doi:10.5598/imafungus.2016.07.01.11

IMA FUNGUS · 7(1): 217–227 (2016)

ARTICLE

IMA Genome-F 6

Draft genome sequences of *Armillaria fuscipes, Ceratocystiopsis minuta, Ceratocystis adiposa*, *Endoconidiophora laricicola, E. polonica* and *Penicillium freii* DAOMC 242723

Brenda D. Wingfield[1], Jon M. Ambler[1], Martin P.A. Coetzee[1], Z. Wilhelm de Beer[2], Tuan A. Duong[1], Fourie Joubert[3], Almuth Hammerbacher[2], Alistair R. McTaggart[2], Kershney Naidoo[1], Hai D.T. Nguyen[4,5], Ekaterina Ponomareva[4], Quentin S. Santana[1], Keith A. Seifert[4], Emma T. Steenkamp[2], Conrad Trollip[1], Magriet A. van der Nest[1], Cobus M. Visagie[4,5], P. Markus Wilken[1], Michael J. Wingfield[1], and Neriman Yilmaz[4,6]

[1]Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, 0028 South Africa; corresponding author: brenda.wingfield@fabi.up.ac.za

[2]Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, 0028 South Africa

[3]Centre for Bioinformatics and Computational Biology, Department of Biochemistry and Genomics Research Institute, University of Pretoria, Private Bag X20, Pretoria 0028, South Africa

[4]Biodiversity (Mycology), Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, 960 Carling Ave., Ottawa, Ontario, K1A 0C6, Canada

[5]Department of Biology, University of Ottawa, 30 Marie-Curie, Ottawa, Ontario, K1N6N5, Canada

[6]Department of Chemistry, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6, Canada

Abstract: The genomes of *Armillaria fuscipes, Ceratocystiopsis minuta, Ceratocystis adiposa*, *Endoconidiophora laricicola, E. polonica,* and *Penicillium freii* DAOMC 242723 are presented in this genome announcement. These six genomes are from plant pathogens and otherwise economically important fungal species. The genome sizes range from 21 Mb in the case of *Ceratocystiopsis minuta* to 58 Mb for the basidiomycete *Armillaria fuscipes*. These genomes include the first reports of genomes for the genus *Endoconidiophora.* The availability of these genome data will provide opportunities to resolve longstanding questions regarding the taxonomy of species in these genera. In addition these genome sequences through comparative studies with closely related organisms will increase our understanding of how these pathogens cause disease.

Key words:
Armillaria root rot
grain spoilage
insect vectored fungi
sap stain fungus
sugarcane root rot

Article info: Submitted: 22 February 2016; Accepted: 15 June 2016; Published: 21 June 2016.

IMA Genome-F 6A

Draft genome sequence of the root rot fungus *Armillaria fuscipes*

Armillaria (*Agaricales; Physalacriaceae*) includes species of the most devastating fungal pathogens of trees and woody plants across the globe (Baumgartner *et al.* 2011). Armillaria root rot, the disease caused by these species, manifests as shoot growth reduction, change in foliage characteristics, crown dieback and ultimately the death of natural forest and planted trees, as well as horticultural crops (Morrison *et al.* 1991). Much attention has been afforded to the study of the taxonomy and biology of disease-causing *Armillaria* species. In an effort to understand the mechanisms of pathogenicity and infection of species that occur in the Northern hemisphere, the genome of *A. mellea* (Collins *et al.* 2013) and transcriptome of *A. solidipes* (Ross-Davis *et al.* 2013) have been sequenced. In addition, draft genomes of *A. gallica* and *A. solidipes* are available in the JGI MycoCosm (Grigoriev *et al.* 2014) fungal genomes database. Despite the importance of *Armillaria* root rot in the Southern Hemisphere (Gregory *et al.* 1991), the genome sequence of an *Armillaria* species native to that hemisphere has not yet been determined.

Armillaria root rot disease in the Southern hemisphere is caused by several species, including *A. fuscipes* (Gregory *et al.* 1991, Coetzee *et al.* 2000). The species is restricted to the African continent where it affects the health of agronomic and timber plantations. In South Africa, Armillaria root rot disease was first reported during the early 1900s (Bottomley 1937), but *A. fuscipes* was only identified as the causal agent of root

rot disease on *Pinus* species in 2000 (Coetzee *et al.* 2000). Subsequent studies showed that this pathogen also affects the health of other tree and woody plant species elsewhere in Africa (Mohammed & Guillaumin 1993, Mwenje *et al.* 2003, Gezahgne *et al.* 2004). Although research has been done to characterise African *Armillaria* species at the molecular level in terms of enzymes (e.g. Agustian *et al.* 1994, Mwenje & Ride 1996, 1997, 1999) - and genomic regions used for taxonomic purposes (e.g. Chillali *et al.* 1997, Pérez-Sierra *et al.* 2004, Coetzee *et al.* 2005), nothing is yet known regarding the genome of *A. fuscipes*. The aim of this study was, therefore, to assemble a draft genome of this fungus with the objective of using the genome in future comparative genome studies.

SEQUENCED STRAIN

South Africa: *Pinus elliottii,* 2001, *M.J. Wingfield* (culture CMW 2740 = CBS 118115; PREM 61366 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The Whole Genome Shotgun project of the *Armillaria fuscipes* genome has been deposited at DDBJ/EMBL/GenBank under the accession number LWUH00000000. The version described in this paper is version LWUH01000000.

METHODS

Isolate CMW 2740 was grown in liquid MYA (2 % Malt extract, 0.2% Yeast extract) at 24 °C in the dark for 6 wk. DNA was extracted from the harvested mycelium using a DNeasy Plant Mini Kit (Qiagen, Aarhus). Whole genome sequencing was performed on an Illumina HiSeq platform with a mate pair insert size of 250 bp, as well as PacBio sequencing with 20 single-molecule real time sequencing (SMRT) cells at the UC Davis Genome Centre (University of California, CA).

Quality scores, nucleotide composition, GC composition, and sequence duplication levels of the Illumina reads, were assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which resulted in the removal of the first 10 bp and last 15 bp of all reads. High quality short reads were employed to conduct error correction on the PacBio reads using the Celera Assembler (CA) pipeline (Miller *et al.* 2008). For efficient error correction of the PacBio reads, coverage of 50x high quality short reads was used. For the *A. fuscipes* genome, 32 million high-quality Illumina reads of length 75 bp were used in the error correction. Assembly of the Illumina sequences was conducted using Velvet optimizer (Victorian Bioinformatics Consortium 2012) and CLC Genomics Workbench v. 5.5.1 (Qiagen). The error corrected PacBio data was assembled with VelvetOptimizer, CLC Genomics Workbench v. 5.5.1 and MIRA (http://www.chevreux.org/projects_mira.html). The quality of the assemblies was estimated based on the N50 value, number of contigs, and by applying the core eukaryotic gene mapping approach (CEGMA) (Parra *et al.* 2007) as well as assessing single-copy orthologs with BUSCO (Simão *et al.* 2015). Based on these metrics, the best assemblies from the corrected PacBio and Illumina read datasets were chosen to create an improved hybrid assembly using Graph Accordance Assembly program (GAA) (Yao *et al.* 2012). Scaffolds were generated with SSpace (Boetzer *et al.* 2011). Gene predication utilized RepeatMasker v. 0.10.1 (described in Tarailo-Graovac & Chen 2009) to remove repeat and low complexity DNA regions followed by Augustus (Keller *et al.* 2011).

RESULTS AND DISCUSSION

The draft genome of *Armillaria fuscipes* (CMW 2740) was estimated at 53 Mb and the combined PacBio – Illumina assembly yielded 24 436 contigs with a N50 of 5 415 nucleotide bases. The GC content was 46.8 %. The CEGMA score indicated a completeness of 91.73 %, while 82 % (1190 of 1438) full-length single-copy orthologs were identified with BUSCO using gene models from *Laccaria bicolor*. In total, 199 (13 %) of these orthologs were fragmented and 49 (3.4 %) were missing. *In silico* gene prediction with Augustus yielded 14 515 genes with an average length of 1 350 bp and an average of 5 introns per gene.

Comparison of the *A. fuscipes* draft genome assembly showed that the genome size and the number of predicted genes are mostly congruent with other species of *Armillaria* (Table 1). The draft genome of *A. gallica* (JGI, MycoCosm, Grigoriev *et al.* 2014) differed greatly from *A. fuscipes*, *A. mellea* (Collins *et al.* 2013) and *A. solidipes* (= *A. ostoyae*) (JGI, MycoCosm, Grigoriev *et al.* 2014) in terms of genome size and number of protein-coding genes (Table 1). The estimated genome size of *A. fuscipes* (53Mb) was somewhat smaller than the 58 Mb determined for *A. mellea* and *A. solidipes*. The number of protein-coding genes correlated with that of *A. mellea*, but was less than that determined for *A. solidipes*. Comparison of average gene length, exon length, intron length, and number of exons per gene, revealed similar results for all four *Armillaria* species.

The genome assembly of *A. fuscipes,* as well as thos of the other *Armillaria* species, were compared to the publicly available assembled genomes of other members of the *Physalacriaceae*, namely *Cylindrobasidium torrendii* (Floudas *et al.* 2015) and *Flammulina velutipes* (Park *et al.* 2014) (Table 1). *Cylindrobasidium torrendii* and *F. velutipes* were characterised with smaller genomes than the *Armillaria* species, but the number of protein coding genes, average exon length, and average number of exons per gene ranges were comparable (Table 1). The genome of *F. velutipes* deviated from the other fungal genomes included in this study by having an average gene length that is much larger (2 294 in comparison to 1350–1655 bp) and an exceptional average intron length (180 in comparison to the 59–73.62 bp for the other species).

The genome sequence of *A. fuscipes* determined in this study is considered a high quality draft genome by virtue of the number of genes that were identified using the CEGMA pipeline and BUSCO analyses as well as the correlation in gene characteristics with *other Armillaria* species and members of

Table 1. Comparison of *Armillaria fuscipes* genome characteristics with those of other *Armillaria* species and members of the family *Physalacriaceae*.

	A. fuscipes	***A. mellea*** **(Collins *et al.* 2011)**	***A. solidipes*** **(JGI Project Id: 1032462)**	***A. gallica*** **(JGI Project Id: 1032466)**	***C. torrendii*** **(Floudas *et al.* 2015)**	***F. velutipes*** **(Park *et al.* 2014)**
Genome Assembly (Mb)	53.00	58.30	58.01	85.34	31.57	35.6
Number of protein-coding genes	14 515	14 473	20 811	25 704	13 940	12 218
Average gene length (bp)	1 350	1 575	1 580	1 459	1 655	2 294
Average exon length (bp)	211.75	217.52	227	217	274	245.7
Average intron length (bp)	68.94	73.62	64	68	59	180.4
Average number of exons per gene	6	6.02	5.69	5.40	5.17	5.81

the family *Physalacriaceae*. *Armillaria fuscipes* forms a basal taxon in the phylogeny of *Armillaria* species (Coetzee *et al.* 2011) and its genome sequence therefore provides a crucial resource for future comparative genome studies.

Authors: **M.P.A. Coetzee*, J.M. Ambler, F. Joubert, Q.S. Santana, and E.T. Steenkamp**
**Contact*: Martin.Coetzee@fabi.up.ac.za

IMA Genome-F 6B

Draft genome sequence of *Ceratocystiopsis minuta*

Ceratocystiopsis minuta was first described as *Ophiostoma minutum* from *Picea abies* in Bialozieza, Poland (Siemaszko 1939). It is the type species of *Ceratocystiopsis*, one of the six currently recognized genera of *Ophiostomatales* (de Beer & Wingfield 2013). *Ceratocystiopsis minuta* has a wide distribution, with isolates reported from various countries on five continents (Siemaszko 1939, Davidson 1942, Mathiesen 1951, Mathiesen-Käärik 1960, Upadhyay 1981, Yamaoka *et al.* 1998, Zhou et *al.* 2001, 2004a, 2004b, Kim *et al.* 2005, Plattner *et al.* 2009). After carrying out a phylogenetic study on a collection of *C. minuta* isolates from various sources, Plattner *et al.* (2009) suggested that isolates identified as *C. minuta* might represent a species complex of several phylogenetic species. The species boundaries and global distribution of *C. minuta* could thus be different from that as currently understood. *Ceratocystiopsis minuta* occurs in association with various bark beetle species that infest coniferous hosts (Siemaszko 1939, Mathiesen-Käärik 1960, Yamaoka *et al.* 1998, Zhou *et al.* 2001, 2004, Xudong *et al.* 2004, Kim *et al.* 2005). An inoculation study with this fungus has shown that it is not a pathogen (Yamaoka *et al.* 1998), and its relevance is likely to be simply as an agent of sap stain.

A number of genome sequences have been generated for species of *Ophiostomatales*. These include genomes from species of *Leptographium s. lat.* (DiGuistini *et al.* 2011, van der Nest *et al.* 2014b, Wingfield *et al.* 2015a), *Ophiostoma* (Forgetta *et al.* 2013, Haridas *et al.* 2013, Khoshraftar *et al.* 2013), *Sporothrix* (Cuomo *et al.* 2014, Teixeira *et al.* 2014, d'Alessandro *et al.* 2016), and *Graphilbum* (Wingfield *et al.* 2015b). Of the current six recognized genera in *Ophiostomatales*, there was no genome sequence available for the genera *Raffaelea* or *Ceratocystiopsis*. The aim of this study was to generate the draft genome sequences of *C. minuta*, the type species of the genus *Ceratocystiopsis* and thus to provide a basis for comparison with other genera in the *Ophiostomatales*.

SEQUENCED STRAIN

Austria: Ehrwald: isol. *Ips cembrae*, Aug. 1997, *T. Kirisits* (culture CMW4352 = CBS138717; PREM 61365 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The genomic sequence of *Ceratocystiopsis minuta* (CMW4352, CBS138717) has been deposited at DDBJ/EMBL/GenBank under the accession LZPB00000000. The version described in this paper is version LZPB01000000.

METHODS

A single hyphal-tip culture was grown in 2 % malt extract and 0.5 % yeast extract broth at 25 °C, shaking at 150 rpm for several days. Mycelium was harvested and subjected to lyophilization. Genomic DNA was extracted from lyophilized mycelium using the method described by Aljanabi & Martinez (1997) with additional modification steps (Duong *et al.* 2013). The methods of genome sequencing, assembly and annotation were similar to those used for *Leptographium lundbergii* (Wingfield *et al.* 2015a). Two pair-end libraries (350 bp and 530 bp average insert size) were prepared and sequenced using the Illumina HiSeq 2000 platform. Obtained

reads were subjected to quality filtering and trimming and assembled using CLC Genomics Workbench v. 8.0.1 (CLCBio, Aarhus). The number of protein coding genes were predicted using the MAKER genome annotation pipeline (Cantarel *et al.* 2008) following similar steps as for *L. lundbergii* (Wingfield *et al.* 2015a). The quality and completeness of the assembly were estimated using BUSCO (Simão *et al.* 2015).

RESULTS AND DISCUSSION

In total, over 23.1 million reads were obtained after quality filtering and trimming, with average read length of 85 bp. The assembled genome of *C. minuta* had an estimated size of 21.3 Mb, which was distributed in 904 scaffolds that were over 500 bp in size. The assembly had a N50 of 63.8 kb, with the longest scaffold was just over 398 kb in size. The mean GC content of the genome was 61.67 %, which is slightly higher than what has been reported for other species in the *Ophiostomatales* (DiGuistini *et al.* 2011, Haridas *et al.* 2013, Wingfield *et al.* 2015a, b). The assessment of the completeness of the assembly by using BUSCO with a fungal dataset resulted in a completeness report of C:96%[D:7.6%],F:2.5%,M:1.2%,n:1348 (C:complete [D:duplicated], F:fragmented, M:missed, n:number of genes). MAKER predicted a total number of 6789 protein coding gene models, of which 5060 were multi-exonic genes. The mean exon length was 646 bp and the mean intron length was 114 bp. From the currently available data, *C. minuta* has the smallest genome and highest gene density for any member of the *Ophiostomatales*. The genome sequence of *C. minuta* presented in this study is the first genome sequence generated for *Ceratocystiopsis* and will serve as a valuable resource to study it's biology and phylogenetic relationship with other members in the *Ophiostomtales*.

Authors: **T.A. Duong*, M.J. Wingfield, Z. W. de Beer, and B.D. Wingfield**
**Contact*: tuan.duong@fabi.up.ac.za

IMA Genome-F 6C

Draft genome sequence of *Ceratocystis adiposa*, causal agent of sugarcane root rot

Genome sequences are available for numerous species of *Ceratocystidaceae* (de Beer *et al.* 2014), providing a valuable resource for genomic studies on this important group of fungi. The first of these genomes to be made available was for *Ceratocystis fimbriata s. str.*, the type species of *Ceratocystis* and the causal agent of rot of sweet potato and other root crops (Wilken *et al.* 2013). This was followed by the genomes of two other pathogenic *Ceratocystis* species (*C. albifundus* and *C. manginecans*) as well as species of the related genus *Huntiella* (*H. moniliformis* and *H. omanensis*) (van der Nest *et al.* 2014a, b). Subsequently, draft genome sequences for an additional six species of *Ceratocystidaceae* have been released (van der Nest *et al.* 2015, Wingfield *et al.* 2015a, b). An additional three, those for *C. adiposa*, *Endoconidiophora laricicola* and *E. polonica* are provided in this announcement. This brings the total number of *Ceratocystidaceae* genomes in the public domain to 14, including representatives of the genera *Ceratocystis*, *Endoconidiophora*, *Davidsoniella*, *Thielaviopsis* and *Huntiella* (van der Nest *et al.* 2015; Wingfield *et al.* 2015a, b). Although this resource has been available for only a relatively short time, it has contributed significantly to a diverse array of studies including work on the mating system of these fungi (Wilken *et al.* 2014, Wilson *et al.* 2015) as well as studies on ecological adaptation (van der Nest *et al.* 2015).

The aim of this study was to produce a draft nuclear genome sequence for *C. adiposa*. Best known as the causal agent of root rot on sugarcane (Butler 1906), this species has also been reported from both *Prunus* (Paulin-Mahady *et al.* 2002) and *Pinus* (Talbot 1956) species. Since its first description as *Ceratostomella adiposum* (Sartoris 1927), this species has also been treated as *Ophiostoma adiposum* and *E. adiposa*, and in 1981 *C. adiposa* was also suggested to be synonymous with *C. major* (de Beer *et al.* 2014). This long and confused taxonomic history has yet to resolve the exact placement of *C. adiposa* within *Ceratocystidaceae* (de Beer *et al.* 2014). Therefore, apart from adding to the *Ceratocystidaceae* genome resource, the availability of a genome sequence for this species could also provide useful tools to address its taxonomy.

SEQUENCED STRAIN

Japan: isol. ex *Saccharum officinarum*, T. Miyake (culture CMW 2574 = CBS 136.34; PREM 61367 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The Whole Genome Shotgun project of the *Ceratocystis adiposa* genome has been deposited at DDBJ/EMBL/GenBank under the accession no. LXGU00000000. The version described in this paper is version LXGU01000000.

METHODS

Ceratocystis adiposa isolate CMW 2574 was obtained from the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute of the University of Pretoria and grown on 2 % malt extract agar (20 g ME and 20 g Agar, Biolab, South Africa) supplemented with 100 µg/L thiamine for 2 wk. The plates were subjected to a previously described DNA isolation procedure (Roux *et al.* 2004), before the purified nucleic acid was sequenced at the Genome Centre, University of California at Davis (CA) using a Genome Analyzer IIx platform (Illumina). For sequencing, paired-end libraries with approximately 350 and 600 base insert sizes were produced, and used to produce reads with an average length of 100 bases. Poor-quality reads and/or terminal

nucleotides were discarded using the software package CLC Genomics Workbench v. 8.5.1 (Qiagen, Aarhus). A draft genome sequence was produced using the *de novo* assembly option in CLC Genomics Workbench under default options. The contigs produced were subjected to scaffolding using SSPACE and to fill the resultant gaps using GapFiller (Boetzer *et al.* 2011, Boetzer & Pirovano 2012). Putative open reading frames (ORFs) were predicted using the online version of the *de novo* prediction software AUGUSTUS based on the gene models for *Fusarium graminearum* (Stanke *et al.* 2006), while genome completeness was assessed with the Benchmarking Universal Single-Copy Orthologs (BUSCO) software on all contigs greater than 500 bases in size (Simão *et al.* 2015).

RESULTS AND DISCUSSION

The draft nuclear genome of *Ceratocystis adiposa* has an estimated size of 28 447 711 bp as assessed through the summation of all contigs. The genome assembly has a N50 value of 114 645, an average coverage of 107x with a mean GC content of 47 %. Of the 951 contigs produced during the assembly process, 644 were 500 bp or larger and were retained in the final contig list. This draft assembly is predicted to contain 6 830 ORFs based on an AUGUSTUS analysis and has a density of 240 ORFs/Mb. The assembly also appeared to have a high degree of completeness with a BUSCO score of 97 %. The assembly contained 1 405 Complete Single-Copy BUSCOs, 93 Complete Duplicated BUSCOs, 26 Fragmented BUSCOs and 7 missing BUSCO orthologs out of the 1 438 BUSCO groups searched.

The overall genome statistics (i.e. genome size, ORF density and GC content) for the *C. adiposa* assembly is typical of what has been reported previously for *Ceratocystidaceae* (Wingfield *et al.* 2015b). That our assembly consists of a relatively small number of contigs, combined with the high completeness score, will be advantageous to any downstream analysis of the *C. adiposa* genome, especially in studies aimed at resolving its unclear evolutionary relationship with other *Ceratocystidaceae* (de Beer *et al.* 2014). Apart from contributing to the ever-growing *Ceratocystidaceae* genome resource, the draft assembly presented here should be a useful resource for studying the biology *C. adiposa*.

Authors: **P.M. Wilken*, E.T. Steenkamp, K. Naidoo, C. Trollip, M.J. Wingfield, Z.W. de Beer, and B.D. Wingfield**
**Contact*: Markus.Wilken@fabi.up.ac.za

IMA Genome-F 6D

Draft genome sequences of the *Ips* bark beetle symbionts *Endoconidiophora laricicola* and *E. polonica*

The family *Ceratocystidaceae* (*Microscales*) includes numerous insect-vectored, wood-staining and pathogenic species of significant economic importance (Seifert *et al.* 1993, Marín & Wingfield, 2006, de Beer *et al.* 2014). This diverse group of fungi occurs on both angiosperms and gymnosperms and was initially treated together in the genus *Ceratocystis* due to superficial similarities in their morphological features and mutualistic insect associations (Wingfield *et al* 1993, de Beer *et al.* 2014). The taxonomy of these fungi has recently been revised using a phylogenetic species concept, in addition to morphological, ecological and biological characters (de Beer *et al.* 2014). Accordingly, the *Ceratocystidaceae* now incorporates several genera, including *Ceratocystis s.str*.

One of the newly recognised genera, *Endoconidiophora,* corresponds to what was previously known as the "*C. coerulescens* complex" (de Beer *et al.* 2014). This genus includes both *E. polonica* and *E. laricicola*, two of the most aggressively pathogenic fungi associated with the bark beetles (*Coleoptera*: *Scoltyinae*) of *Ips typographus* and *Ips cembrae* that infest spruce (*Picea abies*) and larch (*Larix decidua*), respectively (Christiansen & Horntvedt 1983, Redfern *et al*. 1987). As opposed to many other species of *Ceratocystidaceae* that have non-specific associations with fungivorous or sapfeeding insects, *Endoconidiophora* species have specific mutualistic relationships with their bark beetle vectors (Kile 1993, Wingfield *et al.* 1997). In these associations, the fungi are thought to facilitate growth and reproduction of the insect, while the insect disperses the fungi (Paine *et al.* 1997). However, the role of these fungi in tree death has been questioned (Six & Wingfield 2011).

The aim of this study was to sequence the genomes of *E. laricicola* and *E. polonica* to facilitate comparative studies with the genomes of other member species of *Ceratocystidaceae* already in the public domain (Wilken *et al.* 2013, van der Nest *et al.* 2014a, b, Wingfield *et al.* 2015a, b, Belbahri 2015). This study is also part of a larger objective to expand our knowledge of the biology and evolution of the interactions among fungi, bark beetles and their plant hosts.

SEQUENCED STRAIN

Endoconidiophora laricicola: **United Kingdom**: Scotland: *Larix* sp., Nov. 1997, *D. Refern* (culture CMW 20928 = CBS 100207; PREM 61368 – dried culture).
Endoconidiophora polonica: **Norway**: *Picea abies*, Jan. 1990, *H. Solheim* (culture CMW 20930 = CBS 100205; PREM 61369 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The draft genome sequence of *Endoconidiophora laricicola* (CMW 20928) and *E. polonica* (CMW 20930) have been deposited in DDBJ/ENA/GenBank under the accession numbers LXGT00000000 and LXKZ0000000, respectively. The versions described in this paper are the first version of each genome, versions LXGT0100000 and LXKZ0100000.

METHODS

Genomic DNA of *E. laricicola* isolate CMW 20928 and *E. polonica* CMW 20930 maintained in the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa and the CBS-KNAW Fungal Biodiversity Centre (CBS), Utrecht, The Netherlands were sequenced with the Illumina HiSeq2000 platform at the UC Davis Genome Centre, University of California, Davis (CA). Two runs with 350-bp and 530-bp paired-end reads were performed following standard Illumina protocols to generate sequences with read lengths of approximately 100 bases. Reads were quality controlled and trimmed using CLC Genomics Workbench v. 7.5.1 (CLCBio, Aarhus). A draft genome assembly was performed using the *de novo* assembly option in the CLC Genomics Workbench under default options. SSPACE v. 2.0 (Boetzer *et al.* 2011) was used to assemble contigs into scaffolds and gaps were filled using GapFiller v. 2.2.1 (Boetzer & Pirovano 2012). The Benchmarking Universal Single-Copy Orthologs tool, BUSCO (Software v. 1.1b1 of May 2015) (Simão *et al.* 2015) was used to assess the completeness of the assembled genome. Lastly, the assembly was submitted to AUGUSTUS (Stanke *et al.* 2004) in order to predict putative open reading frames (ORFs) using the gene models of *Fusarium graminearum*.

RESULTS AND DISCUSSION

The draft genomes of *Endoconidiophora laricicola* and *E. polonica* had estimated sizes of 32 785 225 and 32 461 618 bases with an average coverage of 93x and 82x, respectively. The final *E. laricicola* assembly consisted of a total of 879 contigs larger than 500 bases, while the final *E. polonica* assembly consisted of a total of 914 contigs that were 500 bases or more in size. The GC content of both these two genomes was 45 % and the N50 for *E. laricicola* and *E. polonica* were 77 789 and 86 326 bases, respectively. The BUSCO analysis defined the genomes as 98 % complete, indicating that most of the core eukaryotic genes were present. From the *E. laricicola* analysis, we observed 1 415 complete single-copy BUSCOs, 104 complete duplicated BUSCOs, 19 fragmented BUSCOs and only 4 missing BUSCOs out of the possible 1 438 groups searched from the dataset for the fungal lineage. From the *E. polonica* analysis, we observed 1 415 complete single-copy BUSCOs, 97 complete duplicated BUSCOs, 10 fragmented BUSCOs and only 13 missing BUSCOs out of the possible 1 438 groups searched from the dataset for the fungal lineage.

Based on size, gene number and GC content, the *E. laricicola* (33.3 Mb, 6 897 ORFs, 45 % GC) and *E. polonica* (33.5 Mb, 6 792 ORFs, 45 % GC) draft genomes resembled those of other species in the Ceratocystidaceae. These include *Davidsoniella virescens* (33.7 Mb, 6953 ORFs, 44.5 % GC), *C. fimbriata* (29.4 Mb, 7266 ORFs, 48 % GC), *C. eucalypticola* (31.3 Mb, 7353 ORFs, 47.9 % GC), *C. manginecans* (31.7 Mb, 7494 ORFs, 47.9 % GC), *C. platani* (29.2 Mb, 5963 ORF, 48.2 % GC) and *C. albifundus* (27.2 Mb, 6967 ORFs, 48.6 % GC) that infect eudicots. This similarity also extended to *Thielaviopsis musarum* (28.5 Mb, 6963 ORF, 49.2 % GC) and *T. punctuala* (28.1Mb, 5480 ORFs, 48.3 % GC) that infect monocots, as well as the non-pathogens *H. moniliformis* (25 Mb, 6832 ORF, 48 % GC) and *H. omanensis* (31.5 Mb, 8359 ORFs eudicots, 47.6 % GC) (Wilken *et al*. 2013, van der Nest *et al.* 2014a, b, c, Wingfield *et al.* 2015a, b, Belbahri 2015).

Availability of the *E. laricicola* and *E. polonica* draft genomes provides an additional genome resource for members of the *Ceratocystidaceae*. These genomes can be incorporated in future comparative genomics studies to investigate the biology and evolution of species in the *Ceratocystidaceae*, as well as the evolution of the family within the broader *Ascomycota*. This is especially in regard to the interactions of these fungi with their insect vectors and plant hosts.

Authors: **M.A. van der Nest*, T.A. Duong, P.M. Wilken, K. Naidoo, A. Hammerbacher, E.T. Steenkamp, A.R. McTaggart, M.J. Wingfield, and B.D. Wingfield**
**Contact*: magriet.vandernest@fabi.up.ac.za

IMA Genome-F 6E

Draft genome sequence of *Penicillium freii* DAOMC 242723

Penicillium freii is classified in *Penicillium* sect. *Fasiculata* (Houbraken & Samson 2011) together with such well-known species as *P. camemberti* and *P. caseifulvum* (used for making camembert cheese), *P. crustosum* (cause of apple rot; penitrem A and roquefortine C producer), *P. expansum* (cause of apple rot), *P. roquefortii* (used for making blue cheeses; roquefortine C producer), *P. solitum* (cause of fruit and nut spoilage; compactin producer), and *P. verrucosum* (cause of grain spoilage; ochratoxin A producer) (Frisvad & Samson 2004), among many others. *Penicillium freii* is a grain spoilage organism commonly found on barley and wheat in colder climates such as Scandinavia, the UK, and Canada, and produces xanthomegnins (hepato- and nephrotoxins) and penicillic acid (antibiotic, antiviral, antitumor, cytotoxic and a possible mycotoxin) (Frisvad *et al.* 2004). It is part of the *P. aurantiogriseum* species complex, a group of mostly grain species that produce a variety of mycotoxins and have slow growing, blue-green agar colonies, smooth to finely roughened conidiophore stipes and smooth walled, more or less globose conidia (Frisvad & Samson 2004). On Czapek yeast extract agar (CYA), it produces crustose colonies, blue-green conidia and abundant exudate droplets (Fig. 1). In this study, we sequenced and annotated a genome draft of *P. freii* DAOMC 242723.

SEQUENCED STRAIN

Canada: *Ontario*: Ridgetown, isolated from wheat flour, 15 July 2011, collected and isolated by *A. Schaafsma*, identified by *C.M. Visagie* & *N. Yilmaz* (culture DAOMC 242723 – dried specimen deposited at DAOM).

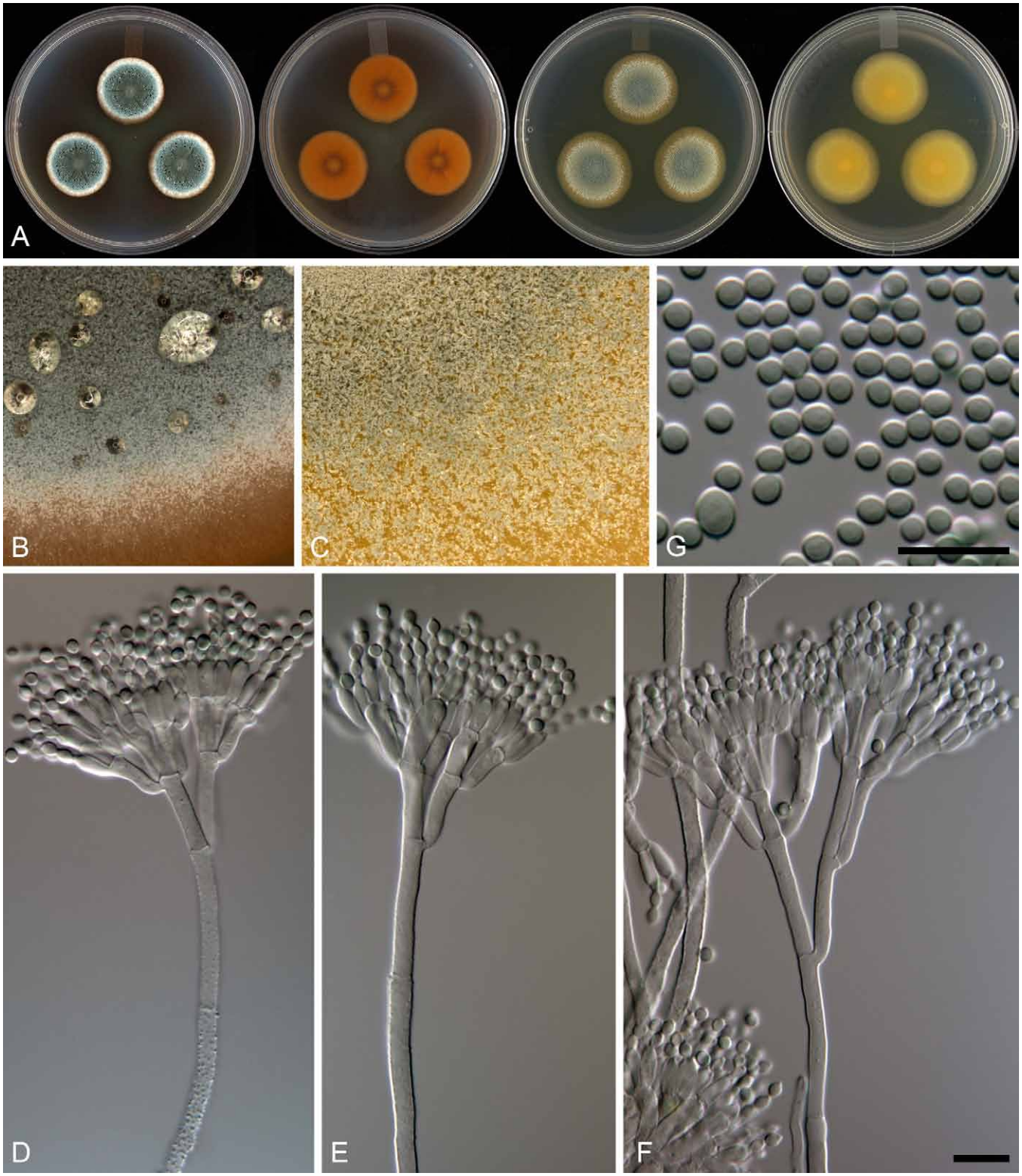

Fig. 1. *Penicillium freii* (DAOMC 242723). A. Colonies (from left to right), CYA, CYA reverse, MEA, MEA reverse. B. Colony texture on CYA. C. Colony texture on MEA. D–F. Conidiophores. G. Conidia. Bars: F = 10 µm (applies to D & E), G = 10 µm.

NUCLEOTIDE SEQUENCE AND RAW READS ACCESSION NUMBERS

This Whole Genome Shotgun project is deposited at DDBJ/EMBL/GenBank under accession LLXE00000000. The version described in this paper is version LLXE01000000. Raw reads were deposited in NCBI SRA (http://www.ncbi.nlm.nih.gov/sra) accession number SRR2146291.

METHODS

Penicillium freii DAOMC 242723 was grown on Blakeslee's malt extract agar for 7 d at 25 °C (Visagie *et al.* 2014) after which a spore suspension was prepared by flooding the colony with 5 mL sterile dH_2O. One mL spore suspension was used to inoculate 100 mL Blakeslee's malt extract broth in shaking flasks incubated at 300 rpm at 25 °C for 6 d. Fungal material was collected by filtration and DNA extracted with the OmniPrep kit for fungi (G-Biosciences) following the manufacturer's protocol. Whole-genome sequencing (paired-end with 101 base pairs, bp) was performed on an Illumina HiSeq 2500 with TrueSeq V3 chemistry at the National Research Council Canada in Saskatoon, Saskatchewan, Canada.

FastQC v. 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to check the quality of genomic reads. Using fastx_trimmer (part of the FASTX-Toolkit v. 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/)), 10 bases from the 5' end were trimmed to yield higher quality reads of 91 bp. Adaptor sequences were removed with Trimmomatic v. 0.33 (Bolger *et al.* 2014). Prior to genome assembly, the optimal k parameter was calculated with KmerGenie v. 1.6950 (Chikhi & Medvedev 2014). Error correction was performed on the trimmed reads with BayesHammer (Nikolenko *et al.* 2013). *De novo* genome assembly was performed with SPAdes v. 3.5.0 (Bankevich *et al.* 2012) with the option to correct mismatches and short indels enabled. Scaffolds shorter than 1000 bp were discarded. Assembly statistics were generated with QUAST v. 2.3 (Gurevich *et al.* 2013). The assembly was checked by alignment of the corrected reads onto the scaffolds using Bowtie2 v. 2.0.0 (Langmead & Salzberg 2012). Alignments produced by Bowtie2 in SAM format were converted to sorted BAM format by SAMtools v. 0.1.19 (Li *et al.* 2009) and statistics for nucleotide coverage were generated with Qualimap v. 2.1 (Garcia-Alcalde *et al.* 2012). To evaluate the completeness of our assembly, CEGMA v. 2.5 (Parra *et al.* 2007) was run on the scaffolds to detect the percentage of conserved eukaryotic genes (CEG's) and BUSCO v. 1.1b1 (http://busco.ezlab.org/) was run on the scaffolds using the fungal profile (Dec 19, 2014 release) to detect Universal Single-Copy Orthologs. Species identification was confirmed by microscopic observation and BLASTing the internal transcribed spacer (JN942696) and beta-tubulin (AY674290) barcode sequences of *P. freii* (Visagie *et al.* 2014) against the assembled genomic scaffolds.

Genome annotation was carried out using webAugustus (Hoff & Stanke 2013) running Augustus v. 3.0.3 (Stanke *et al.* 2006). Predicted proteins were compared against the UniProt/Swiss-Prot manually curated fungal protein data set using BLASTp v. 2.2.28+. The BLAST hits with e-values less than $1.0E^{-100}$ and similarity hits ≥ 90% were assumed to be orthologs and were given protein names in the annotation set. Annotations were validated using Genome Annotation Generator (http://genomeannotation.github.io/GAG/) and tbl2asn (http://www.ncbi.nlm.nih.gov/genbank/tbl2asn2/).

RESULTS AND DISCUSSION

Illumina sequencing of 32 million reads of *Penicillium freii* DAOMC 242723 represented 3.2 Gbp of data, which were assembled into 1923 scaffolds. The whole assembly was 33.5 Mbp with a GC content of 47.4 %. The following statistics for the assembly were obtained: N50 was 61.8 Kbp; the longest scaffold was 325 Kbp; the median nucleotide coverage across the whole assembly was 78.5x. Assessment of the completeness of the genome using BUSCO groups for fungi resulted in values of C:98%[D:6.9%],F:1.1%,M:0.1% ,n:1438 (C:complete [D:duplicated], F:fragmented, M:missed, n:genes) whereas scores of 95.6 % and 97.6 % were obtained from the complete and partial gene set respectively using CEGMA. After annotation and validation, 11 739 predicted gene models were obtained, with 11 418 complete (97.2 %) and 320 lacking a start codon, stop codon or both (2.7 %). Mean gene length was 1489 bp, mean exon length was 483 bp and mean intron length was 79 bp. This draft genome of *P. freii* represents a useful resource for comparative genomic studies and in particular should be useful for those interested in polyketide and other mycotoxin synthesis in *Penicillium* and related genera.

Authors: **H.D.T. Nguyen*, C.M. Visagie, N. Yilmaz, E. Ponomareva, and K.A. Seifert**
**Contact*: hai.nguyen.1984@gmail.com

ACKNOWLEDGEMENTS

Genome sequencing of *Penicillium freii* DAOMC 242723 was funded by Growing Forward 2 funding from Agriculture & Agri-Food Canada, with additional support from Canadian Safety and Security Programme grant CRTI 09-462RD/CSSP 30vv01.

The genome sequencing of the species *Armillaria fuscipes, Ceratocystiopsis minuta, Ceratocystis adiposa, Endoconidiophora laricicola* and *E. polonica* were financed by the University of Pretoria, the Department of Science and Technology (DST)/National Research Foundation (NRF) Centre of Excellence in Tree Health Biotechnology (CTHB), and the Genomics Research Institute (University of Pretoria). This work is based on the research supported in part by a number of grants from the National Research Foundation of South Africa (including Grant specific unique reference number (UID) 83924 and 87332). The Grant holders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author(s), and that the NRF accepts no liability whatsoever in this regard.

REFERENCES

Agustian A, Mohammed C, Guillaumin J-J, Botton B (1994) Discrimination of some African *Armillaria* species by isozyme electrophoretic analysis. *New Phytologist* **128**: 135–143.

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* **25**: 4692–4693.

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.

Baumgartner K, Coetzee MPA, Hoffmeister D (2011) Secrets of the subterranean pathosystem of *Armillaria*. *Molecular Plant Pathology* **12**: 515–534.

Belbahri L (2015) Genome sequence of *Ceratocystis platani*, a major pathogen of plane trees. URL http://www.ncbi.nlm.nih.gov/nuccore/814603118.

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.

Bottomley AM (1937) Some of the more important diseases affecting timber plantations in the Transvaal. *South African Journal of Science* **33**: 373–376.

Butler EJ (1906) Fungus diseases of sugar-cane in Bengal. *Memoirs of the Department of Agriculture in India by Botanical Series* **1**: 1–53.

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.

Chikhi R, Medvedev P (2014) Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**: 31–37.

Chillali M, Idder-Ighili H, Agustian A, Guillaumin J-J, Mohammed C, *et al.* (1997) Species delimitation in the African *Armillaria* complex by analysis of the ribosomal DNA spacers. *Journal of General and Applied Microbiology* **43**: 23–29.

Christiansen E, Horntvedt R (1983) Combined *Ips/Ceratocystis* attack on Norway spruce, and defensive mechanisms of the trees. *Zeitschrift für angewandte Entomologie* **96**: 110–118.

Coetzee MP, Wingfield BD, Bloomer P, Wingfield MJ (2005) Phylogenetic analyses of DNA sequences reveal species partitions amongst isolates of *Armillaria* from Africa. *Mycological Research* **109**: 1223–1234.

Coetzee MPA, Bloomer P, Wingfield MJ, Wingfield BD (2011) Paleogene radiation of a plant pathogenic mushroom. *PLoS ONE* **6**: e28545.

Coetzee MPA, Wingfield BD, Coutinho TA, Wingfield MJ (2000) Identification of the causal agent of Armillaria root rot of *Pinus* species in South Africa. *Mycologia* **92**: 777–785.

Collins C, Keane TM, Turner DJ, O'Keeffe G, Fitzpatrick DA, *et al*. (2013) Genomic and proteomic dissection of the ubiquitous plant pathogen, *Armillaria mellea*: Toward a new infection model system. *Journal of Proteome Research* **12**: 2552–2570.

Cuomo CA, Rodriguez-Del Valle N, Perez-Sanchez L, *et al*. (2014) Genome sequence of the pathogenic fungus *Sporothrix schenckii* (ATCC 58251). *Genome Announcements* **2**: e00446–00414.

D'Alessandro E, Giosa D, Huang L, Zhang J, Gao W, *et al*. (2016) Draft genome sequence of the dimorphic fungus *Sporothrix pallida*, a nonpathogenic species belonging to *Sporothrix*, a genus containing agents of human and feline sporotrichosis. *Genome Announcements* **4**: e00184-00116.

Davidson RW (1942) Some additional species of *Ceratostomella* in the United States. *Mycologia* **34**: 650–662.

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, Wingfield MJ (2013) Emerging lineages in the *Ophiostomatales*, In: *The Ophiostomatoid Fungi: expanding frontiers* (Seifert KA, de Beer ZW, Wingfield MJ eds): 21–46. [CBS Biodiversity Series no. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

DiGuistini S, Wang Y, Liao NY, Taylor G, Tanguay P, *et al*. (2011) Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen. *Proceedings of the National Academy of Sciences, USA* **108**: 2504–2509.

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.

Floudas D, Held BW, Riley R, Nagy LG, Koehler G, *et al.* (2015) Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. *Fungal Genetics and Biology* **76**: 78–92.

Forgetta V, Leveque G, Dias J, Grove D, Lyons R, *et al*. (2013) Sequencing of the Dutch Elm Disease fungus genome using the Roche/454 GS-FLX Titanium system in a comparison of multiple genomics core facilities. *Journal of Biomolecular Techniques* **24**: 39–49.

Frisvad JC, Samson RA (2004) Polyphasic taxonomy of *Penicillium* subgenus *Penicillium* - a guide to identification of food and air-borne terverticillate *Penicillia* and their mycotoxins. *Studies In Mycology* **49**: 1–174.

Frisvad JC, Smedsgaard J, Larsen TO, Samson RA (2004) Mycotoxins, drugs and other extrolites produced by species in *Penicillium* subgenus *Penicillium*. *Studies In Mycology* **49**: 201–241.

Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, *et al.* (2012) Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**: 2678–2679.

Gezahgne A, Coetzee MPA, Wingfield BD, Wingfield MJ, Roux J (2004) Identification of the *Armillaria* root rot pathogen in Ethiopian plantations. *Forest Pathology* **34**: 133–145.

Gregory SC, Rishbeth J, Shaw CG (1991) Pathogenicity and virulence. In: *Armillaria Root Disease* (Shaw CG, Kile GA, eds): 76–87 [Agriculture Handbook no. 691.] Washington D.: Forest Service, United States Department of Agriculture.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, *et al.* (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: D699–D704.

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072–1075.

Haridas S, Wang Y, Lim L, Alamouti SM, Jackman S, *et al*. (2013) The genome and transcriptome of the pine saprophyte *Ophiostoma piceae*, and a comparison with the bark beetle-associated pine pathogen *Grosmannia clavigera*. *BMC Genomics* **14**: 373.

Hoff KJ, Stanke M (2013) WebAUGUSTUS - a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research* **41**: W123–128.

Houbraken J, Samson RA (2011) Phylogeny of *Penicillium* and the segregation of *Trichocomaceae* into three families. *Studies In Mycology* **70**: 1–51.

Keller O, Kollmar M, Stanke M, Waack S (2011) A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**: 757–763.

Khoshraftar S, Hung S, Khan S, Gong Y, Tyagi V, *et al*. (2013) Sequencing and annotation of the *Ophiostoma ulmi* genome. *BMC Genomics* **14**: 162.

Kile GA (1993) Plant diseases caused by species of *Ceratocystis sensu stricto* and Chalara. In: *Ceratocystis and Ophiostoma: taxonomy, ecology, and pathogenicity* (Wingfield MJ, Seifert KA, Webber JF, eds): 173–183. St Paul, MN: American Phytopathological Society Press.

Kim J-J, Allen EA, Humble LM, Breuil C (2005) Ophiostomatoid and basidiomycetous fungi associated with green, red, and grey lodgepole pines after mountain pine beetle (*Dendroctonus ponderosae*) infestation. *Canadian Journal of Forest Research* **35**: 274–284.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Marín M, Wingfield MJ (2006) A review of *Ceratocystis sensu stricto* with special reference to the species complexes *C. coerulescens* and *C. fimbriata*. *Revista Facultad Nacional de Agronomía, Medellín* **59**: 3045–3075.

Mathiesen A (1951) Einige neue *Ophiostoma*-arten in Schweden. *Svensk Botanisk Tidskrift* **45**: 203–232.

Mathiesen-Käärik A (1960) Studies on the ecology, taxonomy and physiology of Swedish insect-associated blue stain fungi, especially the genus *Ceratocystis*. *Oikos* **11**: 1–25.

Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, *et al*. (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* **24**: 2818–2824.

Mohammed C, Guillaumin J-J (1993) *Armillaria* in tropical Africa. In: *Aspects of Tropical Mycology* (Isaac S, Frankland JC, Watling R, Whalley AJS, eds): 207–217. Cambridgel: Cambridge University Press.

Morrison DJ, Williams RE, Whitney RD (1991) Infection, disease development, diagnosis, and detection. In: *Armillaria Root Disease* (Shaw CG, Kile GA, eds): 62–75 [Agriculture Handbook no. 691.] Washington DC: Forest Service, United States Department of Agriculture.

Mwenje E, Ride JP (1996) Morphological and biochemical characterization of *Armillaria* isolates from Zimbabwe. *Plant Pathology* **45**: 1036–1051.

Mwenje E, Ride JP (1997) The use of pectic enzymes in the characterization of *Armillaria* isolates from Africa. *Plant Pathology* **46**: 341–354.

Mwenje E, Ride JP (1999) Purification and characterization of an endo-polygalacturonase (PG1) from a Zimbabwean species of *Armillaria*. *Physiological and Molecular Plant Pathology* **55**: 131–139.

Mwenje E, Wingfield BD, Coetzee MP, Wingfield MJ (2003) Molecular characterisation of *Armillaria* species from Zimbabwe. *Mycological Research* **107**: 291–296.

Nikolenko SI, Korobeynikov AI, Alekseyev MA (2013) BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* **14** (Suppl. 1): S7.

Paine TD, Raffa KF, Harrington TC (1997) Interactions among scolytid bark beetles, their associated fungi, and live host conifers. *Annual Review of Entomology* **42**: 179–206.

Park Y-J, Baek JH, Lee S, Kim C, Rhee H, *et al.* (2014) Whole genome and global gene expression analyses of the model mushroom *Flammulina velutipes* reveal a high capacity for lignocellulose degradation. *PloS One* **9**: e93560.

Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**: 1061–1067.

Paulin-Mahady AE, Harrington TC, McNew D (2002) Phylogenetic and taxonomic evaluation of *Chalara*, *Chalaropsis*, and *Thielaviopsis* anamorphs associated with *Ceratocystis*. *Mycologia* **94**: 62–72.

Pérez-Sierra A, Guillaumin J-J, Spooner BM, Bridge PD (2004) Characterization of *Armillaria heimii* from Africa. *Plant Pathology* **53**: 220–230.

Plattner A, Kim J-J, Reid J, Hausner G, Lim YW, *et al*. (2009) Resolving taxonomic and phylogenetic incongruence within species *Ceratocystiopsis minuta*. *Mycologia* **101**: 878–887.

Redfern DB, Stoakley JT, Steele H, Minter D (1987) Dieback and death of larch caused by *Ceratocystis laricicola sp. nov.* following attack by *Ips cembrae*. *Plant Pathology* **36**: 467–80.

Ross-Davis AL, Stewart JE, Hanna JW, Kim MS, Knaus BJ, *et al*. (2013) Transcriptome of an *Armillaria* root disease pathogen reveals candidate genes involved in host substrate utilization at the host–pathogen interface. *Forest Pathology* **43**: 468–477.

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.

Sartoris GB (1927) A cytological study of *Ceratostomella adiposum* (Butl.) comb. nov., the black-rot fungus of sugar cane. *Journal of Agricultural Research* **35**: 577–583.

Seifert KA, Wingfield MJ, Kendrick WB (1993) A nomenclature for described species of *Ceratocystis*, *Ophiostoma*, *Ceratocystiopsis*, *Ceratostomella* and *Sphaeronaemella*. In: *Ceratocystis and Ophiostoma: taxonomy, ecology, and pathogenicity* (Winfield MJ, Seifert KA, Webber JF, eds): 269–287. St Paul, MN: American Phytopathological Society Press.

Siemaszko W (1939) Fungi associated with bark beetles in Poland. *Planta Polonica* **7**: 1–54.

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.

Six DL, Wingfield MJ (2011) The role of phytopathogenicity in bark beetle-fungus symbioses: A challenge to the classic paradigm. *Annual Review of Entomology* **56**: 255–72.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, *et al.* (2006) AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research* **34**: W435–439.

Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research* **32**: 309–312.

Stanke M, Tzvetkova A, Morgenstern B (2006) AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biology* **7** (Suppl. 1): S11.11–S11.18.

Talbot P (1956) New and interesting records of South African fungi. Part II. *Bothalia* **6**: 489–500.

Tarailo-Graovac M, Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* **25**: 4.10.1–4.10.14.

Teixeira MM, de Almeida LG, Kubitschek-Barreira P, Alves FL, *et al*. (2014) Comparative genomics of the major fungal agents

of human and animal Sporotrichosis: *Sporothrix schenckii* and *Sporothrix brasiliensis*. *BMC Genomics* **15**: 943.

Upadhyay H (1981) *A monograph of Ceratocystis and Ceratocystiopsis*. Athens, GA: University of Georgia Press.

van der Nest MA, Beirn LA, Crouch JA, Demers JE, de Beer ZW, *et al.* (2014b) IMA Genome–F 3: Draft genomes of *Amanita jacksonii, Ceratocystis albifundus, Fusarium circinatum, Huntiella omanensis, Leptographium procerum, Rutstroemia sydowiana,* and *Sclerotinia echinophila*. *IMA Fungus* **5**: 473–486.

van der Nest MA, Bihon W, De Vos L, Naidoo K, Roodt D, *et al.* (2014a) IMA Genome-F 2: *Ceratocystis manginecans, Ceratocystis moniliformis, Diplodia sapinea*: 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.

Victorian Bioinformatics Consortium (2012) *VelvetOptimiser.* bioinformatics.net.au/software.velvetoptimiser.shtml.

Visagie CM, Houbraken J, Frisvad JC, Hong SB, Klaassen CHW, *et al.* (2014) Identification and nomenclature fo the genus *Penicillium*. *Studies In Mycology* **78**: 343–371.

Wilken PM, Steenkamp ET, Wingfield MJ, de Beer ZW, Wingfield BD (2013) IMA Genome–F 1: *Ceratocystis fimbriata*: draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. *IMA Fungus* **4**: 357–358.

Wilken PM, Steenkamp ET, Wingfield MJ, de Beer ZW, Wingfield BD (2014) DNA loss at the *Ceratocystis fimbriata* mating locus results in self-sterility. *PLoS ONE* **9**: e92180.

Wilson AM, Godlonton T, van der Nest MA, Wilken PM, Wingfield MJ*, et al.* (2015) Unisexual reproduction in *Huntiella moniliformis*. *Fungal Genetics and Biology* **80**: 1–9.

Wingfield BD, Ades PK, Al-Naemi FA, Beirn LA, Bihon W*, et al.* (2015a) IMA Genome-F 4: Draft genome sequences of *Chrysoporthe austroafricana*, *Diplodia scrobiculata*, *Fusarium nygamai*, *Leptographium lundbergii*, *Limonomyces culmigenus*, *Stagonosporopsis tanaceti*, and *Thielaviopsis punctulata*. *IMA Fungus* **6**: 233–248.

Wingfield BD, Barnes I, de Beer ZW, De Vos L, Duong TA*, et al.* (2015b) IMA Genome-F 5: Draft genome sequences of *Ceratocystis eucalypticola*, *Chrysoporthe cubensis*, *C. deuterocubensis*, *Davidsoniella virescens*, *Fusarium temperatum*, *Graphilbum fragrans*, *Penicillium nordicum*, and *Thielaviopsis musarum*. *IMA Fungus* **6**: 493–506.

Wingfield MJ, Harrington TC, Solheim H (1997) Two species in the *Ceratocystis coerulescens* complex from conifers in western North America. *Canadian Journal of Botany* **75**: 827–834.

Wingfield MJ, Seifert KA, Webber JF (eds) (1993) *Ceratocystis* and *Ophiostoma*: *taxonomy, ecology and pathogenicity*. St Paul, MN: American Phytopathological Society Press.

Yamaoka Y, Wingfield MJ, Ohsawa M, Kuroda Y (1998) Ophiostomatoid fungi associated with *Ips cembrae* in Japan and their pathogenicity of Japanese larch. *Mycoscience* **39**: 367–378.

Yao G, Ye L, Gao H, Minx P, Warren WC, Weinstock GM (2012) Graph accordance of next-generation sequence assemblies. *Bioinformatics* **28**: 13–16

Zhou XD, de Beer ZW, Wingfield BD, Wingfield MJ (2001) Ophiostomatoid fungi associated with three pine-infesting bark beetles in South Africa. *Sydowia* **53**: 290–300.

Zhou XD, de Beer ZW, Ahumada R, Wingfield BD, Wingfield MJ (2004a) *Ophiostoma* and *Ceratocystiopsis* spp. associated with two pine-infesting bark beetles in Chile. *Fungal Diversity* **15**: 253–266.

Zhou XD, de Beer ZW, Cibrian D, Wingfield BD, Wingfield MJ (2004b) Characterisation of *Ophiostoma* species associated with pine bark beetles from Mexico, including *O. pulvinisporum* sp. nov. *Mycological Research* **108**: 690–698.