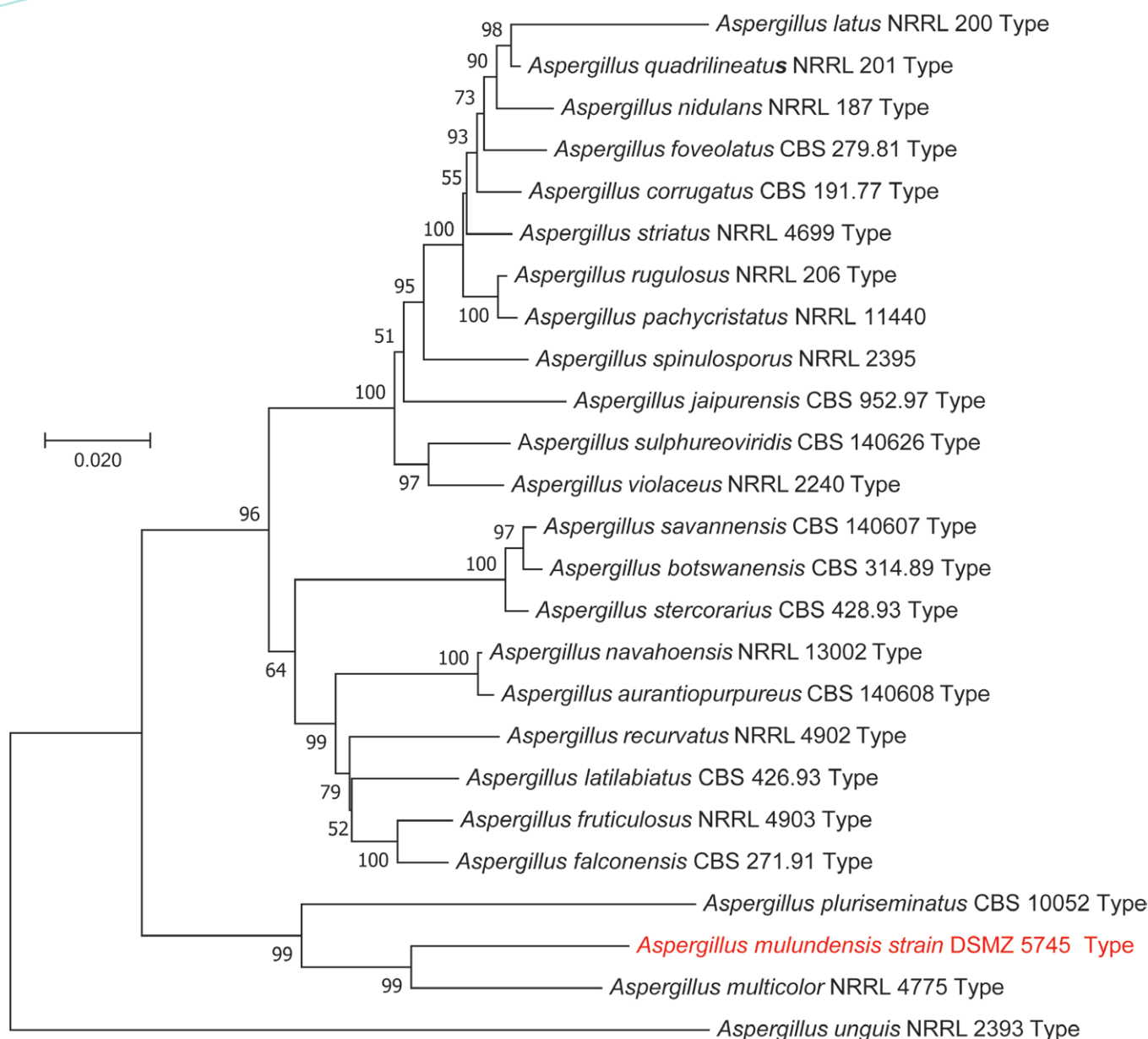
FR190293  
from *Phialophora cf. hyalina* BP5553



Mulundocandin R=OH  
Deoxymulundocandin R=H  
from *Aspergillus mulundensis* DSMZ 5745

**Fig. 2.** Some naturally occurring echinocandins described in the patent literature.





**Fig. 3.** Maximum Likelihood tree of ex-type and authentic strains of *Aspergillus* sect. *Nidulantes* (25 strains) inferred based on an alignment of the concatenated sequences of the ITS-28S rDNA, ribosomal polymerase II,  $\beta$ -tubulin, and calmodulin genes. Data were resampled from Chen *et al.* (2016). DMSZ 5745 is labelled in red, and *A. unguis* was positioned as the outgroup. The Maximum Likelihood tree was based on the Tamura-Nei model. The tree with the highest log likelihood ( $-13\ 959.85$ ) is shown. Branches are labelled with the percentage of trees in which the associated taxa clustered together. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.2371)). Branch lengths were measured in the number of substitutions/site. The dataset included 3329 positions. Data were analyzed in MEGA7 (Kumar *et al.* 2016). "Type" = ex-type cultures.

## RESULTS AND DISCUSSION

The genome of DSMZ 5745 was sequenced to 100-fold coverage, yielding 160 scaffolds with N50 of 2.8 Mb (Table 2). The assembled genome size was 45 Mb, and a total of 11603 genes were predicted. The GC content of this genome is 43.2 %. The genome contains 53 core catalytic genes associated with putative secondary metabolite biosynthetic gene clusters. These clusters include 25 PKSs, 19 NRPSs, one PKS-NRPS hybrids, four dimethylallyl tryptophan synthases, and four terpene synthases. These genes are distributed among 45 putative

gene clusters that also include genes encoding tailoring enzymes, regulators, transporters, and other auxiliary genes. In addition to these gene clusters, 14 secondary metabolite gene clusters containing PKS-like or NRPS-like enzyme genes, or other secondary metabolic-related genes were identified by antiSMASH. In addition a gene cluster containing close orthologues of the pneumocandin gene cluster from *Glarea lozoyensis* (Yue *et al.* 2015) was recognized, and predicted to be responsible for the biosynthesis of muludocandins. The nuclear-encoded secondary metabolomes of *A. mulundensis* and *A. nidulans* FGSC A4 were compared previously (Bills *et al.* 2016). A

phylogenetic tree reflecting the position of this genomes in relation to other *Aspergillus* species is presented (Fig. 3.)

Authors: Q. Yue, Y. Li, L. Chen, X. Zhang, K. Li, J. Sun, X. Liu, Z. An, and G.F Bills\*  
\*Contact: billsge@vt.edu

## IMA Genome-F 9C

# Draft genome sequence of the root pathogen *Berkeleyomyces basicola* (syn. *Thielaviopsis basicola*)

## INTRODUCTION

*Berkeleyomyces basicola* (Ascomycota: Microascales), previously known as *Thielaviopsis basicola* (Nel *et al.* 2017), is an important plant pathogen responsible for root rot of many important agricultural and ornamental plants (Johnson 1916, Stover 1950, Nehl *et al.* 2004, Pereg 2013). Since its description in the mid-1800s (Berkeley & Broome 1850), there has been considerable debate surrounding its appropriate taxonomic placement resulting in numerous name changes. The phylogenetic re-evaluation of *Ceratocystidaceae* by De Beer *et al.* (2014) raised new questions regarding the appropriate taxonomic placement of the species. Their results suggested that *T. basicola* did not group in *Thielaviopsis* or any other genus described in the family. Because the authors included only the sequence data of a single isolate in their analyses they concluded that no taxonomic changes could be made without further study. In a recent investigation, Nel *et al.* (2017) confirmed that *T. basicola* represented a distinct generic lineage in *Ceratocystidaceae* and introduced the new generic name *Berkeleyomyces*. In addition, they showed that isolates of *Berkeleyomyces* represented two cryptic sister species for which they provided the names *B. basicola* and *B. rouxiae*.

The aim of this study was to generate a high-quality genome sequence for *B. basicola*. This would allow for comparisons to be made with the available genomes of other species in *Ceratocystidaceae*, including those in the genera *Ceratocystis*, *Huntia*, *Davidsoniella*, *Thielaviopsis*, *Chalaropsis*, *Endocnidiophora* and the recently described *Bretziella* (De Beer *et al.* 2017). Here we report the complete genome sequence of isolate CMW 49352, the designated reference specimen for *B. basicola* logged in CBS (Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands), and the culture collection of the Forestry and Agricultural Biotechnology Institute (CMW), University of Pretoria, South Africa.

## SEQUENCED STRAIN

**The Netherlands:** *South Holland*, Boskoop, isol ex *Betula* sp., June 1974. S.G. De Hoog (CMW 49352 = CBS 142796; PREM 62125 = dried culture).

## NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The draft genome sequence of *Berkeleyomyces basicola* (CMW 49352 = CBS 142796) has been deposited at DDBJ/ENA/GenBank under the accession number PJAC00000000. The version presented here is PJAC00000000.

## MATERIALS AND METHODS

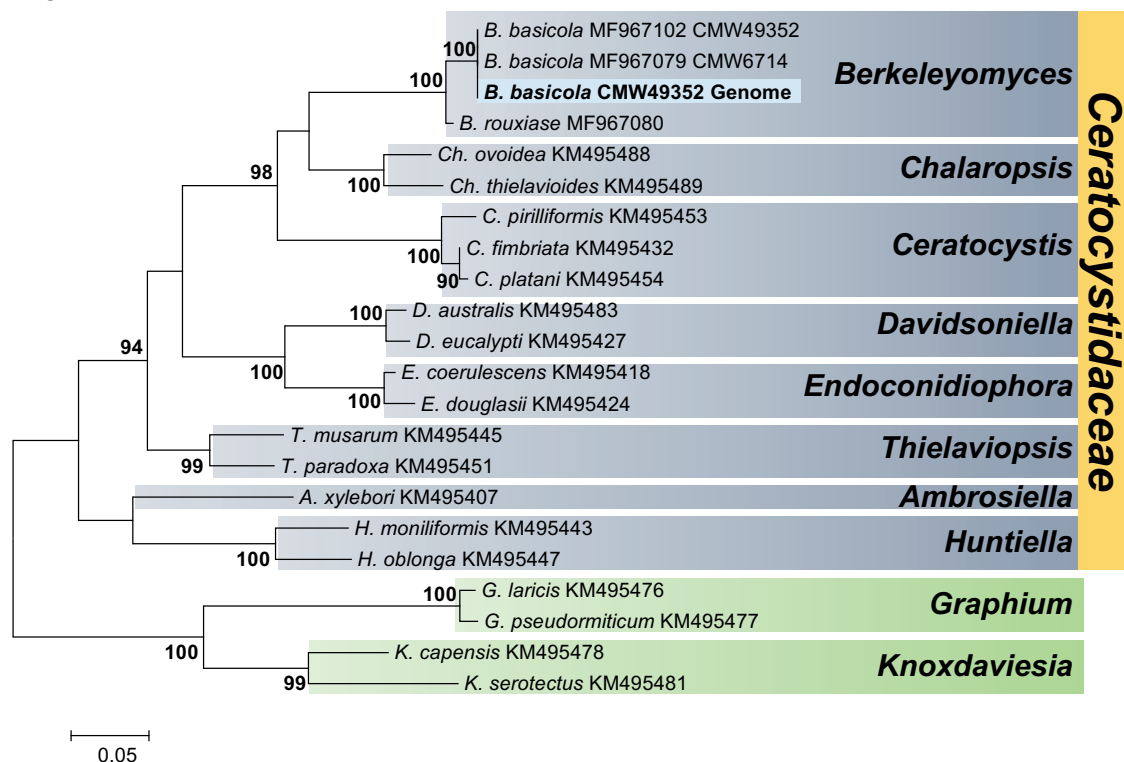
Genomic DNA was extracted from lyophilized mycelium of *Berkeleyomyces basicola* isolate CMW 49352 grown in malt yeast broth (2 % Malt extract, 0.5 % yeast extract; Biolab, Midrand, South Africa) using the method described by Duong *et al.* (2013). A paired-end library was prepared (350 bp average insert sizes) and sequenced using the Illumina HiSeqX Platform. A mate-pair library was prepared (10 Kb average insert size) and sequenced using the Illumina HiSeq2500 platform. Long reads were also generated using one cell of the Single-molecule real time (SMRT or PacBio) sequencing platform (Pacific BioScience). All sequencing was conducted at Macrogen (Seoul, Korea). Quality and adapter trimming of pair-end and mate-pair reads was carried out using Trimmomatic v. 0.36 (Bolger *et al.* 2014). *De novo* assembly of the genome was carried out using SPAdes v. 3.9 (Bankevich *et al.* 2012) using all pair-end, mate-pair and PacBio data. Contigs smaller than 500 bp were removed from the dataset. Initial scaffolding was done using SSPACE-standard v. 3.0 (Boetzer *et al.* 2011) with the paired-end and mate-pair reads. A second round of scaffolding was done using SSPACE-Longread with the PacBio reads. Assembly gaps were filled using GapFiller v. 1.10 (Boetzer & Pirovano 2012) with the paired-end and mate-pair reads, and using PBjelly (English *et al.* 2012) with PacBio reads. Final genome polishing was done using Pilon (Walker *et al.* 2014). Genome completeness was assessed with the Benchmarking Universal Single-Copy Orthologs (BUSCO v. 1.1b1) tool using the Ascomycota dataset (Simão *et al.* 2015). The number of protein coding genes was determined using Augustus v. 3.3.2 (Stanke *et al.* 2004) using pre-optimised species models for *Fusarium graminearum*.

## RESULTS AND DISCUSSION

The paired-end, mate-pair, and PacBio sequencing yielded 431 141 384, 60 673 400 and 42 422 reads, respectively. Final assembly consisted of 81 contigs, with the largest around 3.8 Mb and an N50 of 1.2 Mb. The estimated size of the genome is around 25.1 Mb with a GC content of 52 %. This estimated size is similar to that of other species in *Ceratocystidaceae*, which range between 25.4 Mb for *Huntia moniliformis* and 33.6 Mb for *Davidsoniella virescens* (Wilken *et al.* 2013, Van der Nest *et al.* 2014a, b, Wingfield *et al.* 2015a, b, 2016a, b). The phylogenetic position of *Berkeleyomyces basicola* is presented in Fig. 4.

BUSCO analysis predicted an assembly completeness of 97.4 %. The assembly contained 1280 complete single-copy BUSCOs, one complete and duplicated BUSCOs, 10 fragmented BUSCOs and 24 missing BUSCOs out of a total 1315 BUSCO groups searched. AUGUSTUS annotation predicted 10 074 putative coding regions, corresponding to

## MCM7



**Fig. 4.** Maximum Likelihood (ML) phylogram derived from the analyses of the partial MCM7 gene sequences for species in *Ceratocystidaceae*. CLCbio Genomics Workbench v. 9.5 (CLCbio, QIAGEN, Aarhus, Denmark) was used to screen the genome of *B. basicola* isolate CMW 49352 to identify and extract the MCM7 gene using an available reference sequence for the gene from *B. basicola* (Accession: MF967102). A dataset was prepared based on the phylogenies of Nel *et al.* (2017) and sequences were downloaded from NCBI GenBank. DNA sequence alignments of the dataset were done using the online version of MAFFT v. 7 (Kato & Standley, 2013). The ML analyses were performed in MEGA v. 6.06 (Tamura *et al.* 2013) using the GTR model. Values shown at nodes are confidence values >75 %. The sequence from the *B. basicola* genome is indicated in bold.

around 401 ORFs/Mb. The availability of the genome for *B. basicola* will make possible genome comparisons with other species in *Ceratocystidaceae* and facilitate investigations into factors involved in pathogenicity, ecology, mating, and evolution of this important plant pathogen.

Authors: W.J. Nel, T.A. Duong, A. Hammerbacher, B.D. Wingfield, M.J. Wingfield, and Z.W. De Beer  
\*Contact: janine.nel@fabi.up.ac.za

## IMA Genome-F 9D

### Draft nuclear genome assembly for *Ceratocystis smalleyi*

#### INTRODUCTION

The genus *Ceratocystis* as defined by De Beer *et al.* (2014) is a diverse assemblage of species that are best known as pathogens of angiosperm trees and commercially grown root crops (De Beer *et al.* 2014, Li *et al.* 2016, Seifert *et al.* 2013, Van Wyk *et al.* 2012). Among these, *C. fimbriata* s. lat. is arguably the best-known pathogen, and has been associated

with diseases of sweet potato (Halsted & Fairchild 1891), taro (Huang *et al.* 2008), pomegranate (Somasekhara 1999) and kiwifruit (Piveta *et al.* 2016). The related *C. manginecans* causes disease on mango and *Acacia mangium* trees in Oman and Pakistan (Al-Subhi *et al.* 2006, Al Adawi *et al.* 2013), while *C. eucalypticola* is responsible for mortality on commercially planted eucalypt trees in South Africa (Van Wyk *et al.* 2012). These fungi all share a very similar morphology, making their species boundaries difficult to determine (Fourie *et al.* 2014, Harrington *et al.* 2014). In contrast, several other species in the genus are clearly defined, with universally accepted species status (Engelbrecht & Harrington 2005). These include *C. albifundus* (a pathogen of commercially propagated *Acacia mearnsii* and *Protea cynaroides* in South Africa; Lee *et al.* 2016), *C. cacaofunesta* (causing cacao wilt in the Caribbean and Central and South America; Engelbrecht *et al.* 2007), and *C. smalleyi* (agent of hickory decline in the USA; Johnson *et al.* 2005).

*Ceratocystis smalleyi* was first isolated from a hickory tree (*Carya* sp.) that had been infested by the hickory bark beetle *Scolytus quadrispinosus* (Johnson *et al.* 2005). In 2005, *C. smalleyi* was formally named and described after additional isolates were collected from *Carya* trees that had been attacked by the hickory bark beetle across parts of the eastern US (Johnson *et al.* 2005). The authors subsequently linked *C. smalleyi* with the decline of hickory through a possible



association with the bark beetle *S. quadrispinosus* (Johnson *et al.* 2005). Later studies have confirmed *C. smalleyi* as a pathogen on *Carya* species (Park *et al.* 2010, 2013), and established the close association between the fungus and the bark-beetle (Juzwik *et al.* 2010). This makes *C. smalleyi* the only known *Ceratocystis* species to be associated with a bark-beetle. In other *Ceratocystis* species, the production of volatiles is linked to attracting insects for dispersal (Van Wyk *et al.* 2009, 2012). The specific association between *C. smalleyi* and the vector *S. quadrispinosus* would eliminate the need for producing volatile attractants, and could explain the inability of this species to produce the fruity odours characteristic of other *Ceratocystis* species (Harrington 2009; Johnson *et al.* 2005).

In this study, we aimed to produce a draft genome assembly for *C. smalleyi*. This assembly would be the seventh *Ceratocystis* species for which a genome sequence is published, and adds to the valuable genomic resource available for members of *Ceratocystidaceae* (Molano *et al.* 2018, Van der Nest *et al.* 2014a, b, 2015, Vanderpool *et al.* 2017, Wilken *et al.* 2013, 2018, Wingfield *et al.* 2015a, b, 2016a, b). Furthermore, the availability of a genome assembly will afford the opportunity in future to investigate aspects of the unique biology of *C. smalleyi*.

## SEQUENCED STRAIN

**USA:** *Wisconsin:* Hickory Ridge, isol. *Carya cordiformis*. Oct. 1993, *G. Smalley* (CMW 14800, CBS 114724, BPI 843722 – dried culture).

## NUCLEOTIDE SEQUENCE ACCESSION NUMBER

This Whole Genome Shotgun project for *Ceratocystis smalleyi* isolate CMW 14800 has been deposited at DDBJ/ENA/GenBank under accession NETT00000000. The version described in this paper is version NETT01000000.

## MATERIALS AND METHODS

*Ceratocystis smalleyi* isolate CMW 14800 was obtained from the culture collection of the Forestry and Agricultural Biotechnology Institute (FABI) and grown on 2 % malt extract agar (MEA: 2 % w/v, Biolab, South Africa) at 25 °C. A 14 d old culture was used to isolate genomic DNA using a previously described phenol-chloroform protocol (Roux *et al.* 2004). The isolated DNA was submitted for sequencing on an Illumina Genomics Analyzer Ix at the UC Davis Genome Centre (University of California, Davis). For sequencing, paired-end libraries of 350 bp and 600 bp insert sizes were prepared and sequenced following the protocol provided by Illumina ([www.illumina.com](http://www.illumina.com)). The raw sequencing reads were imported into CLC Genomics Workbench v. 7.5.1 (CLCBio, Aarhus), and default settings were used to both trim the reads for quality and to produce a *de novo* genome assembly using the trimmed reads. Scaffolds were generated from the assembly using

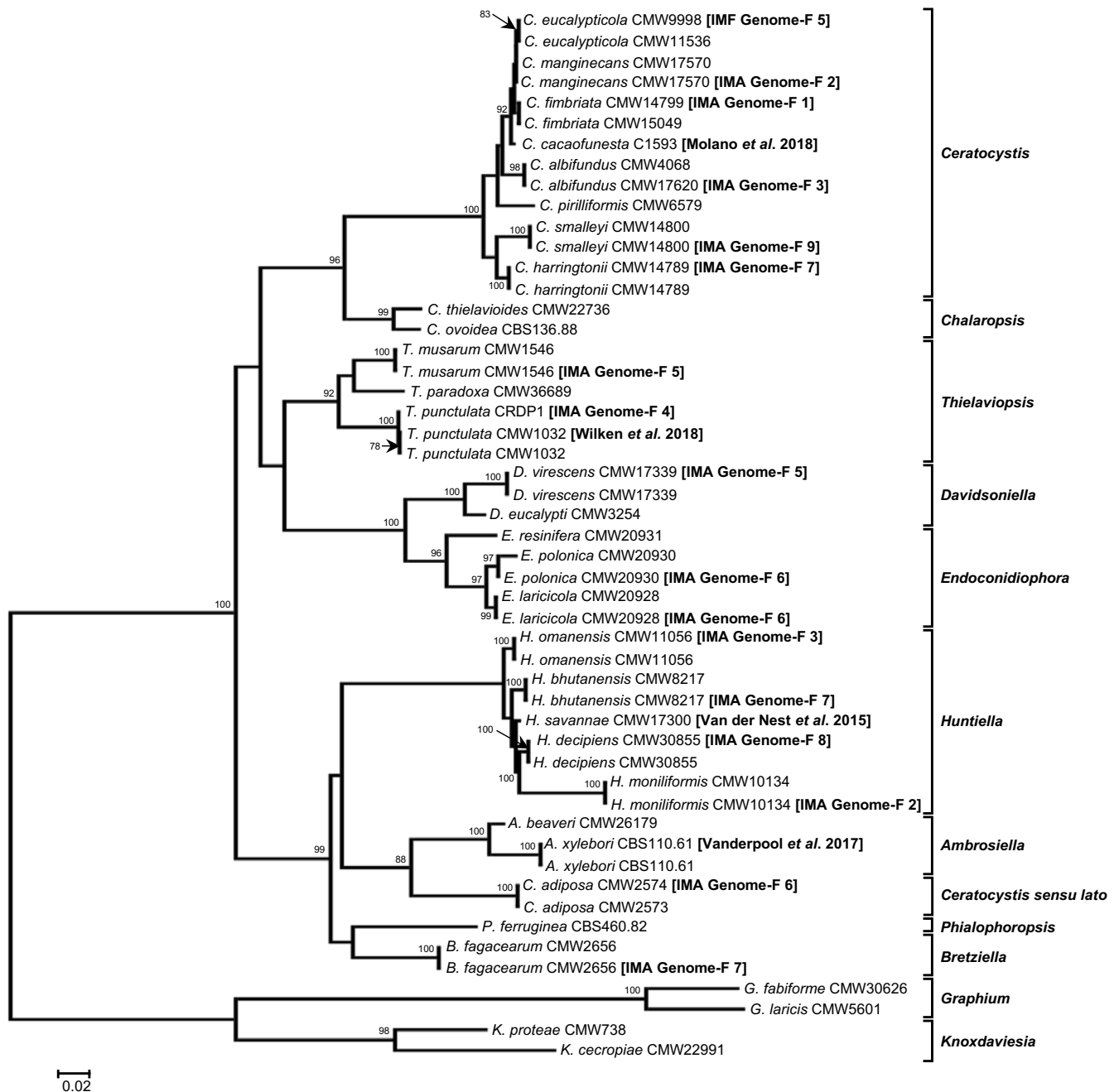
SSPACE v. 2.0 (Boetzer *et al.* 2011), while GapFiller v. 2.2.1 (Boetzer & Pirovano 2012) was used to fill any gaps created during scaffolding. Sequencing coverage was estimated by mapping the trimmed sequencing reads to the contigs, while an estimate of the number of putative open reading frames (ORFs) were obtained through *de novo* gene prediction using the web-based version of AUGUSTUS and gene models from *Fusarium graminearum* (Keller *et al.* 2011). The Benchmarking Universal Single-Copy Orthologs (BUSCO v. 1.22) tool was used in combination with the fungal data set to provide a quantitative measure of the level of genome completeness (Simão *et al.* 2015). The 60S, LSU and MCM7 gene regions were extracted from the genome and, together with these regions from the recently sequenced species *C. cacaofunesta* (Molano *et al.* 2018), *T. punctulata* isolate CMW1032 (Wilken *et al.* 2018), *H. savannae* (Van der Nest *et al.* 2015) and *A. xylebori* (Vanderpool *et al.* 2017) were added to the *Ceratocystidaceae* dataset used for phylogenetic analysis by Wingfield *et al.* (2017). The resulting datasets were aligned using MUSCLE (Edgar 2004), concatenated, and used to construct a Maximum Likelihood phylogeny using PhyML 3.1 (Guindon *et al.* 2010) based on model parameters estimated with jModelTest 2.1.10 (Darriba *et al.* 2012).

## RESULTS AND DISCUSSION

The 27 311 342 bp *Ceratocystis smalleyi* genome was present in 2261 contigs, of which 1242 contigs were larger than 500 bp. The draft assembly yielded a genome with a G/C content of 50.6 %, an average coverage of 84x and 6682 predicted open reading frames at an average gene density of 245 ORFs/Mb. BUSCO analysis indicated a genome completeness of 97 % with 1394 of the 1438 searched orthologs present in the genome being complete. In total 1330 ORFs occurred as single copies while 64 were duplicates. Of the remaining searched homologs, 37 were fragmented while the remaining seven were missing from the genome assembly.

The genome of *C. smalleyi* was comparable in size and gene content to that of other *Ceratocystis* species (Wingfield *et al.* 2015b, 2016a, b). At 27.3 Mb, the *C. smalleyi* genome is slightly larger than that of the related species *C. harringtonii* (genome size of 26 Mb; Wingfield *et al.* 2016b), but smaller than the genome of *C. manginecans* (31.7 Mb; Van der Nest *et al.* 2014b). Gene densities for published *Ceratocystis* genomes range from 204–257 ORF/Mb (Wingfield *et al.* 2015b, 2016a, b), and the *C. smalleyi* gene density falls within this range. In contrast, the 50.6 % G/C content of the *C. smalleyi* genome is unusually high, with all other *Ceratocystis* species showing G/C contents below 49 % (Wingfield *et al.* 2015b, 2016a, b).

The availability of multiple *Ceratocystis* genomes (Fig. 5) provides the opportunity to study the genetic aspects that underlie ecological and life-style differences between members of this genus. Understanding these differences will also be crucial in explaining at least some of the variations in gene content, genome size, and G/C content evident among these genomes. Currently, the published *Ceratocystis* genomes make up the bulk of the *Ceratocystidaceae*



**Fig. 5.** A Maximum Likelihood phylogeny showing *Ceratocystidaceae* isolates for which published whole-genome sequences are available, including that of *C. smalleyi* discussed here. The 60S, LSU, and MCM7 gene regions were used, and was either extracted from the assembled genomes or were obtained from the study of Wingfield et al. (2017). Phylogeny constructed using the TrN+I+G model with confidence values based on 1000 bootstrap replicates. Only bootstrap values  $\geq 75$  are shown.

genome resource with published genomes available for seven species (Fig. 5; Molano et al. 2018, Van der Nest et al. 2014a, b, Wilken et al. 2013, Wingfield et al. 2015b, 2016b). In addition, published genome sequences are available for five *Huntiella* species (Van der Nest et al. 2014a, b, 2015, Wingfield et al. 2016b, 2017), two *Endoconidiophora* species (Wingfield et al. 2016a), three isolates representing two *Thielaviopsis* species (Wilken et al. 2018, Wingfield et al. 2015a; Wingfield et al. 2015b), one *Davidsoniella* species (Wingfield et al. 2015b), one *Bretziella* species (previously *Ceratocystis fagacearum*; De Beer et al. 2017, Wingfield et al. 2016b), one *Ambrosiella* species (Vanderpool et al. 2017)

as well as for *C. adiposa* (Wingfield et al. 2016a). This brings the number of published *Ceratocystidaceae* genomes to 21, with the genome assemblies of several others publicly available ([www.ncbi.nlm.nih.gov/assembly/?term=ceratocystidaceae](http://www.ncbi.nlm.nih.gov/assembly/?term=ceratocystidaceae)). Such a vast genomic resource will prove valuable to future studies on *Ceratocystidaceae*, a family that include fungal species with diverse life-styles and hosts.

Authors: P.M. Wilken\*, M.A. van der Nest, E.T. Steenkamp, K. Naidoo, M.J. Wingfield, and B.D. Wingfield

\* Contact: Markus.Wilken@fab.up.ac.za

## IMA Genome-F 9E

## Draft genome sequences of two *Cercospora beticola* strains from table beet

### INTRODUCTION

The genus *Cercospora* (*Mycosphaerellaceae*) includes several economically important plant pathogens causing leaf and fruit spots on a range of agricultural crops worldwide (Groenewald *et al.* 2013). *Cercospora* species are known to produce cercosporin, a photo-activated toxin that contributes to pathogenicity on a broad range of crops (Daub *et al.* 2000). *Cercospora beticola* is the cause of Cercospora leaf spot (CLS) on sugar and table beet (*Beta vulgaris* ssp. *vulgaris*), and Swiss chard (*Beta vulgaris* ssp. *cicla*) worldwide (Franc 2010). In New York, CLS is the most important disease affecting foliar health of table beet. Symptoms include leaf spots and necrotic lesions with red to purple margins, which coalesce as the disease progresses, and can result in complete defoliation (Pethybridge *et al.* 2017). In broadacre production systems, maintenance of foliar health is important to enable mechanized harvest. For fresh market sales, the presence of CLS lesions on the leaves may result in rejection (Pethybridge *et al.* 2017).

The control of CLS in table beet is dependent on fungicides (Pethybridge *et al.* 2017). However, resistance to single-site mode of action fungicides threatens the durability of CLS control. Recent studies reported a high frequency of isolates with resistance to quinone outside inhibitor fungicides in New York (Vaghefi *et al.* 2016). Moreover, succinate dehydrogenase inhibitor fungicides, which are known to be effective in controlling CLS on sugar beet, failed to provide efficacious control on table beet (Pethybridge *et al.* 2017), and a few isolates with reduced sensitivity to demethylation inhibitors have been detected (Pethybridge, unpubl.). Identifying genomic regions associated with sensitivity to fungicides will enable rapid screening of *C. beticola* populations. Enhanced genomic information for this pathogen will also facilitate studies into the mechanisms of pathogenicity. *De novo* genome assembly of two *C. beticola* strains from table beet are presented here, and made publically available to facilitate genetic studies of this globally important plant pathogen.

### SEQUENCED STRAINS

**USA:** *New York:* western New York, Batavia, from *Beta vulgaris* ssp. *vulgaris* (table beet), 2014, F.S. Hay (Tb14-085 = ICMP 21692); *ibid.* (Tb14-047 = ICMP 21690).

### NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The Whole Genome Shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accessions PDUH00000000 and PDUJ00000000.

### METHODS

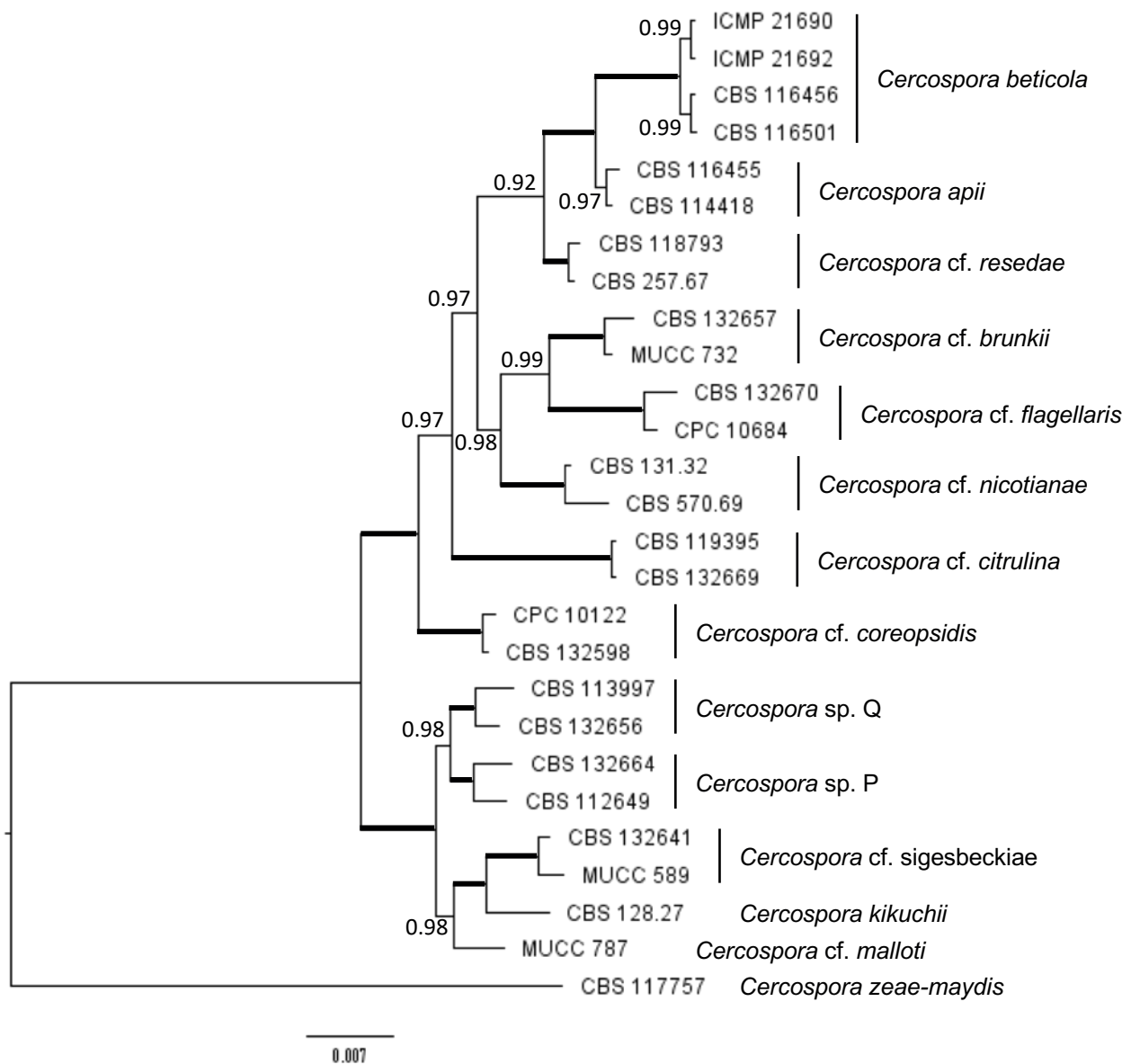
Two *C. beticola* isolates belonging to opposite mating-types (ICMP 21690 [MAT-2] and ICMP 21692 [MAT-1]), collected from table beet in New York, were selected for whole genome sequencing. The identity of the strains as *C. beticola* was confirmed through multi-locus sequence typing of five loci; ITS, actin, calmodulin, histone H3 and translation elongation factor 1-a (Fig. 6). Fungal strains were cultured in clarified V8 broth (10 % (v/v) clarified V8 juice (Campbell's Soup, USA), 0.5 % (w/v) CaCO<sub>3</sub>). Seven-day-old mycelia were harvested, and genomic DNA was extracted as described in Vaghefi *et al.* (2016). The extracted DNA samples were quantified using a Qubit fluorometer (Invitrogen, NY).

A total of 5.6 and 5.0 µg genomic DNA of ICMP 21692 and ICMP 21690 were used to prepare PCR-free libraries with average insert of ~550 bp, using the Illumina paired-end (2×300 bp) MiSeq platform at the Cornell University Institute of Biotechnology Genomics Facility (Ithaca, NY). PCR-free libraries were constructed using Illumina's TruSeq Nano DNA LT Sample Preparation kits, according to the manufacturer's protocol. This yielded 4 607 564 and 4 798 846 paired-end reads, totalling 2.7 and 2.9 Gb data for ICMP 21692 and ICMP 21690, respectively. Quality control of the sequences was conducted using FastQC v.0.11.2 (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>) in the GALAXY portal (Afgan *et al.* 2016). The Kmer counting software Jellyfish v.2.2.3 (Marçais & Kingsford 2011) was used to estimate the genome size.

*De novo* genome assembly was conducted using DISCOVAR *de novo* v.52488; an assembler designed for *de novo* assembly of long Illumina paired-end reads from single PCR-free libraries (Weisenfeld *et al.* 2014). The completeness of the final assemblies was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) v.1.2 (Simão *et al.* 2015). Gene prediction was conducted in the genome annotation pipeline Maker v.2.31.9 (Cantarel *et al.* 2008), using contigs at least 500 bp in length only for ICMP 21692. A preliminary annotation used the *ab initio* gene prediction program SNAP (Korf 2004). The resulting annotation was used to produce a hidden-markov-model (HMM) profile for *C. beticola*, which was further refined with a second stage of SNAP training. The refined HMM file was used for the final annotation (Cantarel *et al.* 2008).

### RESULTS AND DISCUSSION

The Illumina paired-end (2×300 bp) sequencing of *C. beticola* isolates ICMP 21692 and ICMP 21690, resulted in 4 607 564 and 4 798 846 reads for each strain, respectively, with mean base quality of 28.5 and 28.7. The estimated genome size of *C. beticola* was ~37 Mb, based on an approximated genome coverage for both strains of at least 74×. The draft genome of ICMP 21692 had a total assembly size of ~35.03 Mbp (for 1 kb+ scaffolds), a scaffold N50 value of 1 023 488 bp, and maximum contig size of 3 283 856 bp. The draft genome of ICMP 21690 had a total assembly size of ~34.5 Mbp (for 1 kb+ scaffolds), and a scaffold N50 value of 654 439 bp, and maximum contig size of 2 437 838 bp. Both assemblies were



**Fig. 6.** Identity verification of *Cercospora beticola* isolates sequenced in this study. The phylogeny was constructed by Bayesian inference based on the sequences of five loci; ITS, *act*, *cmd*, *his* and *tef1- $\alpha$* . Sequence alignments were produced using MAFFT v. 7 (Kato & Standley 2013) (MrBayes v. 3.1.2; Ronquist & Huelsenbeck, 2003). Branches with posterior probability of 1.00 are thickened. The tree was rooted to *C. zeae-maydis* (CBS 117757).

97 % complete based on the content of BUSCO. All contigs with a length of  $\geq 200$  bp were submitted to the genome database of NCBI. *Ab initio* gene prediction for ICMP 21692 using trained SNAP identified 12 834 Open Reading Frames (ORFs).

The estimated genome size of *C. beticola* (37 Mb) is comparable to that of multiple *Cercospora* species, including *C. canescens* (~34 Mb; Chand *et al.* 2015), *C. cf. sigesbeckiae* (~35 Mb; Albu *et al.* 2017), and *C. zeina* (~37 Mb; Wingfield *et al.* 2017). The draft genome of ICMP 21692 has already

provided the foundation for global population genetics studies of *C. beticola* using microsatellite markers and Genotyping-By-Sequencing (Vaghefi *et al.* 2017a, b). Current studies are focused on identification and characterisation of the genes responsible for sensitivity to fungicides. Availability of genomic data will provide a powerful tool for characterising the genes involved in pathogenicity.

Authors: N. Vaghefi\*, J.R Kikkert, and S.J. Pethybridge

\*Contact: Niloofar.Vaghefi@usq.edu.au



## IMA Genome-F 9F

## Draft genome sequence of *Coleophoma cylindrospora* BP6252, the fungus used to produce sulfated echinocandins FR220897 and FR220899

### INTRODUCTION

Researchers at Fujisawa Pharmaceutical (now Astellas Pharma) isolated strain BP6252 (No. 14573) from an unidentified decaying leaf from Tsushima Island, Japan and identified it as *Coelophoma empetri*. They fermented the strain to produce two water-soluble echinocandin analogues FR220897 and FR220899 (WF14573A and WF14573B) (Hori *et al.* 2004, Kanasaki *et al.* 2006) (Fig. 2). FR220897 and FR220899 are isomers of FR901379 which is used for semisynthesis of micafungin. FR901379 is produced by a different strain of *C. empetri* F-11899 (Iwamoto *et al.* 1994a, b). Differential antifungal activity of these isomers was critical to understanding the effects of the position of the homotyrosine sulfate residue on the antifungal activity (Hino *et al.* 2001, Kanasaki *et al.* 2006). Like other echinocandins, the metabolites strongly inhibited  $\beta$ -1,3-glucan synthase and exhibited potent *in vitro* activity against *Candida albicans* and *Aspergillus fumigatus*, and FR220897 was effective in mouse candidiasis models. The discovery of these echinocandin variants was significant because sulfation of the homotyrosine residue overcomes the inherent poor water-solubility that had previously impeded development of echinocandin-type of antibiotics, including echinocandin B, aculeacins, and the pneumocandins.

*Coleophoma cylindrospora* is a widespread endophyte and leaf saprobe and can be a weak pathogen of leaves and fruits of many woody plants (Sutton 1980, Wu *et al.* 1996, Polashock *et al.* 2009, Crous & Groenewald 2016). The phylogenetic affinity of the strain producing FR220897 and FR220899 was established with multiple phylogenetic marker sequences and was found to be conspecific with other strains of *C. empetri* (Yue *et al.* 2015) (Fig. 7). Subsequently, during a revision of the polyphyletic genus *Coleophoma*, *C. empetri* was found to be phylogenetically indistinct from the similar *C. cylindrospora* and was considered to be a synonym of the latter (Crous & Groenewald 2016).

The primary objective behind the sequencing the genome of *C. cylindrospora* was the identification of the gene cluster-encoding the biosynthesis of FR220897 and FR220899 (Yue *et al.* 2015). The genome sequence will be essential for identifying the mechanism of the regiospecific sulfation reaction. The draft genome also has revealed that the strain harbours an auxiliary copy of  $\beta$ -1,3-glucan synthase that may function as an echinocandin resistance gene (Yue *et al.* 2018). This draft genome will expand genomic data sets for comparative genomics of species in *Leotiomyces*, *Dermataceae*, and endophytic fungi in general.

### SEQUENCED STRAIN

Japan: *Tsushima Island*: Nagasaki Prefecture, isolated from decaying leaf, [no further information] (NBRC-NITE BP6252, Fujisawa No. 14573).

### NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The *C. cylindrospora* isolate BP6252 Whole Genome Shotgun project has been deposited in GenBank under the accession number PDLM00000000.

### MATERIALS AND METHODS

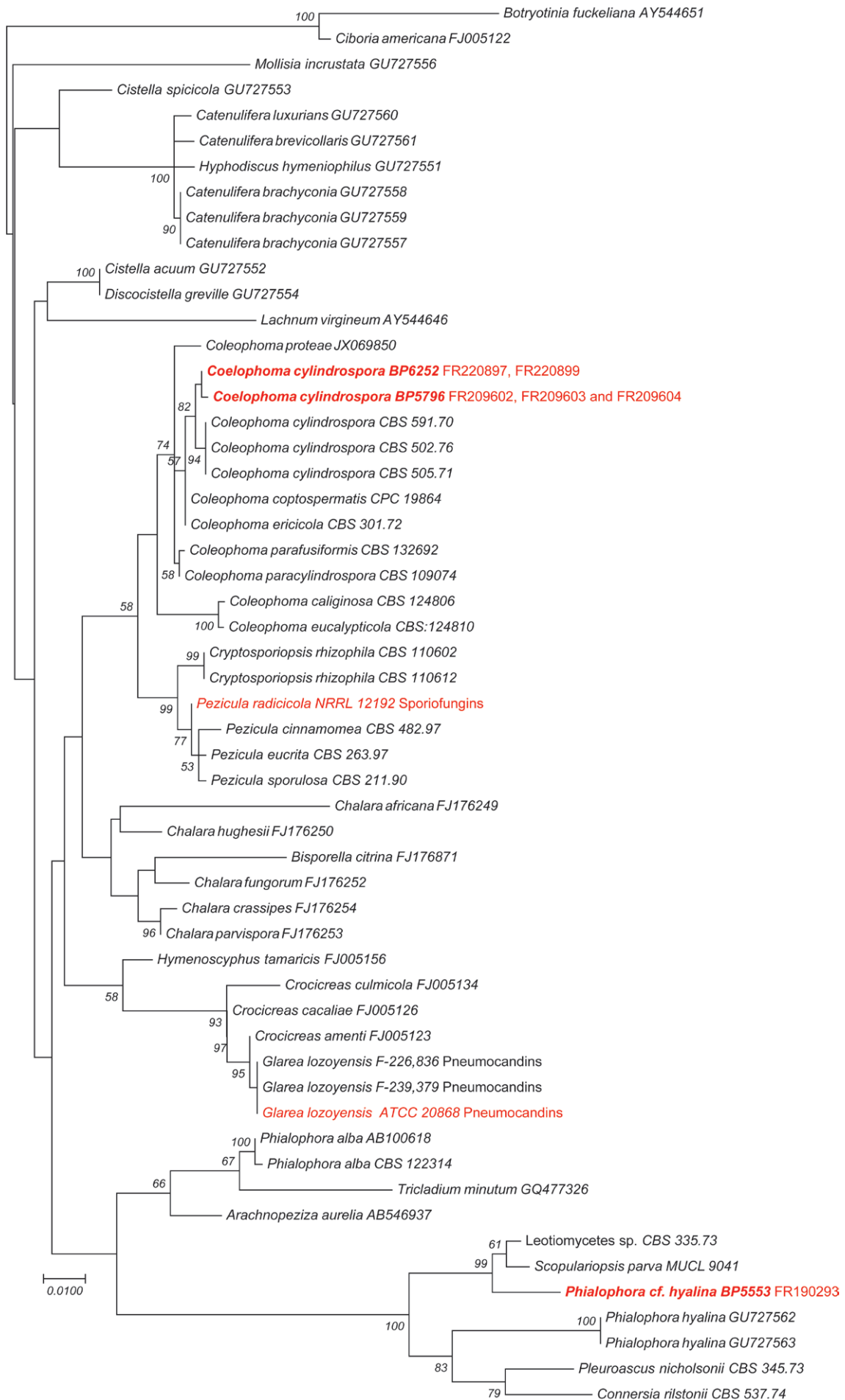
Lyophilized mycelia harvested from liquid cultures were ground in liquid nitrogen and genomic DNA was isolated by using the CTAB protocol (<http://1000.fungalgenomes.org/home/wp-content/uploads/2013/02/genomicDNAProtocol-AK0511.pdf>). A 180 bp insert library and a 5 kb mate-pair library were constructed for Illumina sequencing and were sequenced on an Illumina HiSeq2000 V4 sequencing platform (Yue *et al.* 2015). The Illumina sequencing reads were assembled using Velvet 1.2 (Zerbino & Birney 2008). *Ab initio* gene predictions from the genome assembly were made with Augustus (Stanke *et al.* 2004). Predicted genes were annotated by BLAST searches against UniProt databases (<http://www.uniprot.org/>). Polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs), dimethylallyl tryptophan synthases and related biosynthetic gene clusters were predicted by antiSMASH ver. 3.0 and manual annotation (Weber *et al.* 2015).

### RESULTS AND DISCUSSION

The genome of BP6252 was sequenced to 100-fold coverage, yielding 77 scaffolds with N50 of 2.3 megabases (Mb). The assembled genome size was 42.4 Mb, and a total of 14,177 genes were predicted. The GC content of this genome is 48.7%. The genome contains 26 core catalytic genes associated with putative secondary metabolite biosynthetic gene clusters. These clusters include 15 PKSs, eight NRPSs, two dimethylallyl tryptophan synthases, and one terpene synthase. These genes are distributed among 21 putative gene clusters that also include genes encoding tailoring enzymes, regulators, transporters, and other auxiliary genes. In addition to these gene clusters, nine secondary metabolite gene clusters containing PKS-like or NRPS-like enzyme genes, or other secondary metabolic-related genes were identified by antiSMASH. In addition a gene cluster containing close orthologues of the pneumocandin gene cluster from *Glarea lozoyensis* (Yue *et al.* 2015) was recognized and predicted to be responsible for the biosynthesis of FR220897 and FR220899.

Authors: Q. Yue, Y. Li, L. Chen, X. Zhang, K. Li, J. Sun, X. Liu, Z. An, and G.F Bills\*

\*Contact: billsge@vt.edu



## IMA Genome-F 9G

## Draft genome sequence of *Coleophoma cylindrospora* BP5796, the fungus used to produce the echinocandin variants FR209602 and related echinocandins

### INTRODUCTION

Researchers at Fujisawa Pharmaceuticals (now Astellas Pharma) isolated strain BP5796 (Fujisawa No. 738) from an unidentified leaf sample collected in Japan and identified it as *C. crateriformis* based on comparisons in conidia dimensions with respect to the known *Coleophoma* species at the time (Wu *et al.* 1996). The strain was fermented to produce three water-soluble echinocandin analogues, designated FR209602, FR209603 and FR209604 (Fig. 2). These analogues differ from FR901379 (WF11899A) and its analogues by a substitution of threonine for serine at the peptide's third amino acid and deoxygenation of the homotyrosine residue at C-4. Like other echinocandins, these metabolites strongly inhibited activity of  $\beta$ -1,3-glucan synthase and exhibited potent *in vivo* activity against *C. albicans* and *A. fumigatus* in murine systemic infection models.

The phylogenetic affinity of the strain producing FR209602 and analogues was established with multiple phylogenetic marker sequences (Yue *et al.* 2015) (Fig. 7). Although, we had retained the original identification of *C. crateriformis* in previous work on the evolution of the echinocandin pathways, a multi-gene phylogeny indicated the strain was conspecific with other strains named as *C. empetri*. Subsequently, during a revision of the polyphyletic genus *Coleophoma*, it was noted that an authentic strain of *C. crateriformis*, the type species of the genus *Coleophoma*, was lacking, and thus, its phylogenetic affinities within the genus remained to be determined (Crous & Groenewald 2016). Because strain BP5796 appears to be phylogenetically indistinct from the similar *C. cylindrospora*, we consider it to be conspecific with the latter (Crous & Groenewald, 2016).

The primary motivation for sequencing the genome of *C. cylindrospora* BP5796 was to identify the gene cluster-encoding the biosynthesis of FR209602. The genome sequence will be essential for identification of the mechanism of the regiospecific sulfation reaction. The draft genome also has revealed, that like BP6252, the strain harbours an auxiliary copy of  $\beta$ -1,3-glucan synthase that may function as

an echinocandin resistance gene (Yue *et al.* 2018). This draft genome will expand resources for comparative genomics of species in *Dermataceae* and endophytic fungi.

### SEQUENCED STRAIN

**Japan:** *Toyama Prefecture:* Mount Tateyama, isolated from leaf sample, [no further information] (NBRC-NITE BP5796, Fujisawa No. 738).

### NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The *C. cylindrospora* isolate BP5796 Whole Genome Shotgun project has been deposited in GenBank under the accession number PDLN00000000.

### MATERIALS AND METHODS

The methods for DNA extraction, sequencing, and genome assembly and annotation were essentially the same as for strain BP6252 above.

### RESULTS AND DISCUSSION

The genome of BP5796 was sequenced to 100-fold coverage, yielding 45 scaffolds with N50 of 2.0 Mb. The assembled genome size was 40.4 Mb, and a total of 13257 genes were predicted. The GC content of this genome is 48.5 %. The genome contains 24 core catalytic genes associated with putative secondary metabolite biosynthetic gene clusters. These clusters include 15 PKSs, six NRPSs, two dimethylallyl tryptophan synthases, and one terpene synthase. These genes are distributed among 21 putative gene clusters that also include genes encoding tailoring enzymes, regulators, transporters, and other auxiliary genes. In addition to these gene clusters, seven secondary metabolite gene clusters containing PKS-like or NRPS-like enzyme genes, or other secondary metabolic-related genes were identified by antiSMASH. In addition, a gene cluster containing close orthologues of the pneumocandin gene cluster from *Glarea lozoyensis* (Yue *et al.* 2015) was recognized, and predicted to be responsible for the biosynthesis of FR206902. This gene cluster deviated from other echinocandin gene clusters by the loss of a cytochrome P450 gene orthologous to *htyF* in *A. pachycristatus* and *GLP450-1* in *Glarea lozoyensis* which

**Fig. 7.** Maximum Likelihood tree of genome-sequenced strains producing echinocandins (red) and selected strains of the *Leotiomyces* (55 strains total) based on an alignment of the ITS and 28S rDNA. *Botryotinia fuckeliana* was positioned as the outgroup. The tree was inferred by using the maximum likelihood method based on the Kimura 2-parameter model. The tree with the highest log likelihood (-4229.10) is shown. The percentage of trees in which the associated taxa clustered together is labelled on branch nodes. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.4881)). Branch lengths were measured in the number of substitutions/site. All positions containing gaps and missing data were eliminated. *The dataset included 955 positions. Data were analyzed in MEGA7 (Kumar et al. 2016).*

would account for the absence of the hydroxylation at the homotyrosine C4.

Authors: Q. Yue, Y. Li, L. Chen, X. Zhang, K. Li, J. Sun, X. Liu, Z. An, and G.F. Bills\*  
\*Contact: billsge@vt.edu

## IMA Genome-F 9H

# Draft genome sequence of *Fusarium fracticaudum*

## INTRODUCTION

The genus *Fusarium* contains numerous well-known socio-economically important fungi (Nelson *et al.* 1983). Many of these fungi form part of the *Fusarium fujikuroi* Species Complex (Geiser *et al.* 2013) for which various whole genome sequences have been published, e.g. *Fusarium fujikuroi* (Jeong *et al.* 2013, Wiemann *et al.* 2013, Chiara *et al.* 2015), *Fusarium temperatum* (Wingfield *et al.* 2015b) and *Fusarium circinatum* (Wingfield *et al.* 2012, van der Nest *et al.* 2014a). The latter causes pitch canker, which is a devastating disease of pine (Wingfield *et al.* 2008). Of the five other species found to be associated with *F. circinatum*-like symptoms on pine in Colombia (Herron *et al.* 2015), the genome of *F. pininemorale* has been sequenced (Wingfield *et al.* 2017). In this study, we determined the whole genome sequence for *F. fracticaudum*, which was also described by Herron *et al.* (2015). Like *F. pininemorale*, this species does not seem to be a pathogen of pine as it could not incite lesions on the stems of pine seedlings in standard pathogenicity assays (Herron *et al.* 2015). These differences between *F. circinatum* and these non-pathogenic *Fusarium* species on *Pinus* will provide an opportunity for genome comparisons.

The association of *F. fracticaudum* with diseased pines and the genetic basis of biological traits in *F. fracticaudum* is not yet understood. Availability of various sequenced genomes of species within the FFSC is enabling studies into the biology and evolution of these fungi (Ma *et al.* 2013, De Vos *et al.* 2014, Niehaus *et al.* 2016). Here we determine the genome sequence of *F. fracticaudum*, which will provide an additional resource for comparative genomic studies aimed at understanding the evolution of these fungi and unravelling the molecular basis of their plant interactions.

## SEQUENCED STRAIN

**Colombia:** Angela Maria, Santa Rosa Risalda, 75°36'21" W 4°49'18" N, isolated from diseased *Pinus maximinoi* trees (CMWF25245; FCC5385; CBS137234; Herron *et al.* 2015).

## NUCLEOTIDE ACCESSION NUMBER:

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession PDNT00000000. The version described in this paper is version PDNT01000000.

## METHODS

### Genome sequence

The *F. fracticaudum* isolate was grown on ½ Potato Dextrose agar (PDA; BD Difco™) at 25 °C for 7 d. Genomic DNA was extracted from fungal mycelium following the protocol of Möller *et al.* (1992). Genome sequencing was done with one paired-end (350 bp median insert size) and one mate-pair (5 kb median insert size) library using Illumina HiSeq XTen and HiSeq2000 platforms respectively, at Macrogen (Seoul, Korea). CLC Genomics Workbench v. 8.0.1 (CLCBio, Aarhus Denmark) was used to trim sequences less than 18 bp. The quality filtered reads were subjected to *de novo* assembly in ABySS v. 1.3.7 (Simpson *et al.* 2009), followed by scaffolding with SSPACE v 2.0 (Boetzer *et al.* 2011). The gaps within the sequences were closed using Gapfiller v. 1.11 (Boetzer & Pirovano 2012). To determine the completeness of the genome assembly, BUSCO v2.0.1 (Benchmarking Universal Single Copy Orthologs; Simão *et al.* 2015) was employed using the sordariomycete dataset. Scaffolds were compared to those of the chromosomes of *F. fujikuroi* (Wiemann *et al.* 2013) and *F. temperatum* (Wingfield *et al.* 2015b) using the LASTZ plugin (Harris 2007) of Geneious v 7.0.4 (Kearse *et al.* 2012). WebAUGUSTUS (Hoff & Stanke 2013) was used to predict genes using the *Fusarium graminearum* model (<http://bioinf.uni-greifswald.de/augustus>) and the cDNA data from the *F. circinatum* genome (Wingfield *et al.* 2012) as gene evidences.

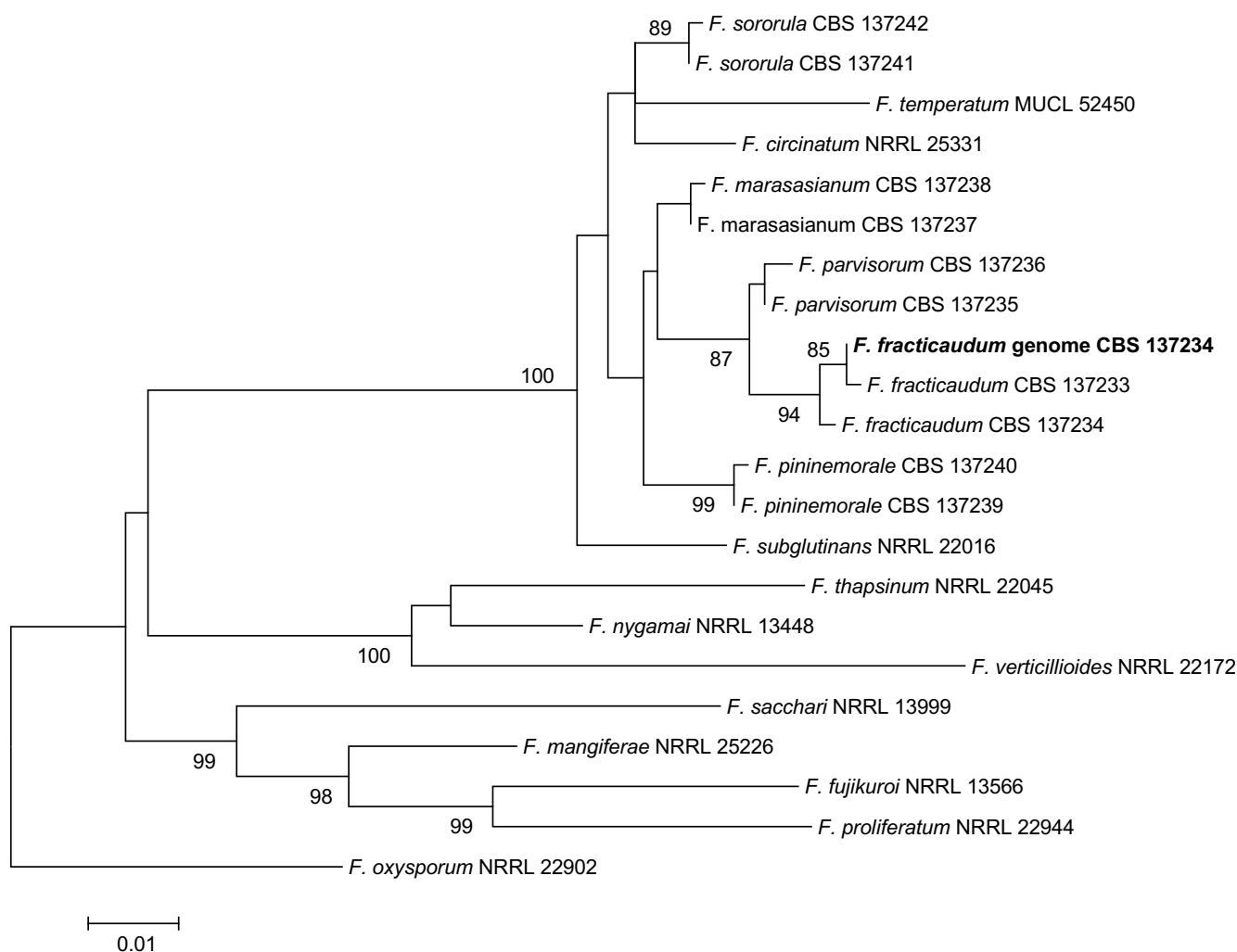
### Phylogenetic analysis

Phylogenetic analysis was conducted using partial sequences of the elongation factor 1- $\alpha$  (ef1- $\alpha$ ) and beta-tubulin genes from other species in the *Fusarium fujikuroi* species complex (Herron *et al.* 2015), including the genome of *F. fracticaudum* determined here. All gene sequences were aligned using MAFFT (Katoh *et al.* 2009). A maximum likelihood phylogenetic analysis was carried out in PhyML v 3.1 (Guindon *et al.* 2010) using the GTR+I+G substitution model with 1000 bootstraps, as determined using jModelTest v 2.1.7 (Darriba *et al.* 2012).

## RESULTS AND DISCUSSION

The assembled genome of *F. fracticaudum* was 46.29 Mb long with a GC content of 47.6 %. The assembly consisted of 50 scaffolds with an N50 value of 4 491 441 bp. WebAUGUSTUS predicted a total of 14 729 open reading frames (ORFs) in the assembly. Based on the BUSCO results, the assembly was 98.8 % complete (i.e., complete and single-copy BUSCOs = 97.6 %; complete and duplicated BUSCOs = 1.2 %; fragmented BUSCOs = 0.9 %; missing BUSCOs = 0.3 %; number of BUSCOs searched = 3725). The phylogeny inferred using two protein-coding genes also shows the previously reported relationships among the FFSC species included (Fig. 8.) (O'Donnell *et al.* 2000, Geiser *et al.* 2005, Herron *et al.* 2015). The sequences extracted from *F. fracticaudum* also grouped with those of another isolate (CBS 137233) and the original GenBank accessions for the isolate sequenced here.





**Fig. 8.** A Maximum Likelihood phylogeny showing the placement of the *F. fracticaudum* isolate (indicated in bold) that was sequenced in this study. The tree was inferred from combined  $\beta$ -tubulin and translation elongation factor 1- $\alpha$  gene sequences (Herron *et al.* 2015). Values at branch nodes are the bootstrapping confidence values with those  $\geq 85\%$  shown. The scale bar indicates substitution per site.

In terms of overall genome statistics, the whole genome sequence of *F. fracticaudum* is similar to those reported for *F. pininemorale*, *F. circinatum*, and *F. temperatum* (Table 3). Also, *F. fracticaudum* contained the reciprocal translocation between chromosome 8 and 11 known in these fungi (De Vos *et al.* 2014). However, sequence comparisons showed that chromosome 12, which is dispensable in other members of the FFSC (Xu *et al.* 1995), is 1 094 708 bp in size in *F. fracticaudum*. This is considerably larger than the 692 922 bp reported for *F. fujikuroi* (Wiemann *et al.* 2013), 986 231 bp in *F. temperatum* (Wingfield *et al.* 2015b), 791 442 bp in *F. nygamai* (Wingfield *et al.* 2015a) and 968 722 bp reported in *F. pininemorale* (Wingfield *et al.* 2017). The differences observed in these genomes highlight the importance of sequencing the genomes of additional species in the FFSC. The *F. fracticaudum* sequenced here, together with those of other FFSC species will undoubtedly provide a platform to answer numerous questions pertaining to the evolutionary history of these fungi and their species-specific traits.

Authors: **B.S Swalarsk-Parry\***, **E.T. Steenkamp**,  
**S. van Wyk**, **T.A Duong**, **B. D. Wingfield**, and **L. De Vos**  
\*Contact: benny.swalarsk@fab.up.ac.za

## IMA Genome-F 9I

### Draft genome sequence of *Phialophora cf. hyalina* BP5553, a fungus that produces the sulfated pneumocandin FR190293

#### INTRODUCTION

Researchers at Fujisawa Pharmaceuticals (now Astellas Pharma) isolated strain BP5553 (Fujisawa No. 16616) from soil collected in Japan and identified it as *Tolyposcladium parasiticum*. They fermented the strain to produce the water-soluble echinocandin analogue FR190293 (Fig. 2). Like other echinocandins, FR190293 strongly inhibited  $\beta$ -1,3-glucan synthase and exhibited potent *in vitro* activity against *Candida albicans* and *Aspergillus fumigatus*. The discovery of this new echinocandin variant was significant because it is the first of the echinocandins to have a dimethyl myristic acid acyl side chain, as in the pneumocandins, in combination with a sulfated homotyrosine residue.

**Table 3.** Genome statistics for *F. fracticaudum* and its close relatives.

	<i>F. fracticaudum</i>	<i>F. pininemorale</i> <sup>1</sup>	<i>F. circinatum</i> <sup>2</sup>	<i>F. temperatum</i> <sup>3</sup>
Genome size (Mb)	46.29	47.83	43.43	45.46
GC content (%)	47.6	46.0	47.4	47.0
Predicted orfs <sup>4</sup>	14 729	14 640	15 056	14 284
Average gene length (bp)	1531	1472	1312	1575
Gene density (ORFs/Mb)	318	306	347	314

<sup>1</sup>Wingfield *et al.* (2017); <sup>2</sup>Wingfield *et al.* (2012); <sup>3</sup>Wingfield *et al.* (2015b); <sup>4</sup>Open reading frames.

As previously reported, in-depth phylogenetic and morphological analysis of BP5553 demonstrated that the identification as the rotifer parasite *T. parasiticum* (syn. *Pochonia parasitica*) was erroneous. Rather than belonging to *Clavicipitaceae*, BP5553 was found to belong in *Helotiales* (Yue *et al.* 2015). Based on rDNA and other protein-encoding sequences, BP5553 falls within a monophyletic lineage along with *ex*-type strains of *Phialophora hyalina*, *Pleuroascus nicholsonii*, *Scopulariopsis parva*, and *Scopulariopsis parvula* (Fig. 7). These strains, along with other species with *Phialophora*-like conidial morphs in *Helotiales* and BP5553 will eventually comprise a new genus in a new family of *Helotiales* (W. Untereiner *et al.*, unpubl.).

The primary objective behind the sequencing the genome of *Phialophora* cf. *hyalina* was the identification of the gene cluster-encoding the biosynthesis of FR190293 (Yue *et al.* 2015). This draft genome will expand genomic data sets for comparative genomics of species in *Leotiomycetes* and *Helotiales*.

## SEQUENCED STRAIN

**Japan:** *Fukushima Prefecture:* Iwaki, isolated from soil, [no further information] (NBRC-NITE BP5553, Fujisawa No. 16616).

## NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The *Phialophora* cf. *hyalina* isolate BP5553 Whole Genome Shotgun project has been deposited in GenBank under the accession number NPIC00000000.

## MATERIALS AND METHODS

The methods for DNA extraction, sequencing, and genome assembly were essentially the same as for strain BP6252 above.

## RESULTS AND DISCUSSION

The genome of BP5553 was sequenced to 102-fold coverage, yielding 32 scaffolds with N50 of 3.8 Mb. The assembled genome size was 33.6 Mb, and a total

of 10 707 genes were predicted. The GC content of this genome is 48.2 %. The genome contains 45 core catalytic genes associated with putative secondary metabolite biosynthetic gene clusters. These clusters include 19 PKSs, 13 NRPSs, six PKS-NRPS hybrids, two dimethylallyl tryptophan synthases, four terpene synthases, and one chalcone synthase. These genes are distributed among 40 putative biosynthetic gene clusters that also include genes encoding tailoring enzymes, regulators, transporters, and other auxiliary genes. In addition to these gene clusters, eight secondary metabolite gene clusters containing PKS-like or NRPS-like enzyme genes, or other secondary metabolic-related genes were identified by antiSMASH. In addition a gene cluster containing close orthologues of the pneumocandin gene cluster from *Glarea lozoyensis* (Yue *et al.* 2015) was recognized, and predicted to be responsible for the biosynthesis of FR190293.

*Authors:* Q. Yue, Y. Li, L. Chen, X. Zhang, K. Li, J. Sun, X. Liu Z. An, and G.F Bills\*

\*Contact: billsge@vt.edu

## IMA Genome-F 9J

### Draft genome sequence of *Morchella septimelata*

## INTRODUCTION

Morels (*Morchella* spp., *Ascomycota*) are a highly desired group of edible fungi with a worldwide distribution (Du *et al.* 2012, Kanwal *et al.* 2011, Richard *et al.* 2015). They have been collected by mycophiles and gourmets for hundreds of years for their delicate taste and unique appearance (Tietel & Masaphy 2017, 2018, Rotzoll *et al.* 2006). Morels are also found containing a variety of secondary metabolites with medicinal properties (Tietel & Masaphy 2018, Shameem *et al.* 2017, Pfab *et al.* 2008, Vieira *et al.* 2016). *Morchella septimelata* is a black morel, belonging to the *Morchella elata* clade (Kuo *et al.* 2012). It was often found in lightly to moderately burned conifer forests, near creek beds, springs and seeps, at an altitude of 1000–2000 m (Pildain *et al.* 2014). The ascomata of *M. septimelata* can be found primarily in years immediately following forest fires, and then often appearing in dwindling numbers for several seasons

thereafter (Hobbie *et al.* 2016). In recent years, artificial cultivation of true morels has made great progress (Masaphy 2010); several *Morchella* species, such as *M. sextelata*, *M. septimelata*, and *M. importuna*, have been successfully cultivated in China, America, and other parts of the world. However, the mechanism of the growth and development of *Morchella* remains unclear, which causes the frequent failure and unstable yield of *Morchella* cultivation (Liu *et al.* 2018).

The genome sequence of *M. septimelata* from this study may reveal the mechanism of secondary metabolites synthesis in *M. septimelata* and provide some insights into the growth, development, and carbohydrate degradation of *M. septimelata*.

## SEQUENCED STRAIN

**China:** Sichuan: Liangshan Yi, N 27°49' E 100°48', alt. 1468 m, isolated from forest soil, 19 Sep. 2015, *Chuan Xiong & Qiang Li* (MG91-dried culture).

## NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The Whole Genome Shotgun project *M. septimelata* isolate (Culture collection number SAAS91) has been deposited at DBJ/EMBL/GenBank under the accession number PYSJ00000000. The version described in this paper is version PYSJ01000000.

## MATERIALS AND METHODS

*Morchella septimelata* MG91 was isolated from forest soil in Liangshan Yi Autonomous Prefecture, Sichuan, China, and was preserved in the Fungal Culture Collection Center of Biotechnology and Nuclear Technology Research Institute (Chengdu, Sichuan). Genomic DNA was extracted from MG91 and subjected to sequencing on the Genome Analyzer Iix next-generation sequencing platform (Illumina) at the BGI (Shenzhen, China). Paired-end libraries with respective insert sizes of 425 bp and 725 bp were used to generate read lengths of 150 bases. The CLC Genomics Workbench v. 6.0.1 (CLCBio, Aarhus, Denmark) was subsequently used to trim reads of poor quality (limit of 0.05) as well as terminal nucleotides. The remaining reads were assembled using the SPAdes 3.0.0 with an optimized k-mer value of 21 (Bankevich *et al.* 2012). Thereafter, scaffolding was completed using SSPACE v. 2.0 (Boetzer *et al.* 2011) and gaps reduced with the use of GapFiller v. 2.2.1 (Boetzer & Pirovano 2012). The completeness of the assembly was evaluated using the BUSCO v3 (Simão *et al.* 2015).

Homology-based gene prediction and *ab initio* prediction were performed to search *M. septimelata* gene models. Homologous protein from *Tuber melanosporum* was used for alignment to the repeat-masked *M. septimelata* genome using Exonerate v 2.2.0 (Slater & Birney 2005). The filtered alignment results (above 300 bp and 90 % coverage) were built as training models for *ab initio* gene prediction. The

*ab initio* prediction was conducted using Augustus v. 3.2.3 (Stanke *et al.* 2008) and GeneMark-ES (Ter-Hovhannisyann *et al.* 2008) guided by training models from homology-based alignments. All gene prediction results were integrated into the final gene models by EVIDENCEModeler (Haas *et al.* 2008). Carbohydrate-active enzymes (CAZyme), including the repertoire of auxiliary enzymes, were predicted using dbCAN (Yin *et al.* 2012).

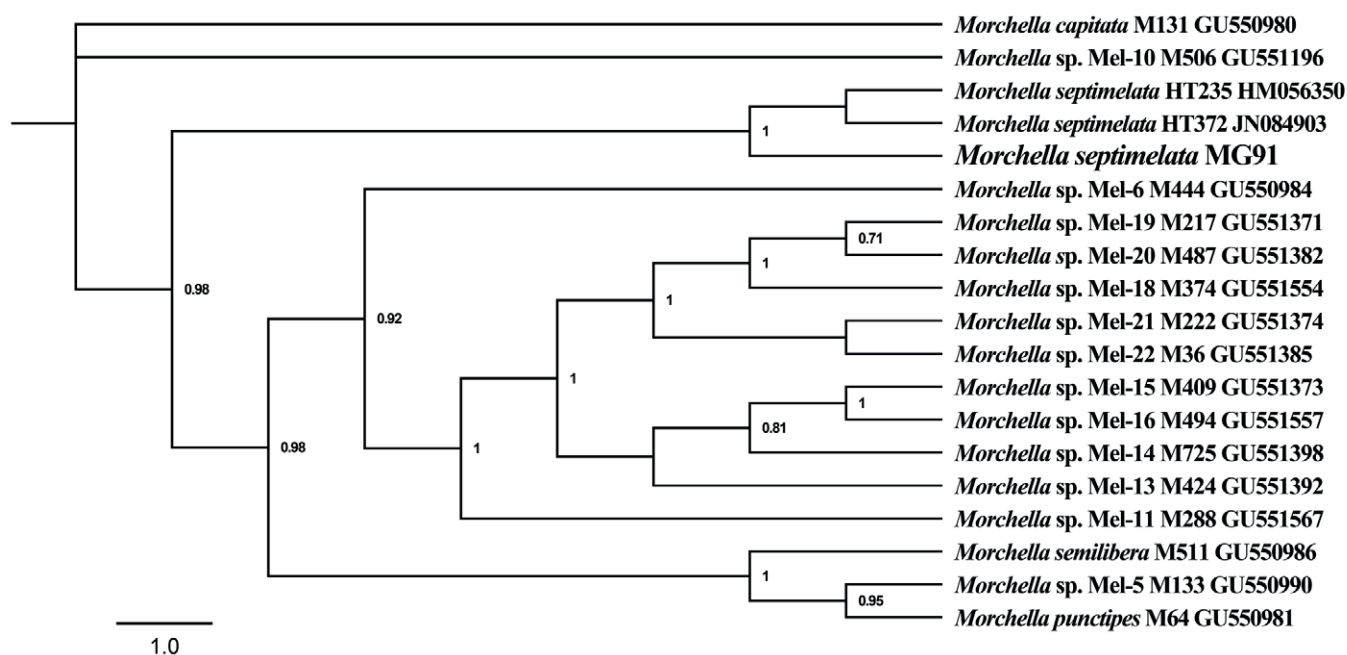
To verify the species identity of the sequenced strain, the Translation Elongation Factor 1-alpha gene for selected *Morchella* species (Fig. 9.) were aligned with mafft (Katoch & Standley 2013). The Bayesian inference (BI) method (Erixon *et al.* 2003) was used to construct the phylogenetic tree of different *Morchella* species. JMODELTEST 2.0.2 was used to ascertain the best-fit model for nucleotide substitutions (Darriba *et al.* 2012). BI analysis was performed with MrBayes v3.2.6 (Ronquist *et al.* 2012). Two independent runs with four chains (three heated and one cold) each were conducted simultaneously for  $2 \times 10^6$  generations. Each run was sampled every 100 generations. We assumed that stationarity had been reached when estimated sample size (ESS) was greater than 100, and the potential scale reduction factor (PSRF) approached 1.0. The first 25 % samples were discarded as burn-in, and the remaining trees were used to calculate Bayesian posterior probabilities (BPP) in a 50 % majority-rule consensus tree.

## RESULTS AND DISCUSSION

The genome of *M. septimelata* had an estimated size of 49.81 Mb with an average coverage of 151.17 times (Table 4). The Scaffold N50 size was 37 734 bases, and the assembly had a mean GC content of 47.40 %. The total number of scaffold generated was 6525. A total of 11 427 genes were predicted with an average length of 1 571 bp. A phylogenetic analysis of the genus *Morchella* is provided to show position of *M. septimelata* (Fig. 9).

The draft genome of *M. septimelata* is larger than that of the closely related species, *M. conica* CCBAS932 (JGI: 1023999) and *M. importuna* SCYDJ1-A1 (JGI: 1047732), which are 48.21 Mb and 48.80 Mb, respectively. Less gene models are found in *M. septimelata* compared to that of these closely related species, *M. conica* CCBAS932 (11 600 gene models) and *M. importuna* SCYDJ1-A1 (11971 gene models). The average gene length of *M. septimelata* is also smaller than that of *M. conica* CCBAS932 (1668 bp) and *M. importuna* SCYDJ1-A1 (1625 bp). The average gene density of *M. septimelata* was 229 genes/Mb, which is smaller than that of *M. conica* CCBAS932 (240 genes/Mb) and *M. importuna* SCYDJ1-A1 (245 genes/Mb).

A total 512 CAZymes were identified in the genome of *M. septimelata*, which is more than that of the closely related species, *M. conica* CCBAS932 (401 CAZymes) and *M. importuna* SCYDJ1-A1 (403 CAZymes), indicating that the carbohydrate degradation ability of *M. septimelata* may be stronger than that of the other two closely related species. A total of 9 secondary metabolite (sM) clusters were found in the *M. septimelata* genome, of which 3 sM clusters were for terpenes. The genome sequence data of *M. septimelata*



**Fig. 9.** A Bayesian inference (BI) phylogenetic analysis of genus *Morchella* using MrBayes v3.2.6 based on partial gene sequences of elongation factor 1-alpha (EF1- $\alpha$ ) gene. Posterior probabilities are shown on the nodes of the tree. The *Morchella septimelata* isolate used for verification was extracted from the assembled genomes. Reference sequences are obtained from the NCBI database with accession number.

**Table 4.** Genome statistics, CAZyme richness and secondary metabolite clusters for the *Morchella septimelata* MG91 genome sequence.

Genome	<i>M. septimelata</i> MG91
Coverage	151.17x
BUSCO	97.7%
Total sequence length (Mb)	49.81
Scaffolds	6 525
Scaffold N50 (bp)	37 734
GC (%)	47.40
Predicted gene models	11 427
Average gene length (bp)	1 571
Average gene density (genes/Mb)	229
<b>Predicted CAZymes</b>	
Total CAZymes	512
Auxiliary activities	75
Pectate lyases	23
Glycosyl transferases	75
Glycoside hydrolases	201
Carbohydrate esterases	72
Carbohydrate binding motifs	66
<b>Predicted secondary Metabolite (sM) Clusters</b>	
Total SM clusters	9
Terpene clusters	3
Type I polyketide synthetases (PKSs)	1
Nonribosomal peptide synthetases (NRPSs)	1
Others	4

presented in this study will provide useful information for understanding the synthesis mechanism of secondary metabolites in *M. septimelata* and lay a foundation for the artificial cultivation of *M. septimelata*.

Authors: Q. Li, C. Xiong, X. Ma, H. Li, Y. Gao, Y. Dong\*, and W. Huang\*

\*Contact: loyalyang@163.com; wenlih11@126.com

## ACKNOWLEDGEMENTS

The authors of the research on the *Berkeleyomyces basicola*, *Ceratocystis smalleyi*, and *Fusarium fructicaudum* genomes thank the University of Pretoria, the Tree Protection Co-operative Programme (TPCP), the DST-NRF Centre of Excellence in Tree Health Biotechnology, the National Research Foundation, the Genomics Research Institute (GRI) and the DST-NRF SARCHI chair on Fungal Genomics for their financial support.

The research on *Cercospora beticola* was supported by the United States Department of Agriculture, National Institute of Food and Agriculture Hatch project NYG-625424, and the Director's Controlled Endowment Fund of The New York Agricultural Experiment Station, Cornell University, Geneva, New York, USA.

The research on the genomes of *Coleophoma cylindrospora*, *Phialophora cf. hyalina*, and *Aspergillus mulundensis* was supported by the National Natural Science Foundation of China (NSFC) grants 31629001 to Z.A., 31741004 to Q.Y., Welch Foundation Grant AU-0042-20030616 to ZA, and the Kay and Ben Fortson Endowment to GFB.



## REFERENCES

- Afgan E, Baker D, Van den Beek M, Blankenberg D, Bouvier D, *et al.* (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research* **44**: W3–W10.
- Al Adawi AO, Barnes I, Khan IA, Al Subhi AM, Al Jahwari AA, *et al.* (2013) *Ceratocystis manginecans* associated with a serious wilt disease of two native legume trees in Oman and Pakistan. *Australasian Plant Pathology* **42**: 179–193.
- Albu S, Sharma S, Bluhm BH, Price PP, Schneider RW, *et al.* (2017) Draft genome sequence of *Cercospora cf. sigesbeckiae*, a causal agent of Cercospora leaf blight on soybean. *Genome Announcements* **5**: e00708–17.
- Al-Subhi AM, Al-Adawi AO, Van Wyk M, Deadman ML, Wingfield MJ (2006) *Ceratocystis omanensis*, a new species from diseased mango trees in Oman. *Mycological Research* **110**: 237–245.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.
- Berkeley MJ, Broome CE (1850) Notices of British fungi. *Annals and Magazine of Natural History*, ser. 2 **5**: 455–465.
- Bills GF, Yue Q, Chen L, Li Y, An Z, *et al.* (2016) *Aspergillus mulundensis* sp. nov., a new species for the fungus producing the antifungal echinocandin lipopeptides, mulundocandins. *Journal of Antibiotics* **69**: 141–148.
- 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 Biology* **13**: R56.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Cantarel BL, Korf I, Robb SM, Parra G, Ross E, *et al.* (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**: 188–196.
- Chand R, Pal C, Singh V, Kumar M, Singh VK, *et al.* (2015) Draft genome sequence of *Cercospora canescens*: a leaf spot causing pathogen. *Current Science* **109**: 2103.
- Chen AJ, Frisvad JC, Sun BD, Varga J, Kocsubé S, *et al.* (2016) *Aspergillus* section *Nidulantes* (formerly *Emericella*): Polyphasic taxonomy, chemistry and biology. *Studies in Mycology* **84**: 1–118.
- Chiara M, Fanelli F, Mule G, Logrieco AF, Pesole G, *et al.* (2015) Genome sequencing of multiple isolates highlights subtelomeric genomic diversity within *Fusarium fujikuroi*. *Genome Biology and Evolution* **7**: 3062–3069.
- Crous, PW, Groenewald, JZ (2016) They seldom occur alone. *Fungal Biology* **120**: 1392–1415.
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772.
- Daub ME, Ehrenshaft M (2000) The photoactivated *Cercospora* toxin cercosporin: contributions to plant disease and fundamental biology. *Annual Review of Phytopathology* **38**: 461–490.
- De Beer ZW, Duong TA, Barnes I, Wingfield BD, Wingfield MJ (2014) Redefining *Ceratocystis* and allied genera. *Studies in Mycology* **79**: 187–219.
- De Beer ZW, Marinowitz S, Duong TA, Wingfield MJ (2017) *Bretziella*, a new genus to accommodate the oak wilt fungus, *Ceratocystis fagacearum* (*Microascales*, *Ascomycota*). *MycKeys* **27**: 1–19.
- De Vos L, Steenkamp ET, Martin SH, Santana QC, Fourie G, *et al.* (2014) Genome-wide macrosynteny among *Fusarium* species in the *Gibberella fujikuroi* complex revealed by amplified fragment length polymorphisms. *PLoS ONE* **9**: e114682.
- Deng Y, van Peer AF, Lan FS, Wang QF, Jiang Y, *et al.* (2016) Morphological and molecular analysis identifies the associated fungus (“Xianghui”) of the medicinal white jelly mushroom, *Tremella fuciformis*, as *Annulohyphoxylon stygium*. *International Journal of Medicinal Mushrooms* **18**: 253–60.
- Du XH, Zhao Q, O'Donnell K, Rooney AP, Yang ZL (2012) Multigene molecular phylogenetics reveals true morels (*Morchella*) are especially species-rich in China. *Fungal Genetics and Biology* **49**: 455–469.
- 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.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.
- Engelbrecht CJ, Harrington TC, Alfenas A (2007) *Ceratocystis* wilt of cacao – a disease of increasing importance. *Phytopathology* **97**: 1648–1649.
- Engelbrecht CJB, Harrington TC (2005) Intersterility, morphology and taxonomy of *Ceratocystis fimbriata* on sweet potato, cacao and sycamore. *Mycologia* **97**: 57–69.
- English AC, Richards S, Han Y, Wang M, Vee V *et al.* (2012) Mind the gap: upgrading genomes with Pacific Biosciences RL long-read sequencing technology. *PLoS ONE* **7**: e47768.
- Erixon P, Svennblad B, Britton T, Oxelman B (2003) Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic Biology* **52**: 665–673.
- Floudas D, Binder M, Riley R, Barry K, Blanchette RA, *et al.* (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**: 1715–1719.
- Fourie A, Wingfield MJ, Wingfield BD, Barnes I (2014) Molecular markers delimit cryptic species in the *Ceratocystis fimbriata sensu lato* complex. *Mycological Progress* **14**: 1–18.
- Franc GD (2010) Ecology and epidemiology of *Cercospora beticola*. In: *Cercospora Leaf Spot of Sugar Beet and Related Species* (RT Lartey, JJ Weiland, L Panella, PW Crous & CE Windels, eds): 7–19. St Paul, MN: Americal Phytopathological Society Press.
- Geiser DM, Aoki T, Bacon CW, Baker SE, Bhattacharyya MK, *et al.* (2013) One fungus, one name: defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology* **103**: 400–408.
- Geiser DM, 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.
- Groenewald JZ, Nakashima C, Nishikawa J, Shin HD, Park JH, *et al.* (2013) Species concepts in *Cercospora*: spotting the weeds among the roses. *Studies in Mycology* **75**: 115–170.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**: 307–321.
- Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, *et al.* *Automated eukaryotic gene structure* annotation using EVIDENCEModeler and

- the program to assemble spliced alignments. *Genome Biology* **11**: R7.
- Halsted BD, Fairchild DG (1891) Sweet potato black rot. *Journal of Mycology* **7**: 1–11.
- Harrington TC (2009) The genus *Ceratocystis*. Where does the oak wilt fungus fit? In: *Proceedings of the 2nd national Oak Wilt symposium* (Appel DN, et al. eds): 1–16. Austin, TX: USDA Forest Service, Forest Health Protection,.
- Harrington TC, Kazmi MR, Al-Sadi AM, Ismail SI (2014) Intraspecific and intragenomic variability of ITS rDNA sequences reveals taxonomic problems in *Ceratocystis fimbriata sensu stricto*. *Mycologia* **106**: 224–242.
- Harris RS (2007) *Improved Pairwise Alignment of Genomic DNA*. Ph thesis, Pennsylvania State University, Department of Computer Science and Engineering.
- Hawser S, Borgonovi M, Markus A, Isert D (1999) Mulundocandin, an echinocandin-like lipopeptide antifungal agent: biological activities *in vitro*. *Journal of Antibiotics* **52**: 305–310.
- Herron DA, Wingfield MJ, Wingfield BD, Rodas CA, Marincowitz S, et al. (2015) Novel taxa in the *Fusarium fujikuroi* species complex from *Pinus* spp. *Studies in Mycology* **80**: 131–150.
- Hino M, Fujie A, Iwamoto T, Hori Y, Hashimoto M, et al. (2001) Chemical diversity in lipopeptide antifungal antibiotics. *Journal of Industrial Microbiology and Biotechnology* **27**: 157–162.
- Hobbie EA, Rice SF, Weber NS, Smith JE (2016) Isotopic evidence indicates saprotrophy in post-fire *Morchella* in Oregon and Alaska. *Mycologia* **108**: 638–645.
- Hoff KJ, Stanke M (2013) WebAUGUSTUS - a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research* **41**: 123–128.
- Ho, Y, Takase S, Tsurumi Y, Hashimoto M, Hino M (2004) WF14573 or its salt, production thereof and use thereof. *US Patent* **6** 730–776.
- Hsieh HM, Ju YM, Rogers JD (2005) Molecular phylogeny of *Hypoxylon* and closely related genera. *Mycologia* **97**: 844–865.
- Huang Q, Wang YY, Zhao YY, Jiao YX, Li XF, et al. (2008) First report of taro black rot caused by *Ceratocystis fimbriata* in China. *Plant Pathology* **57**: 780–780.
- Iwamoto T, Fujie A, Nitta K, Hashimoto S, Okuhara M, et al. (1994a) WF11899A, B and C, novel antifungal lipopeptides – II. Biological properties. *Journal of Antibiotics* **47**: 1092–1097.
- Iwamoto T, Fujie A, Sakamoto K, Tsurumi Y, Shigematsu N, et al. (1994b) WF11899A, B and C, novel antifungal lipopeptides - I. taxonomy, fermentation, isolation and physico-chemical properties. *Journal of Antibiotics* **47**: 1084–1091.
- Jeong H, Lee S, Choi GJ, Lee T, Yun S-H (2013) Draft genome sequence of *Fusarium fujikuroi* B14, the causal agent of the bakanae disease of rice. *Genome Announcements* **1**: e00035–00013.
- Johnson J (1916) Host plants of *Thielavia basicola*. *Journal of Agricultural Research* **7**: 289–300.
- Johnson JA, Harrington TC, Engelbrecht CJB (2005) Phylogeny and taxonomy of the North American clade of the *Ceratocystis fimbriata* complex. *Mycologia* **97**: 1067–1092.
- Juzwik J, Park J-H, Haugen L (2010) Hickory decline and mortality: update on hickory decline research. In: *Iowa's Forest Health Report 2010* (Feeley T, ed.): 53–58. Des Moines, IA.
- Kanasaki R, Abe F, Kobayashi M, Katsuoka M, Hashimoto M, et al. (2006) FR220897 and FR220899, Novel antifungal lipopeptides from *Coleophoma empetri* No. 14573. *Journal of Antibiotics* **59**: 149–157.
- Kanwal HK, Acharya K, Ramesh G, Reddy MS (2011) Molecular characterization of *Morchella* species from the Western Himalayan region of India. *Current Microbiology* **62**: 1245–1252.
- Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. *Methods in Molecular Biology* **537**: 39–64.
- Katoh K, Standley D (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.
- Kearse M, Moir R, Wilson A, Stones–Havas S, Cheung M, et al. (2012) Geneious Basic: a integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.
- Keller O, Kollmar M, Stanke M, Waack S (2011) A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**: 757–763.
- Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* **5**: 59.
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**: 1870–1874.
- Kuo M, Dewsbury DR, O'Donnell K, Carter MC, Rehner SA, et al. (2012) Taxonomic revision of true morels (*Morchella*) in Canada and the United States. *Mycologia* **104**: 1159–1177.
- Lal B, Gund VG, Bhise NB, Gangopadhyay AK (2004) Mannich reaction: an approach for the synthesis of water soluble mulundocandin analogues. *Bioorganic and Medicinal Chemistry* **12**: 1751–1768.
- Lal B, Gund VG, Gangopadhyay AK, Nadkarni SR, Dikshit DK, et al. (2003) Semisynthetic modifications of hemiaminal function at ornithine unit of mulundocandin, towards chemical stability and antifungal activity. *Bioorganic and Medicinal Chemistry* **11**: 5189–5198.
- Lee D-H, Roux J, Wingfield BD, Barnes I, Mostert L, et al. (2016) The genetic landscape of *Ceratocystis albifundus* populations in South Africa reveals a recent fungal introduction event. *Fungal Biology* **120**: 690–700.
- Li H, Lee HS, Kim SH, Moon B, Lee C (2014) Antioxidant and anti-inflammatory activities of methanol extracts of *Tremella fuciformis* and its major phenolic acids. *Journal of Food Science* **79**: C460–468.
- 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 Disease* **100**: 2266–2274.
- Liu Q, Ma H, Zhang Y, Dong C (2018) Artificial cultivation of true morels: current state, issues and perspectives. *Critical Reviews in Biotechnology* **38**: 259–271.
- Ma LJ, Geiser DM, Proctor RH, Rooney AP, O'Donnell K, et al. (2013) *Fusarium* pathogenomics. *Annual Review of Microbiology* **67**: 399–416.
- Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764–770.
- Masaphy S (2010) Biotechnology of morel mushrooms: successful fruiting body formation and development in a soilless system. *Biotechnology Letters* **32**: 1523–1527.
- 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.
- Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, *et al.* (2017) Widespread adenine N6-methylation of active genes in fungi. *Nature Genetics* **49**: 964–968.
- Mukhopadhyay T, Ganguli BN, Fehlhaber HW, Kogler H, Vertesy L (1987) Mulundocandin, a new lipopeptide antibiotic. II. Structure elucidation. *Journal of Antibiotics* **40**: 281–289.
- Mukhopadhyay T, Roy K, Bhat RG, Sawant SN, Blumbach J, *et al.* (1992) Deoxymulundocandin – A new echinocandin type antifungal antibiotic. *Journal of Antibiotics* **45**: 618–623.
- Nehl DB, Allen SJ, Mondal AH, Lonergan PA (2004) Black root rot: a pandemic in Australian cotton. *Australasian Plant Pathology* **33**: 87–95.
- Nel WJ, Duong TA, Wingfield MJ, Wingfield BD, De Beer ZW (2017) A new genus and species for the globally important, multi-host root pathogen *Thielaviopsis basicola*. *Plant Pathology*: DOI 10.1111/ppa.12803.
- Nelson PE, Toussoun TA, Marasas WFO (1983) *Fusarium Species: an illustrated manual for identification*. University Park, PA: Pennsylvania State University Press.
- Niehaus E-M, Münsterkötter M, Proctor RH, Brown DW, Sharon A, *et al.* (2016) Comparative “omics” of the *Fusarium fujikuroi* species complex highlights differences in genetic potential and metabolite synthesis. *Genome Biology and Evolution* **8**: 3574–3599.
- 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.
- Park JH, Juzwik J, Cavender-Bares J (2013) Multiple *Ceratocystis smalleyi* infections associated with reduced stem water transport in bitternut hickory. *Phytopathology* **103**: 565–574.
- Park JH, Juzwik J, Haugen LM (2010) *Ceratocystis* canker of bitternut hickory caused by *Ceratocystis smalleyi* in the north-central and northeastern United States. *Plant Disease* **94**: 277.
- Pereg LL (2013) Black root rot of cotton in Australia: the host, the pathogen and disease management. *Crop and Pasture Science* **64**: 1112–1126.
- Pethybridge SJ, Vaghefi N, Kikkert JR (2017) Management of *Cercospora* leaf spot in conventional and organic table beet production. *Plant Disease* **101**: 1642–1651.
- Pfab R, Habert B, Kleber J, Zilker T (2008) Cerebellar effects after consumption of edible morels (*Morchella conica*, *Morchella esculenta*). *Clinical Toxicology* **46**: 259–260.
- Pildain MB, Visnovsky SB, Barroetavena C (2014) Phylogenetic diversity of true morels (*Morchella*), the main edible non-timber product from native Patagonian forests of Argentina. *Fungal Biology* **118**: 755–763.
- Piveta G, Ferreira MA, Fb Muniz M, Valdetaro D, Valdebenito-Sanhueza R, *et al.* (2016) *Ceratocystis fimbriata* on kiwi fruit (*Actinidia* spp.) in Brazil. *New Zealand Journal of Crop and Horticultural Science* **44**: 13–24.
- Polashock, JJ, Caruso, FL, Oudemans, PV, McManus, PS, Crouch, JA (2009) The North American cranberry fruit rot fungal community: a systematic overview using morphological and phylogenetic affinities. *Plant Pathology* **58**: 1116–1127.
- Richard F, Bellanger JM, Clowez P, Hansen K, O'Donnell K, *et al.* (2015) True morels (*Morchella*, *Pezizales*) of Europe and North America: evolutionary relationships inferred from multilocus data and a unified taxonomy. *Mycologia* **107**: 359–382.
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
- Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.
- Rotzoll N, Dunkel A, Hofmann T (2006) Quantitative studies, taste reconstitution, and omission experiments on the key taste compounds in morel mushrooms (*Morchella deliciosa* Fr.). *Journal of Agricultural and Food Chemistry* **54**: 2705–2711.
- Roux J, van Wyk M, Hatting H, Wingfield MJ (2004) *Ceratocystis* species infecting stem wounds on *Eucalyptus grandis* in South Africa. *Plant Pathology* **53**: 414–421.
- Roy K, Mukhopadhyay T, Reddy GCS, Desikan KR, Ganguli BN. 1987. Mulundocandin, a new lipopeptide antibiotic. I. Taxonomy, fermentation, isolation and characterization. *Journal of Antibiotics* **40**: 275–280.
- Seifert KA, De Beer ZW, Wingfield MJ (2013) *The Ophiostomatoid Fungi: expanding frontiers*. [CBS Biodiversity Series no. 12.] Utrecht: CBS–KNAW Fungal Biodiversity Centre.
- Shameem N, Kamili AN, Ahmad M, Masoodi FA, Parray JA (2017) Antimicrobial activity of crude fractions and morel compounds from wild edible mushrooms of North western Himalaya. *Microbial Pathogenesis* **105**: 356–360.
- 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.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, *et al.* (2009) ABySS: A parallel assembler for short read sequence data. *Genome Research* **19**: 1117–1123.
- Slater GS, Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**: 31.
- Somasekhara YM (1999) New record of *Ceratocystis fimbriata* causing wilt of pomegranate in India. *Plant Disease* **83**: 400.
- Stamets P (2000) Growth parameters for gourmet and medicinal mushroom species. In: *Growing Gourmet and Medicinal Mushrooms* (Stamets P, ed.): 402–405. 3<sup>rd</sup> edn. Berkeley, CA: Ten Speed Press.
- Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637–644.
- Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research* **32**: W309–W312.
- Stover RH (1950) The black rootrot disease of tobacco: I. Studies on the causal organism *Thielaviopsis basicola*. *Canadian Journal of Research* **28**: 445–470.
- Sutton BC (1980) *The Coelomycetes: fungi imperfecti with pycnidia, acervuli and stromata*. Kew: Commonwealth Mycological Institute.
- Tamura K, Stecher G, Peterson D, Filipksi A, Kumar S (2013) MEGA 6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725–2729.
- Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Research* **18**: 1979–90.
- Tietel Z, Masaphy S (2017) True morels (*Morchella*) – nutritional and phytochemical composition, health benefits and flavor: a review.



- Critical Reviews in Food Science and Nutrition*: DOI:10.1080/10408398.2017.1285269.
- Tietel Z, Masaphy S (2018) Aroma–volatile profile of black morel (*Morchella importuna*) grown in Israel. *Journal of the Science of Food and Agriculture* **98**: 346–353.
- Vaghefi N, Hay FS, Kikkert JR, Pethybridge SJ (2016) Genotypic diversity and resistance to azoxystrobin of *Cercospora beticola* on processing table beet in New York. *Plant Disease* **100**: 1466–1473.
- Vaghefi N, Kikkert JR, Bolton MD, Hanson LE, Secor GA, et al. (2017a) *De novo* genome assembly of *Cercospora beticola* for microsatellite marker development and validation. *Fungal Ecology* **26**: 125–134.
- Vaghefi N, Kikkert JR, Bolton MD, Hanson LE, Secor GA, et al. (2017b) Global genotype flow in *Cercospora beticola* populations confirmed through genotyping–by–sequencing. *PLoS One* **12**: 0186488.
- Van der Nest MA, Beirn LA, Crouch JA, Demers JE, De Beer ZW, et al. (2014a) IMA Genome-F 3: Draft genomes of *Amanita jacksonii*, *Ceratocystis albifundus*, *Fusarium circinatum*, *Huntia 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) IMA Genome-F 2: 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 Evolutionary Biology* **15**: 1–20.
- Van Wyk M, Roux J, Nkuekam GK, Wingfield BD, Wingfield MJ (2012) *Ceratocystis eucalypticola* sp. nov. from *Eucalyptus* in South Africa and comparison to global isolates from this tree. *IMA Fungus* **3**: 45–58.
- Van Wyk M, Wingfield BD, Clegg PA, Wingfield MJ (2009) *Ceratocystis larium* sp nov., a new species from *Styrax benzoin* wounds associated with incense harvesting in Indonesia. *Persoonia* **22**: 75–82.
- Vanderpool D, Bracewell RR, McCutcheon JP (2017) Know your farmer: Ancient origins and multiple independent domestications of ambrosia beetle fungal cultivars. *Molecular Ecology*: DOI:10.1111/mec.14394.
- Vieira V, Fernandes A, Barros L, Glamoclija J, Ciric A, et al. (2016) Wild *Morchella conica* Pers. from different origins: a comparative study of nutritional and bioactive properties. *Journal of the Science of Food and Agriculture* **96**: 90–98.
- Walker BJ, Abeel T, Shea T, Priest M, Abouselliel A, 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 HU, et al. (2015) AntiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* **43**: W237–243.
- Weisenfeld NI, Yin S, Sharpe T, Lau B, Hegarty R, et al. (2014) Comprehensive variation discovery in single human genomes. *Nature Genetics* **46**: 1350–1355.
- Wiemann P, Sieber CMK, Von Bargen KW, Studt L, Niehaus E–M, et al. (2013) Deciphering the cryptic genome: genome–wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *PLOS Pathogens* **9**: e1003475.
- Wilken PM, Steenkamp ET, De Beer ZW, Wingfield MJ, Wingfield BD (2013) IMA Genome-F1: Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. *IMA Fungus* **4**: 357–358.
- Wilken PM, Steenkamp ET, Van der Nest MA, Wingfield MJ, De Beer ZW, et al. (2018) Unexpected placement of the *MAT1–1–2* gene in the *MAT1–2* idiomorph of *Thielaviopsis*. *Fungal Genetics and Biology* **113**: 32–41.
- Wingfield BD, Ades PK, Al-Naemi FA, Beirn LA, Bihon W, et al. (2015a) IMA Genome-F4. 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, Ambler JM, Coetzee M, De Beer ZW, Duong TA, et al. (2016a) IMA Genome-F 6: Draft genome sequences of *Armillaria fuscipes*, *Ceratocystiopsis minuta*, *Ceratocystis adiposa*, *Endoconidiophora laricicola*, *E. polonica* and *Penicillium freii* DAOMC 242723. *IMA Fungus* **7**: 217–227.
- Wingfield BD, Barnes I, De Beer ZW, De Vos L, Duong TA, et al. (2015b) IMA Genome-F5: 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, Duong TA, et al. (2017) Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntia decipiens* and *Ophiostoma ips*. *IMA Fungus* **8**: 385–396.
- Wingfield BD, Duong TA, Hammerbacher A, van der Nest MA, Wilson A, et al. (2016b) IMA Genome-F7, Draft genome sequences for *Ceratocystis fagacearum*, *C. harringtonii*, *Grosmannia penicillata*, and *Huntia bhutanensis*. *IMA Fungus* **7**: 317–323.
- Wingfield BD, Steenkamp ET, Santana QC, Coetzee MPA, Bam S, et al. (2012) First fungal genome sequence from Africa: a preliminary analysis. *South African Journal of Science* **108**: 1–9.
- Wingfield MJ, Hammerbacher A, Ganley RJ, Steenkamp ET, Gordon TR, et al. (2008) Pitch canker caused by *Fusarium circinatum* – a growing threat to pine plantations and forests worldwide. *Australasian Plant Pathology* **37**: 319–334.
- Wu W, Davis RW, Tran-Gyamfi MB, Kuo A, LaButti K, et al. (2017) Characterization of four endophytic fungi as potential consolidated bioprocessing hosts for conversion of lignocellulose into advanced biofuels. *Applied Microbiology and Biotechnology* **101**: 2603–2618.
- Wu W, Sutton BC, Gange AC (1996) *Coleophoma fusiformis* sp. nov. from leaves of *Rhododendron*, with notes on the genus *Coleophoma*. *Mycological Research* **100**: 943–947.
- Xu JR, Yan K, Dickman MB, Leslie JF (1995) Electrophoretic karyotypes distinguish the biological species of *Gibberella fujikuroi* (*Fusarium* section *Liseola*). *Molecular Plant-Microbe Interactions* **9**: 74–84.
- Yin Y, Mao X, Yang JC, Chen X, Mao F, et al. (2012) dbCAN: a web resource for automated carbohydrate–active enzyme annotation. *Nucleic Acids Research* **40**: W445–W451.
- Yue Q, Chen L, Zhang X, Li K, Sun J, Liu X, An Z, Bills GF (2015) Evolution of chemical diversity in echinocandin lipopeptide antifungal metabolites. *Eukaryotic Cell* **14**: 698–718.
- Yue Q, Li Y, Chen L, Zhang X, Liu X, An Z, Bills GF (2018) Genomics–driven discovery of a novel self–resistance mechanism in



the echinocandin-producing fungus *Pezicula radicola*. *Environmental Microbiology*: in press: DOI.org/10.1111/1462-2920.14089.

Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* **18**: 821–829.