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*

Brenda D. Wingfield[1], Irene Barnes[1], Z. Wilhelm de Beer[2], Lieschen De Vos[1], Tuan A. Duong[1], Aquillah M. Kanzi[1], Kershney Naidoo[1], Hai D.T. Nguyen[3], Quentin C. Santana[1], Mohammad Sayari[1], Keith A. Seifert[3], Emma T. Steenkamp[2], Conrad Trollip[1], Nicolaas A. van der Merwe[1], Magriet A. van der Nest[1], P. Markus Wilken[1], and Michael J. Wingfield[1]

[1]Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, 0028 South Africa; corresponding author e-mail: 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]Biodiversity (Mycology), Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, 960 Carling Avenue, Ottawa, Ontario, K1A 0C6, Canada

Abstract: The genomes of *Ceratocystis eucalypticola, Chrysoporthe cubensis, Chrysoporthe deuterocubensis, Davidsoniella virescens, Fusarium temperatum, Graphilbum fragrans, Penicillium nordicum* and *Thielaviopsis musarum* are presented in this genome announcement. These seven genomes are from plant pathogens and otherwise economically important fungal species. The genome sizes range from 28 Mb in the case of *T. musarum* to 45 Mb for *Fusarium temperatum*. These genomes include the first reports of genomes for the genera *Davidsoniella, Graphilbum* and *Thielaviopsis.* 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.

IMA Genome-F 5A

Draft genome sequence of *Ceratocystis eucalypticola*

Many species of *Ceratocystidaceae* have been studied extensively due to their significance as pathogens of agricultural forestry crops (Roux & Wingfield 2009), as well as their impact on natural woody ecosystems (Roux *et al.* 2007, Lee *et al.* 2015). The family includes eight genera accommodating more than 80 phylogenetically closely related but often morphologically similar species (van Wyk *et al.* 2013, de Beer *et al.* 2014, Mayers *et al.* 2015). These genera, as defined by de Beer *et al.* (2014), are clearly delimited based on a combination of phylogenetic inference, morphology, and in some cases distinct ecological partitioning. For example, the genus *Huntiella* accommodates species that are saprobes, whereas most species of *Ceratocystis* are pathogens of angiosperms.

Species of *Ceratocystis* include important pathogens of trees propagated as non-natives in plantations in the tropics and Southern Hemisphere (Roux & Wingfield 2013, Wingfield *et al.* 2013), including *Eucalyptus* (Laia *et al.* 1999, Roux *et al.* 2000, Roux *et al.* 2001, Barnes *et al.* 2003). Isolates of *Ceratocystis* from *Eucalyptus* in South Africa, related to those known to kill trees in these plantations, were described as the new species *C. eucalypticola* (van Wyk *et al.* 2012). The taxonomy of this species and some of its relatives remains open to debate (Fourie *et al.* 2015, Oliveira *et al.* 2015) and there is a clear need to gain a deeper understanding of species boundaries as well as issues relating to its biology and ecology.

The aim of this study was to sequence the genome of *C. eucalypticola* in order to allow for genomic analysis and comparisons with already available genomes from other *Ceratocystidaceae* (Wilken *et al.* 2013, van der Nest *et al.* 2014a, b). These comparisons, coupled with phylogenomic studies, will be useful in resolving the taxonomic debates ongoing in *Ceratocystis*. Additionally, these resources will provide a platform to characterise factors associated with pathogenicity and fungal ecological strategy, as well as provide an opportunity to study the evolution of these traits within a family of closely related fungi.

SEQUENCED STRAIN

South Africa: *Mpumalanga*: Sabie, isol. ex artificial wound of *Eucalyptus*, July 2002, *M. van Wyk & J. Roux* (CMW 9998, CBS 124017, PREM 60169 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The draft genome sequence of *Ceratocystis eucalypticola* (CMW9998) has been deposited at DDBJ/EMBL/GenBank with the accession number LJOA00000000. Here we describe version LJOA01000000.

METHODS

Genomic DNA of *Ceratocystsis eucalypticola* isolate CMW 9998 was sequenced using the Illumina HiSeq 2000 platform at the UC Davis Genome Centre, University of California, Davis (CA). Two libraries with medium insert sizes of 350 bp and 530 bp were used to generate pair-end sequences with read lengths of approximately 100 bases. CLC Genomics Workbench v. 7.5.1 (CLCBio, Aarhus, Denmark) was used to analyse the NGS-data, as well as to perform a *de novo* assembly. Reads of low quality (*P error* limit of 0.05) and/or terminal nucleotides were trimmed, with the remaining reads being retained for assembly. *De novo* genome assembly was performed with a word size of 64, and a bubble size of 100 bp. The raw reads were mapped back to the contigs in order to perform scaffolding, with an estimated paired distance ranging from 147 to 654 bp. The completeness of the assembled genome was assessed using the Benchmarking Universal Single-Copy Orthologs tool, BUSCO (Software v. 1.1b1 of May 2015) (Simão *et al.* 2015). BUSCO was performed on all contigs >1 kb, making use of the fungal lineage dataset. 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 genome of *Ceratocystis eucalypticola* had an estimated size of 31 260 284 bases, with an N50 of 116 489 and an average coverage of 80×. The *de novo* assembly generated a total of 2129 contigs, of which 961 were longer than 1 kb. The average scaffold length was 14 676 bases, with the largest scaffold being 726 305 bases in size. The GC content of the assembly was 47.9 %. BUSCO analysis defined the genome as 97 % complete with 1408 single-copy orthologs present, while 92 BUSCO orthologs were found to be duplicated. Only 30 BUSCO orthologs were missing or fragmented out of the possible 1438 groups searched from the fungal lineage dataset. Gene prediction resulted in a total of 7353 putative ORFs, at a gene density of approximately 235 ORFs/Mb.

The assembled genome of *C. eucalypticola,* with a size of approximately 31.2 Mb and 7353 ORFs, closely resembled those of other sequenced *Ceratocystis* spp. (Wilken *et al.* 2013, van der Nest *et al.* 2014a, b). The fungus had a genome size most similar to that of *C. fimbriata* (29.4 Mb, 7266 ORFs) and *C. manginecans* (31.7 Mb, 7494 ORFs), while the *C. albifundus* genome is slightly smaller (27.2 Mb) with only 6967 genes predicted. The genome size statistics for *Ceratocystis* are similar to those found in the genus *Huntiella*, with *H. omanensis* and *H. moniliformis* being 31.5 Mb and 25 Mb, respectively (van der Nest *et al.* 2014a, b). Gene predictions for *Huntiella* showed a slightly higher gene density when compared with those of *Ceratocystis* (243 ORFs/Mb on average), with *H. omanensis* having a density of 266 ORFs/Mb and *H. moniliformis* having 280 ORFs/Mb, respectively.

The availability of these resources will provide opportunities to answer questions regarding the similarities and differences seen in this genus. The genome data for *Ceratocystis s. str.* is particularly useful for exploring the species boundaries through phylogenomic analysis. This, in combination with genomic comparisons to other species within *Ceratocystidaceae,* will lead to a better understanding of the evolution of pathogenicity and other life history traits.

Authors: **C. Trollip*, T.A. Duong, M.A. van der Nest, I. Barnes, M.J. Wingfield, and B.D. Wingfield**
*Contact: Conrad.Trollip@fabi.up.ac.za

IMA Genome-F 5B

Draft genome sequences of *Chrysoporthe cubensis* and *C. deuterocubensis,* causal agents of Eucalyptus canker

Fungi in the genus *Chrysoporthe* are economically important pathogens of plantation grown *Eucalyptus* spp. and other members of *Myrtales* (Gryzenhout *et al.* 2004). These fungi cause serious stem canker disease, referred to as Chrysoporthe canker (Gryzenhout *et al.* 2004)*,* and are predominantly found in tropical and subtropical parts of the world where conditions favour their growth (Alfenas *et al.* 1982). Although Chrysoporthe canker has been successfully managed through propagation of disease resistant clones, it is still considered a threat since it can lead to substantial economic losses where resistance breeding is not in place (Wingfield 2003).

There are eight described species of *Chrysoporthe*, including *C. cubensis* (Hodges *et al.* 1976, 1979, Rodas *et al.* 2005), *C. doradensis* (Gryzenhout *et al.* 2005), *C. inopina* (Gryzenhout *et al.* 2006), and *C. hodgesiana* (Gryzenhout *et al.* 2004) which occur in South and Central America. *Chrysoporthe deuterocubensis* is primarily found in Southeast Asia, although introductions to Australia, China, Hawaii, and parts of East Africa have been suggested (Myburg *et al.* 2002, Nakabonge *et al.* 2006, van der Merwe *et al.* 2010). *Chrysoporthe zambiensis* and *C. syzygiicola* are found in Zambia (Chungu *et al.* 2010), while *C. austroafricana* is found only in southern Africa (Wingfield *et al.* 1989, Gryzenhout *et al.* 2004).

Table 1. Comparison of whole genome sequencing assembly features for *Chrysoporthe cubensis*, *C. Deuterocubensis,* and *C. austroafricana* (Wingfield *et al.* 2015).

Measure	C. austroafricana (CMW 2113)	C. deuterocubensis (CMW 8650)	C. cubensis (CMW 10028)
Genome size	44 669 169	43 969 123	42 624 564
GC	53.9 %	55 %	54.9 %
Contigs	6 416	6 500	3 326
Scaffolds (SSPACE)	3 232	2 574	954
Gene models (Augustus)	13 484	13 772	13 121
Completeness (BUSCO)	94 %	95 %	94 %

Chrysoporthe cubensis, *C. deuterocubensis,* and *C. austroafricana* have been isolated from native trees, suggesting that these fungi might be native to the regions where the trees are found (Myburg *et al.* 2003, Rodas *et al.* 2005, Heath *et al.* 2006). Interestingly, despite the distinct geographical distribution, these species seem to be closely related (Chungu *et al.* 2010, van der Merwe *et al.* 2010). Unfortunately, there is limited available information regarding the evolution of *Chrysoporthe* species.

The genome of *C. austroafricana* was recently sequenced and released in the public domain (Wingfield *et al.* 2015). This is the only whole genome sequence resource available for the genus *Chrysoporthe*. Additional genomic resources could enhance further understanding of the biology of this assemblage of fungi, through genome-wide comparisons. The aim of this study was thus to sequence the genomes of *C. cubensis* (isolate CMW 10028) and *C. deuterocubensis* (isolate CMW 8650).

SEQUENCED STRAINS

Chrysoporthe cubensis: **Colombia**: Timba, 2002, *C.A. Rodas* (CMW 10028, PREM 58311 – dried culture).
Chrysoporthe deuterocubensis: **Indonesia**: Sulawesi, 2001, *M.J. Wingfield* (CMW 8650, CBS 115719, PREM 58018 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The *Chrysoporthe cubensis* isolate number CMW 10028 and *C. deuterocubensis* isolate CMW 8650 Whole Genome Shotgun projects were deposited in GenBank with accession numbers LJCY00000000 and LJDD00000000**,** respectively. The version described here is LJCY00000000 and LJDD00000000 for *C. cubensis* and *C. deuterocubensis*, respectively.

MATERIALS AND METHODS

Genomic DNA was extracted using a modified protocol (Steenkamp *et al.* 1999) from isolate CMW 10028 (*Chrysoporthe cubensis*) and CMW 8650 (*C. deuterocubensis*) mycelium obtained from 7-d-old fungal cultures. The Illumina MiSeq paired-end sequencing protocol at the Agricultural Research Council (ARC, South Africa) was used to obtain whole genome sequence data. To assemble the paired-end MiSeq sequences, CLC Genomics Workbench v. 7.5.1 (CLCBio, Aarhus, Denmark) was used. The assemblies were subsequently scaffolded using SSPACE v. 2.0 (Boetzer *et al.* 2011), which included unused MiSeq reads from the CLC Genomics Workbench assembly. The AUGUSTUS (Stanke & Morgenstern 2005) protein coding gene prediction software was used for *de novo* annotation of protein coding gene models using *Neurospora crassa* and *Fusarium graminearum* as references. Genome completeness was assed using BUSCO (Benchmarking Universal Single-Copy Orthologs) which utilizes single-copy orthologs to predict genome completeness (Simão *et al.* 2015).

RESULTS AND DISCUSSION

The approximate size of the *Chrysoporthe cubensis* genome was 42 624 564 base pairs (bp) including gaps, while the *C. deuterocubensis* assembly was 43 969 123 bp in size. These figures were calculated from 989 and 2 599 scaffolds for *C. cubensis* and *C. deuterocubensis*, respectively. From the AUGUSTUS analysis, 12 435 gene models were predicted from the *C. cubensis* genome, while 13 098 gene models where predicted in the *C. deuterocubensis* genome. Despite the differences observed in the assembly statistics, the CEGMA analysis for genome completeness in both *C. cubensis* and *C. deuterocubensis* was predicted at 95.16 %.

Compared to the closely related *C. austroafricana* genome, that of *C. cubensis* was slightly smaller, while the *C. deuterocubensis* genome was slightly larger. Similarly, *C. cubensis* had fewer predicted gene models than either *C. austroafricana* or *C. deuterocubensis* (Table 1). In terms of gene content, the *Chrysoporthe* spp. genomes were slightly larger than that of the distantly related *Cryphonectria parasitica* (43.9 Mb, 11 184 gene models) (http://genome.jgi.doe.gov/Crypa2/Crypa2.info.html) although the genome sizes were relatively close, and the model filamentous fungi, *Neurospora crassa* (39.9 Mb, 10 082 gene models) (Galagan *et al.* 2003) and *Magnoporthe grisea* (40.3 Mb, 11 109 gene models) (Dean *et al.* 2005).

The significance of differences observed in genome size and the number of predicted genes among the *Chrysoporthe* species is not known. However, it might be speculated that their geographical distribution could have played a role in the evolution of these genomes. The availability of these genomes will make it possible to answer such phylogeographic

Table 2. Summary of whole nuclear genome DNA sequence assembly statistics (Wilken *et al.* 2013, van der Nest *et al.* 2014a, b, Belbahri 2015, Wingfield *et al.* 2015).

Species	*Ceratocystis fimbriata*	*Huntiella moniliformis*	*Ceratocystis manginecans*	*Ceratocystis albifundus*	*Huntiella omanensis*	*Thielaviopsis punctulata*	*Ceratocystis platani*	*Thielaviopsis musarum*	*Ceratocystis eucalypticola*	*Davidsoniella virescens*
Summary Data										
Accession Numbers	APWK00000000	JMSH00000000	JJRZ00000000	JSSU00000000	JSUI00000000	LAEV00000000	LBBL00000000	LKBB00000000	LJOA01000000	LJZU00000000
Genome size (bases)	29 410 862	25 429 610	31 706 104	27 149 029	31 502 652	28 117 660	29 181 400	28 493 324	31 260 284	33 645 160
Genome Coverage	20x	38x	22x	24x	9x	75x	655x	95x	80x	82x
N50 contig size (bases)	42 879	191 280	77 070	58 335	41 324	18 915	77 580	103 017	116 489	118 189
GC content (%)	48.065	48	47.9	48.6	47.6	48.3	48.2	49.17	47.9	44.50
Genome completeness (%)	96.77[1]	96.4[1]	96.4[1]	96.4[1]	90.83 (completed genes)[1] 95.97 (partial genes)[1]	98.38[1]	97[2]	96[2]	97[2]	97[2]
Assembly software	Newbler	AbySS	Velvet	Velvet	Velvet	SPAdes	CLC Genomics	Velvet	CLC Genomics	CLC Genomics
Number of Contigs	3688	680	2234	1958	8127	3751	1213	672	2129	1500
Retained contigs	2641[a]	365[a]	980[a]	939[b]	1638[b]	2379[b]	1213[b]	541[b]	961[b]	561[b]
ORFs	7266	6832	7494	6967	8359	5480	5963	6963	7353	6953
Gene density (ORFs/Mb)	246	269	236	257	266	195	204	244	235	207

[1] CEGMA completeness score (Parra *et al.* 2007).

[2] BUSCO completeness score.

a = contigs greater than 500 nucleotides.

b = contigs greater than 1000 nucleotide.

questions and will aid in addressing questions relating to the biology of *Chrysoporthe* species.

Authors: **A.M. Kanzi, B.D. Wingfield, M.J. Wingfield, E.T. Steenkamp, and N. A. van der Merwe***
*Contact: albe.vdmerwe@up.ac.za

IMA Genome-F 5C

Draft nuclear genome sequence for *Davidsoniella virescens*, the causal agent of sapstreak disease in hardwood trees

The newly recognized genus *Davidsoniella* (de Beer *et al.* 2014) includes species previously accommodated in the *Ceratocystis coerulescens s. lat.* clade. *Davidsoniella virescens* is a tree pathogen that infects hardwood trees such as *Acer saccharum* (sugar maple) in eastern North America (Davidson 1944). This fungus is highly pathogenic to sugar maple in plantations, where it feeds on the sugars and other carbohydrates in the wood of the trees (Bal *et al.* 2013). The disease caused by *D. virescens* is commonly referred to as sapstreak and the fungus affects the internal wood chemistry where it has been implicated in the production of volatiles that can enhance the growth of other fungi (Wargo & Harrington 1991).

The aim of this study was to assemble a draft nuclear genome sequence of *D. virescens,* which would ultimately allow for comparative studies with other sequenced genomes in *Ceratocystidaceae*. The genomes of seven other species of *Ceratocystidaceae* are publically available to aid with such a comparative analyses. These include *Ceratocystis fimbriata*, a pathogen affecting sweet potatoes (Wilken *et al.* 2013); the canker and wilt disease causing *C. albifundus*, occurring on *Acacia mearnsii* trees (van der Nest *et al.* 2014a); the mango wilt pathogen, *C. manginecans* (van der Nest *et al.* 2014b); species in the related genus *Huntiella*, *H. moniliformis* and *H. omanensis*, saprobic fungi usually found on freshly cut or wounded logs (van der Nest *et al.* 2014a), the causal agent of black scorch disease in date palms, *Thielaviopsis punctulata* (Wingfield *et al.* 2015) and the plane tree pathogen *C. platani* (Belbahri 2015). Two additional *Ceratocystidaceae* genomes, those for *C. eucalypticola* and *Thielaviopsis musarum*, are included in this issue and collectively, these will add value to the comparative analysis of the genomes across this family. Understanding the general biology of *D. virescens* will further assist in developing a deeper understanding of sap streak and potentially contribute to disease management strategies.

SEQUENCED STRAIN

USA: *New Hampshire*: isol. ex *Acer saccharum*, Aug. 1987, *D. Houston* (CMW 17339 = CBS 130772; PREM 61293 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The Whole Genome Shotgun project of the *Davidsoniella virescens* genome has been deposited at DDBL/EMBL/GenBank under the accession no. LJZU00000000. The version described in this paper is version LJZU01000000.

METHODS

Davidsoniella virescens isolate CMW 17339 was used in this study. Cultures were grown at 25 °C on 2 % malt extract agar (MEA: 20 % w/v, Biolab, Midrand, South Africa) supplemented with 100 µg/L thiamine. Total genomic DNA was isolated from 2-wk-old cultures using a phenol-chloroform method previously described (Roux *et al.* 2004). Sequencing was carried out on the Genomics Analyzer IIx platform (Illumina) at the Genome Centre (University of California at Davis, CA). Paired-end libraries with insert fragments of 350 and 600 bases were used to generate the read lengths of 100 bases. CLCBio Genomics workbench software v. 7.5.1 (CLCBio, Aarhus, Denmark) was used for quality assessment and *de novo* assembly. Poor-quality reads (limit of 0.05) and/or terminal nucleotides were discarded. The remaining reads were assembled *de novo* using a word size of 64 with a bubble size of 100 base pairs. Scaffolding with an estimated pair distance of 99 to 562 base pairs was performed by mapping raw reads back to the contigs. Only contigs greater than 1000 bases were retained. Predictions of open reading frames (ORFs) based on the gene model for *Fusarium graminearum* (http://bioinf.uni-greifswald.de/augustus) were made using AUGUSTUS (Stanke *et al.* 2006). The quantitative assessment of the genome assembly completeness was assessed against the Benchmarking Universal Single-Copy Orthologs software program, BUSCO (Simão *et al.* 2015) using contigs greater than 1000 bases in length.

RESULTS AND DISCUSSION

Davidsoniella virescens had an estimated nuclear genome size of 33 645 160 bases. The N50 value was determined to be 118 189 bases generating a mean GC content of 44.50 %. A total of 563 contigs were produced from the CLCBio assembly, of which 561 were retained after excluding the mitochondrial scaffolds. The AUGUSTUS gene prediction pipeline estimated 6 953 ORFs. This draft genome assembly had a BUSCO completeness score of 97 % indicating that the core eukaryotic genes were present. From this analysis, 1404 single-copy genes were observed, of which 73 were duplicated genes. Of the 1 438 genes searched, only 2.2 % were classified as fragmented or missing. A gene density of 207 ORFs per Mb was observed for the 6 953 genes predicted.

Davidsoniella virescens has the largest estimated genome size (33 Mb) of all *Ceratocystidaceae* genomes sequenced thus far (Table 2). The dissimilarity in the coverage, N50 values and number of contigs can be attributed to the different sequencing and assembly

platforms used to generate the data (Table 2). The retention of the final number of contigs processed in gene prediction tools also differed because some researchers have chosen to retain smaller contigs (greater than 500 nucleotides). *Davidsoniella virescens* had a similar GC content and genome completeness to the genomes sequenced of the other species in *Ceratocystidaceae* (Table 2). Adding to the growing number of sequenced and assembled genomes, the *D. virescens* genome provides a powerful resource to aid in its phylogenetic classification in *Ceratocystidaceae*. Similar or shared biological features can now be identified due to the availability of these genomes.

Authors: **K. Naidoo*, C. Trollip, P.M. Wilken, M.J. Wingfield, and B.D. Wingfield**
*Contact: Kershney.Naidoo@fabi.up.ac.za

IMA Genome-F 5D

Nuclear genome assembly for the maize pathogen *Fusarium temperatum*

Fusarium temperatum (formerly *F. subglutinans* group 1, de Vos *et al.* 2014) is an important mycotoxin-producing pathogen of maize (Scauflaire *et al.* 2011). This fungus is a member of the *Fusarium fujikuroi* complex which includes numerous pathogens responsible for destructive diseases of many plants (Kvas *et al.* 2009). Due to the economic importance of the complex, the whole genome sequences for several of its members have been determined and are publicly available. These include *F. verticillioides* (*Fusarium* Comparative Sequencing Project, Broad Institute of Harvard and MIT; http://www.broad.mit.edu), *F. circinatum* (Wingfield *et al.* 2012), *F. fujikuroi* and *F. mangiferae* (Wiemann *et al.* 2013), as well as *F. nygamai* (Wingfield *et al.* 2015).

To complement these genomic resources, genetic linkage maps for some of these fungi are also available (Jurgensen *et al.* 2002, de Vos *et al.* 2007). For example, a genetic linkage map available for a hybrid cross between *F. circinatum* and *F. temperatum* (de Vos *et al.* 2007) has been used as a framework in the analyses of certain loci and traits in these fungi (de Vos *et al.* 2011, 2013). Most recently, an analysis of the genomic architecture of species in this complex, allowed the anchoring of this genetic linkage map to the genomic sequence data for *F. verticillioides* and *F. fujikuroi* (de Vos *et al.* 2014). The aim of this study was therefore to determine the whole genome sequence for the other parent (*F. temperatum*) in this hybrid cross. The availability of genome data for this fungus would allow comparisons to the other sequenced members of the *F. fujikuroi* complex as well as contribute to improving our knowledge of the genetic processes and properties underlying the biology of these important fungi.

SEQUENCED STRAIN

Mexico: *Texcoco*: isol. *Zea mays* ssp. *mexicana* seeds (teosinte), Nov. 1996*, A.E. Desjardins, R.D. Plattner & T.R. Gordon* (CMW 40964, CBS 138287; PREM 61039 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The *Fusarium temperatum* genomic sequence data has been deposited at DDBJ/EMBL/GenBank under the accession LJR00000000. The version described in this paper is version LJR01000000.

METHODS

DNA was extracted from *Fusarium temperatum* grown on ½ PDA (Iturritxa *et al.* 2011). One mate-pair (2 840 bp average insert size) and two paired-end (average insert sizes of 213 and 476 bp) libraries were prepared and subjected to 100 bp Illumina HiSeq 2000 sequencing at Fasteris (Geneva). After removing poor quality reads using CLC Genomics Workbench v. 6.5 (CLCbio, Aarhus, Denmark), sequences were assembled using ABySS v. 1.3.7 (Simpson *et al.* 2009). Closing of gapped regions within the scaffolds was done using GapFiller v. 1.11 (Boetzer & Pirovano 2012). The completeness of the genome assembly was evaluated using CEGMA (Parra *et al.* 2008) and putative open reading frames (ORFs) were predicted using AUGUSTUS (Hoff & Stanke 2013) together with the gene models for *F. graminearum* and cDNA data from the closely related *F. circinatum* (Wingfield *et al.* 2012). By making use of MUMmer v. 3.22 (Kurtz *et al.* 2004), the *F. temperatum* scaffold sequences were compared to the chromosomes of two other sequenced members in the *F. fujikuroi* complex, *F. fujikuroi* (Wiemann *et al.* 2013) and *F. verticillioides* (*Fusarium* Comparative Sequencing Project) (De Vos *et al.* 2014).

RESULTS AND DISCUSSION

Assembly of 188 294 812 good quality reads yielded a draft genome for *Fusarium temperatum* that consisted of 45 458 781 bp with 414x coverage. This assembly consists of 43 scaffolds with an N50 of 4 506 647 bp and an average scaffold size of 1 057 181 bp. Based on the CEGMA analysis, this draft genome is 97.38 % complete (Parra *et al.* 2008). The GC content is 47 %. The assembly contains 14 284 putative ORFs with an average length of 1576 bp and an average density of 314 ORFs/Mb. These genome statistics for *F. temperatum* are comparable to those of the other sequenced *Fusarium* members (*Fusarium* Comparative Sequencing Project, Wingfield *et al.* 2012, 2015, Wiemann *et al.* 2013), which highlights the genomic similarities amongst the members in the *F. fujikuroi* complex.

Sequence comparisons of the sixteen largest scaffolds (which accounts for 99.56 % of the total genome size) to
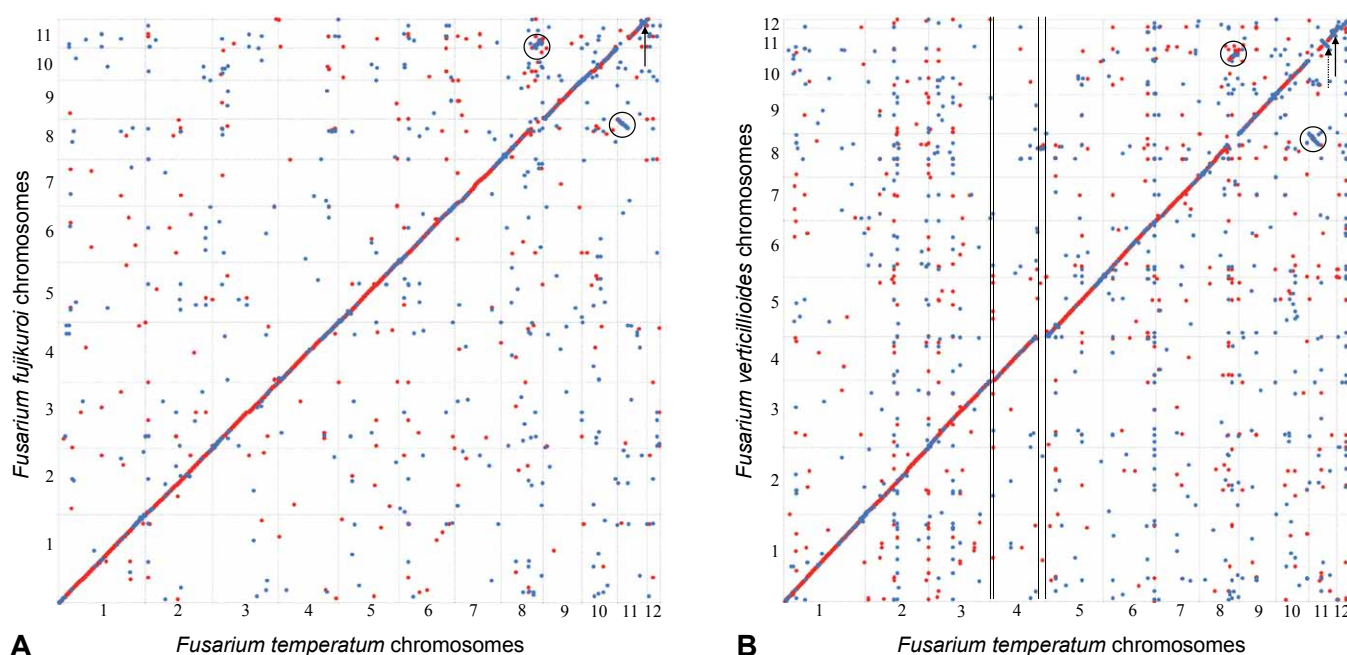
Fig. 1. Whole genome comparisons of (A) *Fusarium verticillioides* to *F. temperatum,* and (B) *F. fujikuroi* to *F. temperatum*. Dotplot alignments of *F. verticillioides* scaffolds (placed into chromosomes; de Vos *et al*. 2014) and *F. fujikuroi* chromosomes (Wiemann *et al*. 2013) against the 16 largest *F. temperatum* scaffolds. Forward matches are indicated by red dots, reverse matches with blue dots. The black circles show the reciprocal translocation between chromosome 8 and 11 – the single scaffold representing chromosome 11 has a portion that aligns to chromosome 8 and a portion that align to chromosome 11 of *F. verticillioides* and *F. fujikuroi*. Solid arrows are indicative of inversions in *F. temperatum*, while the dotted arrow indicates an inversion in *F. fujikuroi*, when compared to the two other *Fusarium* spp.

the information for the *F. verticillioides* and *F. fujikuroi* genomes suggests that these scaffolds likely make up the 12 chromosomes predicted for species in the *F. fujikuroi* complex (Xu *et al.* 1995). This was further illustrated by the alignments of the *F. temperatum* scaffolds to the chromosome sequences for *F. verticillioides* and *F. fujikuroi* (Fig. 1). These alignments also confirmed the reciprocal translocation in *F. temperatum* and *F. circinatum* observed by de Vos *et al.* (2014) between chromosomes 8 and 11 (Fig. 1). The subtelomeric regions missing from chromosome 4 in *F. fujikuroi* (Wiemann *et al.* 2013) are present in *F. temperatum* (Fig. 1B), confirming that the shortened chromosome 4 is *F. fujikuroi*-specific.

Like *F. verticillioides*, *F. temperatum* also harboured the large inversion previously reported in chromosome 11 between *F. verticillioides* and *F. fujikuroi* (Wiemann *et al.* 2013, de Vos *et al.* 2014) (Fig. 1B), although *F. temperatum* appears to have an additional inversion in this chromosome when compared to *F. verticillioides* and *F. fujikuroi* (Fig. 1). Sequence comparisons also revealed that chromosome 12 is present in this *F. temperatum* assembly, albeit 1.42 times larger than its counterpart in *F. fujikuroi* (Wiemann *et al.* 2013). Within the *F. fujikuroi* complex, chromosome 12 has also been shown to be dispensable as well as strain-specific (Xu *et al.* 1995, Jurgenson *et al.* 2002, Ma *et al.* 2010, Wiemann *et al.* 2013, van der Nest *et al.* 2014a). Collectively, chromosome 11 and 12 therefore seems to be the most variable of the chromosomes in this complex. The addition of the whole genome sequence of *F. temperatum*, to the other sequenced members of the *F. fujikuroi* complex, would assist phylogenomic studies into the evolution and biology of these important fungi.

Authors: **L. De Vos*, Q.C. Santana, B.D. Wingfield, M.A. van der Nest, M.J. Wingfield, and E.T. Steenkamp**
Contact: Lieschen.devos@up.ac.za

IMA Genome-F 5E

Draft genome sequence of *Graphilbum fragrans*

Graphilbum is one of six currently recognized genera in *Ophiostomatales* (*Ascomycota, Sordariomycetes*) (de Beer & Wingfield 2013). The genus includes nine named species and some undescribed taxa (de Beer & Wingfield 2013). As with most other species of *Ophiostomatales*, species of *Graphilbum* are commonly found associated with coniferous hosts. *Graphilbum fragrans* was first described in 1954 from Sweden (Mathiesen-Käärik 1954), where it was initially treated in *Graphium* (as *G. fragrans*). This species was later reported from conifers or conifer-infesting beetles from various other countries including Australia, Canada, China, New Zealand, Korea, Poland, South Africa, Spain, and the USA (Harrington *et al.* 2001, Jacobs *et al.* 2003, Zhou *et al.* 2006, Kim *et al.* 2007, Romon *et al.* 2007, Paciura *et al.* 2010, Jankowiak & Bilański 2013).

The availability of whole genome sequences and recent advancements in genome analyses have contributed to a better understanding of the biology, pathogenicity and evolutionary processes in fungi. A number of genomes of species in *Ophiostomatales* have been sequenced and analysed, however, these include only species in *Leptographium*, *Ophiostoma,* and *Sporothrix* (DiGuistini *et al.* 2011, Haridas *et al.* 2013, Khoshraftar *et al.* 2013, Teixeira *et al.* 2014, van der Nest *et al.* 2014, Wingfield *et al.* 2015). The aim of this study was to generate the genome sequence for *G. fragrans*, the first genome available for the genus *Graphilbum* and thus to provide a basis for comparison between the genera of Ophiostomatales.

SEQUENCED STRAIN

South Africa: *Mpumalanga*: from *Hylastes angustatus* infesting *Pinus patula*, 1999, *X.D.* Zhou (culture CMW 19357 = CBS 138720; PREM 61294 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The genomic sequence of *Graphilbum fragrans* (CMW 19357, CBS 138720) has been deposited at DDBJ/EMBL/GenBank under the accession LLKO00000000. The version described in this paper is version LLKO01000000.

METHODS

Methods for DNA extraction, genome sequencing, assembly and annotation were similar to those used for *Leptographium lundbergii* (Wingfield *et al.* 2015). Total genomic DNA was extracted following the protocol of Duong *et al.* (2013). 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 first subjected to quality filtering and trimming, followed by *de novo* assembly using CLC Genomics Workbench v. 8.0.1 (CLCBio, Aarhus, Denmark). Genome completeness was estimated using BUSCO (Simão *et al.* 2015). Total number of gene models was predicted using the MAKER genome annotation pipeline (Cantarel *et al.* 2008).

RESULTS AND DISCUSSION

Over 26.2 million reads were obtained after filtering and trimming. *De novo* assembly using CLC Genomic Workbench resulted in 80 scaffolds that were over 500 bp in size. The assembly had a N50 value of 973.6 kb and the longest scaffold was 2.66 Mb. The genome of *Graphilbum fragrans* was estimated to be 34.26 Mb, with the mean GC content of 55.7 %. We assessed the completeness of the obtained genome by running BUSCO on the resulting assembly using the fungal reference dataset and obtained BUSCO values of C: 97 % [D: 5.8 %], F: 1.8 %, M: 0.6 %, n: 1348

(C: complete, [D: duplicated], F: fragmented, M: missed, n: genes), indicating that the obtained genome sequence should cover most of the organism's gene space. Genome annotation using MAKER resulted in 10 633 gene models filtered based on MAKER max build (8 942 gene models if MAKER standard build was applied) (Campbell *et al.* 2014). Of 10 633 gene models predicted using MAKER max build, 8102 were multi-exonic genes, mean intron length was 121.1 bp and mean exon length was 552.8 bp. The genome of *G. fragrans*, which is the first genome reported for *Graphilbum*, represents a useful resource for various comparative genomic and systematic studies in *Ophiostomatales*.

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 5F

Draft genome sequence of *Penicillium nordicum* DAOMC 185683

Penicillium nordicum is classified in the subgenus *Penicillium* section *Fasciculata* (Houbraken & Samson 2011) and is commonly isolated from cheese, nuts and other fat and protein rich substrates like salami and ham (Frisvad & Samson 2004). The importance of this fungus relates to its production of the regulated mycotoxin ochratoxin A (OTA), which is hepatoxic, nephrotoxic, teratogenic and immunotoxic in animals (Pitt *et al.* 2012), known to promote oxidative DNA damage by the production of reactive oxygen species and to generate DNA adducts (Hadjeba-Medjdoub *et al.* 2012), and is classified as a possible human renal carcinogen (group 2B) by the International Agency for Research on Cancer (IARC, Pitt *et al.* 2012).

OTA is also produced by *P. verrucosum*, the sister species to *P. nordicum* (Samson *et al.* 2004), and by several species of *Aspergillus* (Visagie *et al.* 2014a). Despite the importance of OTA in grain, coffee and grape products, its biosynthetic pathway has yet to be fully elucidated. However, there is evidence that a gene cluster including an alkaline serine protease, a polyketide synthase and a non-ribosomal peptide synthase may play a role in OTA production in *P. nordicum* (Karolewiez & Geisen 2005, Geisen *et al.* 2006). In this study, we sequenced and annotated a genome draft of *P. nordicum* DAOMC 185683, as part of our investigation of genes regulating OTA production in *Penicillium* species.

SEQUENCED STRAIN

Canada: *Alberta*: Brooks, isolated from *Lycopersicon esculentum* (tomato), collected and isolated 24 Jan. 1983, *R.J. Howard GT-78.* (DAOMC 185683). Originally identified as *Penicillium aurantiogriseum* by John D. Bissett; re-identified as *P. nordicum* by Keith A. Seifert in 2012.

NUCLEOTIDE SEQUENCE AND RAW READS ACCESSION NUMBERS

This Whole Genome Shotgun project was deposited at DDBJ/EMBL/GenBank under accession LHQQ00000000. The version described in this paper is version LHQQ01000000. Raw reads were deposited in NCBI SRA (http://www.ncbi.nlm.nih.gov/sra) accession number SRR2146067.

DNA EXTRACTION, WHOLE GENOME SEQUENCING AND ASSEMBLY

Penicillium nordicum DAOMC 185683 was grown on Blakeslee's malt extract agar for 7 d at 25 °C (Visagie *et al.* 2014b). To make a spore suspension, the colonies were flooded with 5 mL of sterile distilled water. One mL of this spore suspension was inoculated in 100 mL of Blakeslee's malt extract broth and was left shaking at 300 rpm at 25 °C for 6 d. To obtain fungal tissue for DNA extraction, cells were removed from the liquid culture by filtration. DNA was extracted with the OmniPrep kit for fungi (G-Biosciences) following the manufacturer's protocol. Whole-genome sequencing (paired-end with 101 bp) was performed on an Illumina HiSeq 2500 with TrueSeq V3 chemistry at the National Research Council Canada in Saskatoon (Saskatchewan, Canada).

The quality of genomic reads was determined with FastQC v. 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). 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. Species identification was confirmed by comparing the internal transcribed spacer (KJ834513) and beta-tubulin (KJ834476) barcode sequences of *P. nordicum* (Visagie *et al.* 2014b) against the assembled genomic scaffolds using BLASTn. Assembly statistics were generated with QUAST v. 2.3 (Gurevich *et al.* 2013).

The assembly was assessed 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 genome 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.

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 UniProt/Swiss-Prot manually curated fungal protein data set by 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. Genome Annotation Generator (http://genomeannotation.github.io/GAG/) and tbl2asn (http://www.ncbi.nlm.nih.gov/genbank/tbl2asn2/) were used to validate annotations.

RESULTS AND DISCUSSION

Approximately 22 million reads, comprising 2.2 Gbp of data, were assembled into 996 scaffolds resulting in an assembly of 30.8 Mb with a GC content of 47.8 %. The N50 value was 92.3 Kb and the longest scaffold was 391 Kb. The median nucleotide coverage across the whole assembly was 57x. The assembled genome had a CEGMA score of 96.8 % when calculated from the complete gene set and 98.4 % when calculated from both partial and complete gene sets. Assessment of the completeness of the genome using BUSCO groups for fungi resulted in values of C: 99 % , [D: 6.8 %], F: 0.7 %, M: 0.1 %, n: 1438 (C: complete, [D: duplicated], F: fragmented, M: missed, n: genes). Therefore, the assembled genome covered most of the organism's gene content. After annotation and validation, the genome contained 12 959 protein-coding genes. Of all suggested gene models, 12 448 were complete (96.0 %), but 511 gene models lacked a start codon, stop codon or both (4.0%). Mean gene length was 1388 bp, mean exon length was 437 bp and mean intron length was 85 bp. One other *P. nordicum* genome is accessioned in NCBI (JNNR), sequenced from a strain isolated from crop fields in Karlsruhe, Germany (UASWS BFE487). A comparative analysis has not yet been published, but as with our strain, the genome size was 30.4 Mb contained in 915 scaffolds, but the genome has less than half the coverage (at 20×) and only 46 genes were annotated.

This draft genome of a North American strain of *P. nordicum,* the first record of this species from Canada, represents a useful resource for biogeographical and comparative genomic studies of OTA (ochratoxin A) producing species of *Penicillium*, *Aspergillus,* and other related fungi. It will facilitate future gene knockout studies aiming to uncover the full OTA biosynthetic pathway in *P. nordicum*.

Authors: **H.D.T. Nguyen* and K.A. Seifert**
*Contact: hai.nguyen.1984@gmail.com

IMA Genome-F 5G

Draft genome sequence of the banana pathogen *Thielaviopsis musarum*

Thielaviopsis musarum is a pathogen of banana (Mitchell 1937, Riedl 1962) that typically infects banana fruits during maturation. This is especially true under conditions of high

humidity, darkness and moderately steady temperatures (Riedl 1962).

Thielaviopsis musarum was previously treated as *Ceratocystis musarum,* but was transferred to *Thielaviopsis* as part of a major revision of the family *Ceratocystidaceae* by de Beer *et al.* (2014). The fungus was first reported as a new variety of *C. paradoxa* causing stem-end rot of banana in Australia (Mitchell 1937). Riedl (1962) isolated a similar fungus from banana stems in Vienna although the plant material probably originated elsewhere, and described it as a new species distinct from *C. paradoxa*. Although some authors regarded the species from banana as distinct from *C. paradoxa* (de Hoog 1974, Nag Raj & Kendrick 1975), others viewed *C. musarum* as a synonym of *C. paradoxa* (Upadhyay 1981). These disputes have, however, been settled with DNA-based studies (Harrington 2009, de Beer *et al.* 2013) and *T. musarum* is now recognised as a distinct species in *Thielaviopsis.* The aim of this study was to sequence and assemble the whole genome of an isolate of *T. musarum*. This was undertaken to provide information allowing for the recognition of fungal genes that are associated with pathogenicity and other important biological properties in members of *Ceratocystidaceae*.

SEQUENCED STRAIN

New Zealand: on *Musa* sp., *T.W. Canter Vischer* (PREM 60962 – epitype, dried culture; CMW 1546 – ex-epitype culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBERS

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LKBB00000000. The version described in this paper is version LKBB00000000.

METHODS

Isolate CMW1546 (CBS 139399) of *Thielaviopsis musarum* was grown in malt extract agar (MA). High quality DNA was isolated from harvested mycelium (Raeder & Broda 1985) and sequencing was performed using the Genomics Analyzer IIx platform (Illumina) using paired-end libraries with insert sizes of approximately 350 and 600 bases. Reads with an average length of 97 bases were quality-trimmed using the software package CLC Genomics Workbench v. 6.0.1 (CLCBio, Aarhus, Denmark). The quality-filtered reads were assembled using the Velvet *de novo* assembler (Zerbino & Birney 2008), with an optimized k-mer size of 77. We used SSPACE v.2.0 (Boetzer *et al.* 2011) to assemble contigs into scaffolds and gaps were filled using GapFiller v. 2.2.1 (Boetzer & Pirovano 2012). The completeness of the assembled genome was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) tool, (Software v. 1.1b1 of May 2015; Simão *et al.* 2015). The BUSCO analysis was performed on all contigs >1 kb, making use of the fungal lineage dataset.

AUGUSTUS (Hoff & Stanke 2013) and the gene models for *Fusarium graminearum* were used to identify putative open reading frames (ORFs).

RESULTS AND DISCUSSION

The *Thielaviopsis musarum* draft genome had an estimated size of 28 493 324 bases, a 95× coverage, N50 contig size of 103 017 bases and a mean GC content of 49.17 %. The assembly was composed of 672 contigs, of which 541 were larger than 1 kb. Based on the BUSCO analysis, this assembly is 96 % complete. A total of 1392 single-copy BUSCO orthologs were present, of which 78 were duplicated. Out of a possible 1438 BUSCO groups searched, 11 BUSCO groups were missing or fragmented. The final assembly was predicted to encode 6 963 putative ORFs at a density of 244 ORFs/Mb.

The *T. musarum* genome appears to be relatively small and harbours fewer genes than other *Sordariomycetes* (*e.g., Fusarium fujikuroi,* 43.8 Mb with 14813 ORFs; *Cryphonectria parasitica,* 43.9Mb with 11184 ORFs) (Wiemann *et al.* 2013; http://genome.jgi.doe.gov/Crypa2/Crypa2.home.html).   The genome size of *T. musarum* was, however, in the same range of some species of *Ceratocystidaceae* such as *Ceratocystis manginecans* (of 31.7 Mb with 7494 ORFs*), C. albifundus albifundus* (27.1 Mb with 6967 ORFs)*, C. fimbriata* (29.4 Mb with 7266 ORFs)*, Huntiella omanensis* (31.5 Mb with 8395 ORFs*)* and *H. moniliformis* (25.5 Mb with 6832 ORFs) (Wilken *et al.* 2013, van der Nest *et al.* 2014a, b)*.*

The *Thielaviopsis musarum* genome was only marginally larger than that of *T. puntulata* (accession number: LAEV00000000) with its 28.1 Mb genome. However, *T. puntulata* was reported to encode 5480 ORFs (Wingfield *et al.* 2015) as opposed to the 6963 of *T. musarum*, suggesting a higher ORF density for the latter (i.e., 244 ORFs/Mb for *T. musarum vs.* 195 ORFs/Mb for *T. puntulata*). Future research should, therefore, consider whether these differences in genome size and ORF density could be ascribed to differences in the methodologies used to sequence and annotate the respective genomes. Overall, these genomes will provide interesting perspectives regarding the development and evolution of important biological traits in these fungi.

Authors: **M.A Sayari*, C. Trollip, K. Naidoo, B.D. Wingfield, and M.J. Wingfield**
*Contact: Mohammad.Sayari@fabi.up.ac.za

ACKNOWLEDGEMENTS

REFERENCES

Alfenas A, Hubbes M, Couto L (1982) Effect of phenolic compounds from *Eucalyptus* on the mycelial growth and conidial germination of *Cryphonectria cubensis*. *Canadian Journal of Botany* **60**: 2535–2541.

Bal TL, Richter DL, Storer AJ, Jurgensen MF (2013) The relationship of the sapstreak fungus, *Ceratocystis virescens*, to sugar maple dieback and decay in Northern Michigan. *American Journal of Plant Sciences* **4**: 436–43.

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.

Barnes I, Roux J, Wingfield BD, O'Neill M, Wingfield MJ (2003) *Ceratocystis fimbriata* infecting *Eucalyptus grandis* in Uruguay. *Australasian Plant Pathology* **32**: 361–366.

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.

Campbell MS, Holt C, Moore B, Yandell M (2014) Genome annotation and curation using MAKER and MAKER-P. *Current Protocols in Bioinformatics*: 4.11. 1-4.11. 39.

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.

Chungu D, Gryzenhout M, Muimba-Kankolongo A, Wingfield MJ, Roux J (2010) Taxonomy and pathogenicity of two novel *Chrysoporthe* species from *Eucalyptus grandis* and *Syzygium guineense* in Zambia. *Mycological Progress* **9**: 379–393.

Davidson RW (1944) Two American hardwood species of *Endoconidiophora* described as new. *Mycologia* **45**: 579–86.

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, Seifert KA, Wingfield MJ (2013) A nomenclator for ophiostomatoid genera and species in the *Ophiostomatales* and *Microascales*. In: *The Ophiostomatoid Fungi: expanding frontiers* (Seifert KA, de Beer ZW, Wingfield MJ, eds): 245–322. [CBS Biodiversity Series vol. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

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 vol. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

de Hoog GS (1974) The genera *Blastobotrys*, *Sporothrix*, *Calcarisporium* and *Calcarisporiella* gen. nov. *Studies in Mycology* **7**: 1–84.

de Vos L, Myburg AA, Wingfield MJ, Desjardins AE, Gordon TR, *et al.* (2007) Complete genetic linkage maps from an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. *Fungal Genetics and Biology* **44**: 701–714.

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.

de Vos L, van der Merwe NA, Wingfield MJ, Myburg AA, Wingfield BD (2013) Transmission ratio distortion in an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. *Genes and Genomics* **35**: 177–183.

de Vos L, van der Nest MA, van der Merwe NA, Myburg AA, Wingfield MJ, *et al.* (2011) Genetic analysis of growth, morphology and pathogenicity in the F_1 progeny of an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. *Fungal Biology* **115**: 902–908.

Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, *et al.* (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**: 980–986.

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.

Fourie A, Wingfield MJ, Wingfield BD, Barnes I (2015) Molecular markers delimit cryptic species in *Ceratocystis sensu stricto*. *Mycological Progress* **14**: 1–18.

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.

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, *et al.* (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**: 859–868.

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.

Geisen R, Schmidt-Heydt M, Karolewiez A (2006) A gene cluster of

the ochratoxin A biosynthetic genes in *Penicillium. Mycotoxin Research* **22**: 134–141.

Gryzenhout M, Myburg H, van der Merwe NA, Wingfield BD, Wingfield MJ (2004) *Chrysoporthe*, a new genus to accommodate *Cryphonectria cubensis*. *Studies in Mycology* **50**: 119–142.

Gryzenhout M, Myburg H, Wingfield BD, Montenegro F, Wingfield MJ (2005) *Chrysoporthe doradensis* sp. nov. pathogenic to *Eucalyptus* in Ecuador. *Fungal Diversity* **20**: 39–57.

Gryzenhout M, Rodas CA, Portales MJ, Clegg P, Wingfield BD, *et al.* (2006) Novel hosts of the Eucalyptus canker pathogen *Chrysoporthe cubensis* and a new *Chrysoporthe* species from Colombia. *Mycological Research* **110**: 833–845.

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

Hadjeba-Medjdoub K, Tozlovanu M, Pfohl-Leszkowicz A, Frenette C, Paugh RJ, *et al.* (2012) Structure-activity relationships imply different mechanisms of action for ochratoxin A-mediated cytotoxicity and genotoxicity. *Chemical Research in Toxicology* **25**: 181–190.

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.

Harrington TC (2009) The genus *Ceratocystis*. Where does the oak wilt fungus fit? In: *Proceedings of the 2nd National Oak Wilt Sym* (Appel DN, Billings RF, eds). Austin, TX: USDA Forest Service.

Harrington TC, McNew D, Steimel J, Hofstra D, Farrell R (2001) Phylogeny and taxonomy of the *Ophiostoma piceae* complex and the Dutch Elm disease fungi. *Mycologia* **93**: 111–136.

Heath RN, Gryzenhout M, Roux J, Wingfield MJ (2006) Discovery of the canker pathogen *Chrysoporthe austroafricana* on native *Syzygium* spp. in South Africa. *Plant Disease* **90**: 433–438.

Hodges C, Reis M, Ferreira F, Henfling J (1976) O cancro do eucalipto causado por *Diaporthe cubensis* [*Eucalyptus* spp.; Fungo; Brasil]. *Fitopatologia Brasileira* **1**: 129–170.

Hodges CS, Geary TF, Cordell CE (1979) The occurrence of *Diaporthe cubensis* on *Eucalyptus* in Florida, Hawaii, and Puerto Rico. *Plant Disease Reporter* **63**: 216–220.

Hoff KJ, Stanke M (2013) WebAUGUSTUS – a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acid Research* **41**: W123–W128.

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

Iturritxa E, Ganley RJ, Wright J, Heppe E, Steenkamp ET, *et al*. (2011) A genetically homogenous population of *Fusarium circinatum* causes pitch canker of *Pinus radiate* in the Basque Country, Spain. *Fungal Biology* **115**: 288–295.

Jacobs K, Seifert KA, Harrison KJ, Kirisits T (2003) Identity and phylogenetic relationships of ophiostomatoid fungi associated with invasive and native *Tetropium* species (*Coleoptera*: *Cerambycidae*) in Atlantic Canada. *Canadian Journal of Botany*. **81**: 316–329.

Jankowiak R, Bilański P (2013) Diversity of ophiostomatoid fungi associated with the large pine weevil, *Hylobius abietis*, and infested Scots pine seedlings in Poland. *Annals of Forest Science* **70**: 391–402.

Jurgenson JE, Zeller KA, Leslie JF (2002) Expanded genetic map of *Gibberella moniliformis* (*Fusarium verticillioides*). *Applied and Environmental Microbiology* **68**: 1972–1979.

Karolewiez A, Geisen R (2005) Cloning a part of the ochratoxin A biosynthetic gene cluster of *Penicillium nordicum* and characterization of the ochratoxin polyketide synthase gene. *Systematic and Applied Microbiology* **28**: 588–595.

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.

Kim G-H, Kim J-J, Breuil C (2007) Sap-staining fungi from logs and boards of two commercially important pines in Korea. *Holzforschung*. **61**: 333–336.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Sumway M, *et al*. (2004) Versatile and open software for comparing large genomes. *Genome Biology* **5**: R12.

Kvas M, Marasas WFO, Wingfield BD, Wingfield MJ, Steenkamp ET (2009) Diversity and evolution of *Fusarium* species in the *Gibberella fujikuroi* complex. *Fungal Diversity* **34**: 1–21.

Laia M, Alfenas A, Harrington T (1999) Isolation, detection in soil, and inoculation of *Ceratocystis fimbriata*, causal agent of wilting, die-back and canker in *Eucalyptus.* In: *Proceedings of the 12th biennial conference of the Australasian Plant Pathology Society*. eds): 77.

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

Lee DH, Roux J, Wingfield BD, Wingfield MJ (2015) Variation in growth rates and aggressiveness of naturally occurring self-fertile and self-sterile isolates of the wilt pathogen *Ceratocystis albifundus*. *Plant Pathology* **64**: 1103–1109.

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

Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J., *et al.* (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367–373.

Mathiesen-Käärik A (1953) Eine Übersicht über die gewöhnlichsten mit Borkenkäfer assoziierten Bläuepilze in Schweden und einige für Schweden neue Bläuepilze. *Meddelanden från Statens Skogsforskiningsinstitut*. **43**: 1–74.

Mayers CG, McNew DL, Harrington TC, Roeper RA, Fraedrich SW*, et al.* (2015) Three genera in the *Ceratocystidaceae* are the respective symbionts of three independent lineages of ambrosia beetles with large, complex mycangia. *Fungal Biology* doi:10.1016/jfunbio.2015.1008.1002.

Mitchell RS (1937) Stem end rot of bananas with special reference to the physiological relationships of *Thielaviopsis paradoxa* (de Seynes) von Hohn. *Journal of the Council for Scientific and Industrial Research* **10**: 123–130.

Myburg H, Gryzenhout M, Wingfield BD, Wingfield MJ (2002) β-Tubulin and histone H3 gene sequences distinguish *Cryphonectria cubensis* from South Africa, Asia, and South America. *Canadian Journal of Botany* **80**: 590–596.

Myburg H, Gryzenhout M, Wingfield BD, Wingfield MJ (2003) Conspecificity of *Endothia eugeniae* and *Cryphonectria cubensis*: a re-evaluation based on morphology and DNA sequence data. *Mycoscience* **44**: 187–196.

Nag Raj TR, Kendrick WB (1975) *A Monograph of Chalara and allied Genera*. Laurier W, Waterloo, ON: Waterloo University Press.

Nakabonge G, Roux J, Gryzenhout M, Wingfield MJ (2006). Distribution of *Chrysoporthe* canker pathogens on Eucalyptus

and *Syzygium* spp. in Eastern and Southern Africa. *Plant Disease Reporter* **90**: 734–740.

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

Oliveira L, Harrington TC, Ferreira MA, Damacena M, Al-Sadi AM*, et al.* (2015) Species or genotypes? Reassessment of four recently described species of the *Ceratocystis* wilt pathogen, *Ceratocystis fimbriata*, on *Mangifera indica*. *Phytopathology* **105**: 1229–1244.

Paciura D, Zhou X, de Beer ZW, Jacobs K, Ye H, *et al.* (2010) Characterisation of synnematous bark beetle-associated fungi from China, including *Graphium carbonarium* sp. nov. *Fungal Diversity* **40**: 75–88.

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

Pitt JI, Wild CP, Baan RA, Gelderblom WCA, Miller JD, *et al.* (2012) *Improving Public Health through Mycotoxin Control.* [Scientific Publications Series no. 158.] Lyon:. International Agency for Research on Cancer.

Raeder U, Broda P (1985) Rapid preparation of DNA from filamentous fungi. *Letters in Applied Microbiology* **1**: 17–20.

Riedl H (1962) *Ceratocystis musarum* sp. n., die heuptfruchtform der *Thielaviopsis*–Art von bananenstielen. *Sydowia* **15**: 247–251.

Rodas CA, Gryzenhout M, Myburg H, Wingfield BD, Wingfield MJ (2005) Discovery of the Eucalyptus canker pathogen *Chrysoporthe cubensis* on native *Miconia* (*Melastomataceae*) in Colombia. *Plant Pathology* **54**: 460–470.

Romon P, Zhou X, Iturrondobeitia J, Wingfield M, Goldarazena A (2007) *Ophiostoma* species (*Ascomycetes: Ophiostomatales*) associated with bark beetles (*Coleoptera*: *Scolytinae*) colonizing *Pinus radiata* in northern Spain. *Canadian Journal of Microbiology* **53**: 756–767.

Roux J, Coutinho T, Mujuni Byabashaija D, Wingfield MJ (2001) Diseases of plantation *Eucalyptus* in Uganda: research in action. *South African Journal of Science* **97**: 16–18.

Roux J, Heath RN, Labuschagne L, Nkuekam GK, Wingfield MJ (2007) Occurrence of the wattle wilt pathogen, *Ceratocystis albifundus* on native South African trees. *Forest Pathology* **37**: 292–302.

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.

Roux J, Wingfield M (2013) *Ceratocystis* species on the African continent, with particular reference to *C. albifundus*, an African species in the *C. fimbriata sensu lato* species complex. In: *The Ophiostomatoid Fungi: expanding frontiers* (Seifert KA, de Beer ZW, Wingfield MJ, eds): 131–138. [CBS Biodiversity Series vol. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

Roux J, Wingfield MJ (2009) *Ceratocystis* species: emerging pathogens of non-native plantation *Eucalyptus* and *Acacia* species. *Southern Forests* **71**: 115–120.

Roux J, Wingfield MJ, Bouillet JP, Wingfield BD, Alfenas AC (2000) A serious new wilt disease of *Eucalyptus* caused by *Ceratocystis fimbriata* in Central Africa. *Forest Pathology* **30**: 175–184.

Samson RA, Seifert KA, Kuijpers AFA, Houbraken J, Frisvad JC (2004) Phylogenetic analysis of *Penicillium* subgenus *Penicillium* using partial β-tubulin sequences. *Studies in Mycology* **49**: 175–200.

Scauflaire J, Gourgue M, Munaut F (2011) *Fusarium temperatum* sp. nov. from maize, an emergent species closely related to *Fusarium subglutinans*. *Mycologia* **103**: 586–597.

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 SJM, *et al.* (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research* **19**: 1117–1123.

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

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* **33** (Suppl 2): W465–W467.

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, Morgenstem B (2006) AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biology* **7**: S11.1-S.8.

Steenkamp ET, Wingfield BD, Coutinho TA, Wingfield MJ, Marasas WFO (1999). Differentiation of *Fusarium subglutinans* f. sp. *pini* by Histone Gene Sequence Data. *Applied and Environmental Microbiology* **65**: 3401–3406.

Teixeira MM, de Almeida LG, Kubitschek-Barreira P, Alves FL, Kioshima ÉS*, 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 HP (1981) *A Monograph of Ceratocystis and Ceratocystiopsis*. Athens, GA: University of Georgia Press.

van der Merwe NA, Gryzenhout M, Steenkamp ET, Wingfield BD, Wingfield MJ (2010) Multigene phylogenetic and population differentiation data confirm the existence of a cryptic species within *Chrysoporthe cubensis*. *Fungal Biology* **114**: 966–979.

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, 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.* (2014b) 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 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, Wingfield MJ (2013) *Ceratocystis* species in the *Ceratocystis fimbriata* complex. In: *The Ophiostomatoid Fungi: expanding frontiers* (Seifert KA, de Beer ZW, Wingfield MJ, eds): 65–73. [CBS Biodiversity Series vol. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

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

Visagie CM, Varga J, Houbraken J, Meijer M, Kocsube S, *et al.* (2014a) Ochratoxin production and taxonomy of the yellow aspergilli (*Aspergillus* section *Circumdati*). *Studies in Mycology* **78**: 1–61.

Wargo PM, Harrington TC (1991) Host stress and susceptibility. In: *Armillaria Root Disease* (Shaw CG Kile GA, eds): 88–101. [Agriculture Handbook no. 691.] Washington, DC: USDA Forest Service.

Wiemann P, Sieber CM, von Bargen KW, Studt L, Niehaus EM, *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 Pathogen* **9**: e1003475.

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.

Wingfield BD, Ades PK, Al-Naemi FA, Beirn LA, Bihon W*, et al.* (2015) 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, 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**: 1–2.

Wingfield BD, van Wyk M, Roos H, Wingfield MJ (2013) *Ceratocystis*: emerging evidence for discrete generic boundaries. In: *The Ophiostomatoid Fungi: expanding frontiers* (Seifert KA, de Beer ZW, Wingfield MJ, eds): 57–64. [CBS Biodiversity Series vol. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

Wingfield MJ (2003) Increasing threat of diseases to exotic plantation forests in the Southern Hemisphere: lessons from Cryphonectria canker. *Australasian Plant Pathology* **32**: 133–139.

Wingfield MJ, Swart WJ, Abear BJ (1989) First Record of Cryphonectria canker of Eucalyptus in South Africa. *Phytophylactica* **20**: 311–313.

Xu J-R, Yan K, Dickman MB, Leslie JF (1995) Electrophoretic karyotypes distinguish the biological species of *Gibberella fujikuroi* (*Fusarium* section *Liseola*). *Molecular Plant-Microbe Interactions* **8**: 74–84.

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

Zhou X, de Beer ZW, Wingfield MJ (2006) DNA sequence comparisons of *Ophiostoma* spp., including *Ophiostoma aurorae* sp. nov., associated with pine bark beetles in South Africa. *Studies in Mycology* **55**: 269–277.

ARTICLE