IMA Genome-F 4

Draft genome sequences of *Chrysoporthe austroafricana, Diplodia scrobiculata, Fusarium nygamai, Leptographium lundbergii, Limonomyces culmigenus, Stagonosporopsis tanaceti,* and *Thielaviopsis punctulata*

Brenda D. Wingfield[1], Peter K. Ades[2], Fatima A. Al-Naemi[3], Lisa A. Beirn[4], Wubetu Bihon[5,6], Jo Anne Crouch[7], Z. Wilhelm de Beer[5], Lieschen De Vos[1], Tuan A. Duong[1], Christopher J. Fields[8], Gerda Fourie[5], Aquillah M. Kanzi[1], Martha Malapi-Wight[7], Sarah J. Pethybridge[9], Osman Radwan[10], Gloria Rendon[8], Bernard Slippers[1], Quentin C. Santana[1], Emma T. Steenkamp[5], Paul W.J. Taylor[11], Niloofar Vaghefi[9], Nicolaas A. van der Merwe[1] , Daniel Veltri[7,12], and Michael J. Wingfield[5]

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

[2]Department of Forest and Ecosystem Science, The University of Melbourne, Victoria, 3010, Australia

[3]Department of Biological and Environmental Sciences, College of Arts and Sciences, Qatar University, Doha, Qatar

[4]Department of Plant Biology and Pathology, Rutgers University, 59 Dudley Road, New Brunswick, NJ 08901, USA

[5]Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, P. bag x20, Pretoria 0028, South Africa

[6]Agricultural Research Council, Vegetable and Ornamental Plant Institute, P. Bag X293, Pretoria 0001, South Africa

[7]Systematic Mycology and Microbiology Laboratory, U.S. Department of Agriculture (USDA), Agricultural Research Service (ARS), Beltsville, MD 20705, USA

[8]High Performance Biological Computing Group, Roy J. Carver Biotechnology Center/W.M. Keck Center, University of Illinois at Urbana-Champaign, IL, USA

[9]School of Integrative Plant Sciences, Plant Pathology & Plant-Microbe Biology Section, Cornell University, Geneva, NY, 14456, USA

[10]Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana-Champaign, IL, USA and Department of Plant Production, College of Technology, Zagazig University, Sharkia, Egypt

[11]Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Victoria, 3010, Australia

[12]Oak Ridge Laboratories ARS Research Participation Program, USDA-ARS, Beltsville, MD 20705, USA

Abstract: The genomes of *Chrysoporthe austroafricana, Diplodia scrobiculata, Fusarium nygami, Leptographium lundbergii, Limonomyces culmigenus, Stagonosporopsis tanaceti,* and *Thielaviopsis punctulata* are presented in this genome announcement. These seven genomes are from endophytes, plant pathogens and economically important fungal species. The genome sizes range from 26.6 Mb in the case of *Leptographium lundber*gii to 44 Mb for *Chrysoporthe austroafricana*. The availability of these genome data will provide opportunities to resolve longstanding questions regarding the taxonomy of species in these genera, and may contribute to our understanding of the lifestyles through comparative studies with closely related organisms.

IMA Genome-F 4A

Draft genome sequence of the pine pathogen *Diplodia scrobiculata*

Diplodia scrobiculata was originally considered to be a morphotype (B) of *D. sapinea* (syn. *Diplodia pinea*) until marker technology made it possible to recognise it as a discrete species (de Wet *et al*. 2003). The fungus is an endophyte and latent pathogen of *Pinus* species known especially from the USA and some countries of western Europe (Palmer 1987, Blodgett & Stanosz 1997, Stanosz *et al*. 1999, Burgess *et al*. 2004). It was recently reported from the Southern Hemisphere for the first time when it was found in South Africa (Bihon *et al*. 2011).

Diplodia scrobiculata is a weak pathogen that has been isolated from various *Pinus* species and some other conifers where it often co-exists with *D. sapinea* (de Wet *et al*. 2003). A population genetics study on this fungus has shown that isolates are genetically diverse where it is native, and that

there is strong geographic isolation between populations from different regions of North America (Burgess *et al.* 2004). With the exception of a few studies on the epidemiology and population structure, very little is known of the biology of *D. scrobiculata*.

Recently, the genome sequence of *D. sapinea* became available for public use (van der Nest *et al.* 2014a). Together with this and other sequences for species of *Botryosphaeriaceae* (Phillips *et al.* 2013) the availability of a genome sequence for *D. scrobiculata* would promote detailed biological studies and provide the basis for interesting genome comparisons including for example those considering pathogenicity factors.

SEQUENCED STRAIN

South Africa: Mpumanlanga: Sappi, isol. ex *Pinus patula*, Oct. 2008, *W. Bihon* (CMW 30223 = CBS139796; PREM 61246 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The Whole Genome Shotgun projects have been deposited at DBJ/EMBL/GenBank under the accessions LAEG00000000. The version described in this paper is version LAEG01000000.

METHODS

DNA from a single spore culture of *Diplodia scrobiculata* (CMW30223 = CBS139796) was extracted and sequenced using the Illumina: HiSeq genome analyser platform at Inqaba Biotechnology Industries, South Africa. Reads were subjected to the necessary sequence quality analysis and those of less than 30 bases and poor quality regions of reads were discarded. Remaining reads were assembled into a draft genome using CLC Genomic *de novo* assembler 7.0, with the minimum contig size set to 500 bases (CLCbio, Aarhus, Denmark). Gene predictions were done using AUGUSTUS (Stanke *et al.* 2006) based on the gene models for *Aspergillus oryzae* (http://bioinf.uni-greifswald. de/webaugustus/prediction). Completeness of the genome was evaluated using the Core Eukaryotic Genes Mapping Approach (CEGMA) analysis (Parra *et al.* 2007). Contigs of ≥ 500 bases were submitted to the genome database of NCBI.

RESULTS AND DISCUSSION

The draft nuclear sequence of *D. scrobiculata* isolate CMW30223 = CBS139796 resulted in an estimated genome size of 35.85 megabases (Mb). De novo assembly using CLC Genomics produced a total of 3789 contigs with N50 of 16 kilobases (Kb) and maximum contig size of 119715 bases. Mapping to the more conserved set of 248 Core Eukaryotic Genes (CEGs) using CEGMA pipeline (Parra *et al.* 2007), more than 95.2 % of the genome sequence was complete.

Putative gene prediction using AUGUSTUS (Stanke *et al.* 2006) predicted 13624 genes.

Diplodia scrobiculata has a genome size similar to that of *D. sapinea* (36.9Mb) and *Alternaria alternata* (33 Mb) (van Der Nest *et al.* 2014a, http://genome.jgi.doe.gov/). It has a similar number of genes to *A. alternata* (13469) but more than the most closely *D. sapinea* (13020). Although, *D. scrobiculata* is commonly found in association with *D. sapinea*, it is not clear why it is mainly confined to the Northern Hemisphere and that it has not moved globally as its sibling species has. It is also not known why it is considerably less pathogenic than *D. sapinea* (de Wet *et al.* 2002). Public access to the genome sequence of *D. scrobiculata* together with other related species will facilitate answering these questions as well as others, such as those relating to the association of endophytic latent pathogens and their tree hosts.

Authors: **W. Bihon, M.J. Wingfield, B. Slippers, and B.D. Wingfield**

Draft genome sequence of the *Eucalyptus* canker pathogen *Chrysoporthe austroafricana*

Species of *Chrysoporthe* (family *Cryphonectriaceae*) are important tree pathogens (Wingfield 2003), occasionally causing serious canker disease on *Myrtales* (Gryzenhout *et al.* 2004). *Chrysoporthe austroafricana* is mainly found in southern African countries, where it affects susceptible plantation-grown *Eucalyptus* species as well as native *Syzygium* species (Wingfield *et al.* 1989, Myburg *et al.* 2002, Roux *et al.* 2005, Heath *et al.* 2006, Nakabonge *et al.* 2006, Vermeulen *et al.* 2011). Symptoms on *Eucalyptus* often include cankers at the base and root collars of trees, causing wilt and subsequent death of young trees. Older trees show swollen basal cankers which girdle the stem and cause breakage during high winds.

The genus *Chrysoporthe* also includes *C. cubensis* (South and Central America) and *C. deuterocubensis* (south-east Asia) (Hodges *et al.* 1976, 1979, Myburg *et al.* 2002, Rodas *et al.* 2005, Gryzenhout *et al.* 2006, van der Merwe *et al.* 2010). Five other species of *Chrysoporthe* are currently recognized: *C. doradensis* (Ecuador) (Gryzenhout *et al.* 2005), *C. syzygiicola* and *C. zambiensis* (Zambia) (Chungu *et al.* 2010), and *C. inopina* and *C. hodgesiana* (Columbia) (Gryzenhout *et al.* 2004, 2006).

Previous studies on *Chrysoporthe* species have focused on their taxonomy (Gryzenhout *et al.* 2004), population genetics (Nakabonge *et al.* 2007, van der Merwe *et al.* 2010, Vermeulen *et al.* 2013), and geographical distribution (Myburg *et al.* 2002, Rodas *et al.* 2005, Heath *et al.* 2006, Nakabonge *et al.* 2006, Vermeulen *et al.* 2011). However, in order to further explore the genetics and evolution of these species in relation to their mating systems, population biology, phylogenetics and potential biotechnological applications, we have determined the whole genome sequence of *C. austroafricana* isolate

CMW2113. This genome resource will not only broaden the scope of research on the eight *Chrysoporthe* species, but also other fungi in the *Cryphonectriaceae*.

SEQUENCED STRAIN

South Africa: KwaMbonambi, 1989, *M.J. Wingfield* (culture CMW2113 = CBS 112916; PREM 58023 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The *Chrysoporthe austroafricana* isolate CMW2113 Whole Genome Shotgun project was deposited in GenBank with accession number JYIP00000000. The version described here is JYIP01000000.

METHODS

Genomic DNA was extracted from *C. austroafricana* isolate CMW2113. Genome sequencing was done at the Agricultural Research Council (ARC, South Africa) using the Illumina MiSeq paired-end protocol (Illumina, CA). Additionally, three independent genomic DNA samples were sequenced using the single-read IonTorrent protocol (Life Technologies Carlsbad, CA) at the University of Pretoria. These three batches consisted of two batches sequenced using a 100-base sequencing kit, and one batch using a 300-base kit. CLC Genomics Workbench v. 7.5.1 (CLCbio, Aarhus, Denmark) was used to assemble the paired-end MiSeq sequences after quality assessment. Additional scaffolding was performed with the unused MiSeq and IonTorrent data using SSPACE 2.0 (Boetzer et al. 2011). Protein coding gene models were annotated *de novo* using AUGUSTUS v. 3.0.3 with *Fusarium graminearum* as the model organism (Stanke & Morgenstern 2005). Lastly, the completeness of the genome was assessed using the occurrence of core eukaryotic orthologous genes, as implemented in the CEGMA v. 2.4 software (Parra *et al*. 2007).

RESULTS AND DISCUSSION

The assembled and scaffolded *C. austroafricana* genome was 44 669 169 bases in size, including gaps. The coverage calculated from 6 416 contigs was 40X, with an N50 value of 61 627 bases. Further, the GC content was estimated at 53.9 %. Analysis of the genome completeness using CEGMA showed that 94.35 % of 445 core orthologous genes were present. The AUGUSTUS gene prediction pipeline predicted a total of 13 484 protein coding gene models.

The total *C. austroafricana* genome size, predicted at 44.6 Mb is slightly larger than that of *Cryphonectria parasitica* (43.9Mb; 11 184 genes) (http://genome.jgi.doe.gov/Crypa2/Crypa2.info.html), which is the only other member of *Cryphonectriaceae* for which genome information is available. The evolutionary significance of these differences is yet to be determined, including whether these variations played a role in the diversification of the species. The *C. austroafricana* genome size and number of predicted protein coding genes were also compared to model ascomycetes such as *Neurospora crassa* (39.9 Mb, 10 082 genes; Galagan *et al*. 2003), *Podospora anserina* (36Mb, 10 545 genes) (Espagnel *et al*. 2008) and *Magnaporthe grisea*, (40.3Mb, 11 109 genes; Dean *et al*. 2005), all of which were smaller with fewer predicted protein coding genes.

The continued availability of fungal genomes provides an opportunity to broaden the scope of research in fungal biology, taxonomy, and evolution. This is especially true for species of *Chrysoporthe*, since the lack of genomic resources has limited the scope of previous studies. The *C. austroafricana* WGS is an invaluable resource that will certainly advance our knowledge of the biology of *Chrysoporthe* species.

Authors: **A.M. Kanzi, B.D. Wingfield, M.J. Wingfield, E.T. Steenkamp, and N.A. van der Merwe**

IMA Genome-F 4C

Draft genome sequence of *Fusarium nygamai*

The *Fusarium fujikuroi* species complex (FFSC, previously referred to as the *Gibberella fujikuroi* species complex) includes a range of important pathogens as well as species of human and animal health concern because of their ability to produce highly toxic compounds (Leslie & Summerell 2006, Kvas *et al.* 2009). To allow genomic comparisons of these important fungi, the genomes of a growing number of FFSC species are available, e.g. *F. verticillioides, F. circinatum, F. fujikuroi,* and *F. mangiferae* (Fusarium Comparative Sequencing Project, Wingfield *et al.* 2012, Wiemann *et al.* 2013). These fungi are, respectively, pathogens of maize, pine, rice, and mango and are by no means representative of the FFSC diversity known to be associated with cultivated plants (Leslie & Summerell 2006, Kvas *et al.* 2009). This is true, not only in terms of plant host, but also other biological properties (e.g. morphology, mycotoxicology) typically linked with this complex. For example, none of the four species for which whole genome sequences are capable of forming chlamydospores (Kvas *et al.* 2009). In fact, the ability to produce these thick-walled vegetative structures is not a common FFSC feature and only a few species have been shown to have this capability (Leslie & Summerell 2006, Kvas *et al.* 2009).

The chlamydospore formers in the FFSC include *F. dlamini*, *F. napiforme*, *F. nygami*, *F. acutatum*, *F. pseudoanthophilum*, *F. udum,* and *F. xylarioides* (O'Donnell *et al.* 1998, 2000).

Among these, the important wilt pathogens, *F. udum*, *F. xylarioides* (Booth 1971), and *F. nagamai* have probably been most extensively studied. The aim of this study was to sequence the whole genome of *F. nagamai.* The species represents one of the so-called mating populations or

biological species of the complex (Klaasen & Nelson 1996) and is often associated with sorghum and other field crops. The species may be dominant in native grassland soils, especially in hot and arid locations (Sangalang *et al.* 1995). *Fusarium nygami* has also been demonstrated to be the causal agent of basal stalk and root rot of grain sorghum (Trimboli & Burgess 1983), while certain isolates have been reported to cause disease on broad beans, asparagus, cotton, maize, millet, and rice (Leslie & Summerell 2006). Overall, the genome sequence of this species would allow for more extensive genomic comparisons across the FFSC and facilitate in depth explorations of the genetic basis for the diverse biological characters found in this complex.

SEQUENCED STRAIN

USA: *Pennsylvania*: laboratory strain, progeny of a sexual cross between two *F. nygamai* strains (M-1564 x M-2370) isolated from soil in South Africa and soil in Australia. MAT-1 mating type tester strain for mating population G (Klaasen & Nelson 1996), June 1996, *J.A. Klaasen* & *P.E. Nelson* (MRC8546, CMW12847, KSU G-05111, CMWF1209, M-7492, K-148, PREM – 61277 dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The *Fusarium nygamai* genome sequence data has been deposited at DDBJ/EMBL/GenBank under the accession number LBNR 0000 0000. The first version is described in this paper.

METHODS

Genomic DNA was isolated from 1-wk-old single-spore cultures incubated on half strength potato dextrose agar (PDA) (Biolab Diagnostics, Wadeville, South Africa) at 25 °C as described previously (Groenewald *et al*. 2006). One mate-pair and two pair-end libraries with insert sizes of 3000 bp, 250–350bp and 560–650 bp, respectively, were prepared and subjected to Illumina HiSeq 2000 sequencing at Fasteris (Genèva, Switzerland). After exclusion of poor quality reads using CLC Genomics Workbench v. 8.0 (CLCbio, Aarhus, Denmark), sequences were assembled into scaffolds using ABySS v. 1.5.2 (Simpson *et al.* 2009) and gapped regions within scaffolds were closed using GapFiller v1.11 (Boetzer & Pirovano 2012). The completeness of the genome assembly was evaluated using CEGMA v. 2.5 (Parra *et al.* 2007) and BUSCO using conserved orthologous gene sets specific to fungi (Simão *et al*. 2015). Putative open reading frames (ORFs) were predicted using WebAUGUSTUS (Hoff & Stanke 2013) together with the gene models for *F. graminearum* (Fusarium Comparative Sequencing Project http//www.broad.mit.edu) and *F. circinatum* (Wingfield *et al.* 2012).

RESULTS AND DISCUSSION

Assembly of the draft genome of *Fusarium nygamai* yielded a genome size of 51 615 029 bp with an average coverage of 333x. The genome was assembled into 409 scaffolds ranging in size from 200 bp to 5 319 023 bp with an N50 scaffold length of 2 761 656 bp and an average GC content of 47.46 %. Of the 409 contigs, 15 scaffolds were larger than 1 000 000 bases in size and represented approximately 78 % of the complete genome sequence. Based on the occurrence of a set of genes shared by all eukaryotes (Parra *et al.* 2007) the *F. nygamai* genome is 96.37 % complete and 99 % complete based on the conserved fungal gene sets. The genome of *F. nygamai* is predicted to contain 15 780 putative ORFs.

The genome of *F. nygamai* is larger than those of *F. verticillioides* (41.8 Mb), *F. circinatum* (44.3 Mb), *F. fujikuroi* (43.9 Mb), and *F. mangiferae* (45.6 Mb) (Wiemann *et al.* 2013). The genome size, however, is similar to that of *F. solani* (51.3 Mb) (Wiemann *et al.* 2013) with the predicted number of ORFs in agreement to not only *F. solani* but also the other smaller genomes of species in the FFSC (Wiemann *et al.* 2013). Preliminary BLAST analysis also suggests that the *F. nygamai* isolates sequenced in this study contains twelve chromosomes. The twelfth chromosome appears to be dispensable in some members of the genus *Fusarium* and was found to be absent from a laboratory strain of *F. circinatum* (van der Nest *et al.* 2014b). The availability of this genome, should therefore not only assist in future phylogenomic studies of the FFSC species which would allow for a better understanding of the overall evolution of this complex, but could also assist in a better understanding of the role and importance of dispensable chromosomes in the FFSC.

Authors: **Q.S. Santana, G. Fourie, L. de Vos, N.A. van der Merwe, M.J. Wingfield, B.D. Wingfield, and E.T. Steenkamp**

IMA Genome-F 4D

Draft genome sequence of *Leptographium lundbergii*

Leptographium lundbergii, described in 1927 as a cause of blue stain in spruce in Sweden (Lagerberg *et al*. 1927), is the type species of *Leptographium s. lat*. This is an ascomycete genus residing in *Ophiostomatales* that currently accommodates more than 100 species (de Beer & Wingfield 2013). Most species of *Leptographium s. lat*. are associates of bark and root-infesting beetles found on conifers and are best known for their ability to cause sap stain in recently killed trees (Jacobs & Wingfield 2001). Species of *Leptographium s. lat*. are mostly considered to be saprobes, but a small number are important tree pathogens (Harrington 1993, Jacobs & Wingfield 2001). Phylogenetically, the pathogens are closely related to those species that have a saprobic lifestyle (Linnakoski *et al*. 2012, de Beer & Wingfield 2013).

ARTICLE

The genomes of some *Leptographium s.lat.* associated with tree disease have been sequenced and characterized. These include those of *Grosmannia clavigera* and *Leptographium procerum* (DiGuistini *et al.* 2011, van der Nest *et al.* 2014). The availability of a genome sequence of a saprobic species such as *L. lundbergii* would make possible opportunities for genome-level comparisons between pathogenic and non-pathogenic species, and possibly provide insights into how pathogenic species cause disease. The aim of this study was thus to generate genome data of *L. lundbergii*, a saprobe and importantly the type species of *Leptographium s. lat.*

SEQUENCED STRAIN

Norway: *Pinus sylvestris*, 1989, *H. Roll-Hansen* (culture CMW 2190 = CBS 138716; PREM 61278 – dried culture)

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The genomic sequence of *Leptographium lundbergii* (CMW 2190 = CBS 138716) has been deposited at DDBJ/EMBL/GenBank under the accession LDEF00000000. The version described in this paper is version LDEF01000000.

METHODS

Genomic DNA was extracted from a single conidium culture of *L. lundbergii* CMW 2190 (CBS 138716) following the method described by Duong *et al*. (2013). Genomic DNA was submitted for Illumina Hiseq sequencing at the Genome Centre, University of California at Davis (CA). Two insert libraries (350 bp and 530 bp average insert size) were prepared and 100 bp pair-end sequencing was carried out on the Illumina HiSeq 2000 platform. Reads with low quality were filtered out and 15 bases of sequence were trimmed from 5' ends of all reads. De-novo assembly was carried out with reads after filtering and trimming using the CLC genomic work bench v. 8.0.1 (CLCBio, Aarhus, Denmark). The quality and completeness of the assembled genome was estimated by calculating CEGMA score (Parra *et al.* 2007). The protein coding genes were predicted using the MAKER genome annotation and curation pipeline (Cantarel *et al*. 2008). Firstly, RepeatScout (Price *et al*. 2005) was used to search for repetitive regions of the genome. These regions were then masked from the genome using RepeatMasker (Smit *et al*. 2013–15). For protein coding gene prediction, SNAP (Korf 2004), Augustus (Stanke & Waack 2003) and GeneMark (Lukashin & Borodovsky 1998) gene prediction programmes were used. Two rounds of self-training were conducted for SNAP using proteins from *Grossmania clavigera* (DiGuistini *et al*. 2011), *Ophiostoma ulmi* (Khoshraftar *et al*. 2013), *O. novo-ulmi* (Comeau *et al*. 2015), *O. piceae* (Haridas *et al.* 2013), *Magnaporthe grisea* (Dean *et al.* 2005), and *Neurospora crassa* (Galagan *et al*. 2003) as protein homology evidence. The species gene model of *M. grisea* was used for Augustus

prediction and the HMM file generated with a separated run was used for GeneMark. InterProScan (Zdobnov & Apweiler 2001) was used to screen the predicted proteins for the presence of PfamA domains to use as additional support for resulting gene models. The completeness of the genome was estimated using CEGMA (Parra *et al.* 2007, 2009) and BUSCO (Simão *et al*. 2015). The fungal mating type locus of the sequenced isolate was identified and characterized by using Blast against known mating type genes from closely related species (Duong *et al*. 2013, Tsui *et al*. 2013).

RESULTS AND DISCUSSION

About 22 million reads with an average read length of 84.5 bp were obtained after quality filtering and trimming. De-novo assembly using the CLC Genomics Workbench resulted in 412 scaffolds with over 500 bp in size. The N50 value was 212 Kb and the longest scaffold was just over 1 Mb. The coverage across the whole assembly was around 70 times. The estimated genome size of *L. lundbergii* is 26.6 Mb, smaller than the closely related species, *L. procerum* (28.6 Mb) and *G. clavigera* (29.8 Mb) (DiGuistini *et al*. 2011, van der Nest *et al*. 2014a). The mean GC content of the *L. lundbergii* genome was 56.8 %, which is a slightly higher than that of *G. clavigera* (53.4 %), and *L. procerum* (54.77 %). The assembled *L. lundbergii* genome had a CEGMA completeness score of 95.97 % when calculated from the complete gene set and 97.18 % when calculated from both partial and complete gene sets. Assessment of the completeness of genome using 1438 BUSCO groups for fungi resulted in BUSCO values of C:97%[D:5.5%],F:2.0%,M:0.4% ,n:1438 (C:complete [D:duplicated], F:fragmented, M:missed, n:genes). These results from CEGMA and BUSCO indicated that the assembled genome covered most of the organism's gene space. MAKER predicted a total of 9547 protein coding genes (MAKER's max build), 8035 of which were supported by protein and/or PfamA domain evidence (MAKER's standard build) (Campbell *et al.* 2014). *Leptographium lundbergii* had a gene density of 1/2.78 kb and the mean GC of the coding regions was 54 % (minimum of 27.5 % and maximum of 70.2 %). Of all suggested gene models from MAKER, 2070 were single exon (21.7 %) and 7477 gene models were multi-exon (78.3 %). The median exon length was 544 bp and the median intron length was 126 bp.

The sequenced *L. lundbergii* isolate harboured the *MAT1-2* idiomorph with a typical *MAT1-2-1* gene and a truncated version of *MAT1-1-1* gene. This is similar to the *MAT1-2* idiomorphs characterized in a number of closely related species of *Leptographium s. lat.* (Duong *et al*. 2013, Tsui *et al*. 2013). The draft genome of *L. lundbergii* represents a useful resource that will enable future comparative genomic studies to be conducted between saprophytic and pathogenic species of *Leptographium* and related fungi. As the type species of the genus, the genome sequence of *L. lundbergii* will also be valuable for future systematic studies on the relationships between members of this genus and other species in the *Ophiostomatales*.

Authors: **T.A. Duong, M.J. Wingfield, Z.W. de Beer, and B.D. Wingfield**

Draft genome sequence of *Stagonosporopsis tanaceti*

The genus *Stagonosporopsis* from *Didymellaceae* (de Gruyter *et al.* 2009) includes several economically important plant pathogens such as *S. cucurbitacearum* (syn. *Didymella bryoniae*), the cause of gummy blight of cucurbits, and *S. andigena* and *S. crystalliniformis*; destructive quarantine pathogens of potato and tomato, respectively (Aveskamp *et al.* 2010, de Gruyter *et al.* 2012). *Stagonosporopsis tanaceti* (formerly known as *Phoma ligulicola* var. *inoxydabilis*) is the cause of ray blight, the most damaging disease of pyrethrum (*Tanacetum cinerariifolium*) in Australia (Pethybridge & Wilson 1998, Vaghefi *et al.* 2012). The fungus was first detected in the Australian pyrethrum fields in 1995, and soon became a major threat, causing devastating epidemics in 1999, which led to losses of to 100 % (Pethybridge *et al.* 2008). The origin and global distribution of *S. tanaceti* is unknown but it is believed to have close evolutionary affinity to *S. chrysanthemi* and *S. inoxydabilis*, which infect various *Asteraceae* in Europe and the USA (Stevens 1907, van der Aa *et al.* 1990).

The homothallic *S. chrysanthemi* and *S. inoxydabilis* have well-described sexual morphs in *Didymella* (de Gruyter *et al.* 2002). However, no sexual morph has been detected in *S. tanaceti*, and it is believed to reproduce mainly asexually in pyrethrum fields based on the detection of only one mating-type gene in the population as well as significant linkage disequilibrium of neutral genetic markers (Pethybridge *et al.* 2012, Vaghefi *et al.* 2015a, b). It causes deformation and dieback of the leaves, stem lesions, and flower blight, which can result in flower death and complete yield loss if not controlled with multiple fungicide applications (Pethybridge *et al.* 2008).

Stagonosporopsis tanaceti is also known to live as an endophyte on symptomless plant tissue, becoming pathogenic when environmental conditions are conducive to disease development (unpublished data). However, the infection process and pathogenicity mechanisms of *S. tanaceti* are not yet fully understood. Many plant pathogenic *Dothideomycetes* are known to produce toxins that contribute to host-specific or non-specific pathogenicity (Daub & Ehrenshaft 2000, Friesen *et al.* 2008). Indeed, *S. chrysanthemi* produces phytotoxic compounds upon infection of chrysanthemum (Schadler & Bateman 1974, 1975) but these compounds have not been described in the *S. tanaceti*-pyrethrum pathosystem. Availability of genomic data will provide a powerful tool for understanding the mechanisms of pathogenicity and infection process of *S. tanaceti* through studying the genes underlying such mechanisms.

Moreover, genome sequence data provides a valuable resource for detection of the genes responsible for sensitivity to fungicides. Disease management strategies for ray blight in Australian pyrethrum fields are heavily reliant upon fungicides (Pethybridge *et al.* 2005, 2008). However, the use of fungicides with a high risk of resistance development (e.g. strobilurins and succinate dehydrogenase inhibitors; FRAC 2013) is a major concern to the durability of disease management. Identifying genomic regions associated with fungicide sensitivity will enable rapid screening of *S. tanaceti* populations for characterizing temporal changes in resistant allele frequencies in response to specific selection pressures. A *de novo* genome assembly of *S. tanaceti* is presented and made publically available to facilitate understanding the genetic mechanisms underlying the biology and evolution of this important plant pathogen.

SEQUENCED STRAINS

Australia: *Tasmania*: northern Tasmania, Scottsdale, from *Tanacetum cinerariifolium* (pyrethrum), 2004, *S.J. Pethybridge* (CBS H-20947 – holotype; CBS 131484 – culture ex-holotype); from *T. cinerariifolium*, 2004, *S.J. Pethybridge* (CBS 131485); from *T. cinerariifolium*, 2004, *S.J. Pethybridge* (TAS 55503); from *T. cinerariifolium*, 1998, *S.J. Pethybridge* (DAR 70020).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession JUDZ00000000. The version described in this paper is version JUDZ01000000.

METHODS

The genome of *Stagonosporopsis tanaceti* (a pooled DNA sample of four isolates) was sequenced through the Illumina platform (100 bp paired-end reads) at the Savannah River Ecology Laboratory (SREL, University of Georgia, GA). This yielded 29 395 716 paired reads, totalling 5.8 Gb. Trimmomatic v.0.22 (http://www.usadellab.org/cms/?page=trimmomatic) was used for quality trimming with a 4-base long sliding window, a minimum quality value of 30, and minimum length of 30 bases. A total of 21 553 834 paired reads (73.32 %) and 5 519 748 single reads (18.78 %) were retained for assembly after quality trimming (7.9 % discarded). Partial genome assembly was conducted using SOAPdenovo v. 1.05 (Li *et al.* 2008) with the kmer length parameter tested between 21 and 51 in 2 bases increments. The optimal kmer resulting in the maximal N50 length occurred at k = 35. The completeness of the genome was assessed using the Core Eukaryotic Genes Mapping Approach (CEGMA v. 1.0) (Parra *et al.* 2007). Gene prediction was conducted in Maker v. 2.31.7 (Cantarel *et al.* 2008); a genome annotation pipeline that combine *ab initio* gene prediction and evidence-based annotations, suitable for novel genomes with no training data available. Only contigs with a length of ≥ 500 bp were used for annotation. A preliminary annotation utilized *ab initio* gene prediction programs SNAP (Korf 2004) and Augustus v. 2.5.5 (Stanke *et al.* 2006) as well as evidence-based gene prediction

using the ESTs and protein sequences of *Didymella exigua* genome (http://genome.jgi-psf.org/Didex1/Didex1.home.html). The resulting annotation was used to produce a hidden-markov-model (HMM) profile for *S. tanaceti*, which was further refined with a second stage of SNAP training. The refined HMM file was used for the final annotation (Cantarel *et al.* 2008).

RESULTS AND DISCUSSION

The draft genome had a total assembly size of 40.8 Mbp with a genome coverage of at least 100×, N50 value of 98 157 bp and maximum contig size of 112 607 bp. All contigs with a length of ≥ 200 bp (9 834 contigs) were submitted to the genome database of NCBI. Mapping to the conserved set of 248 Core Eukaryotic Genes (CEGs) using CEGMA (Parra *et al.* 2007) estimated the genome to be at least 94 % complete, as the analysis revealed hits for 234 CEGs (94.35 %) with complete match and 244 with partial match (98.39 %). *Ab initio* gene prediction using trained SNAP identified 17 219 Open Reading Frames (ORFs).

The estimated genome size of *S. tanaceti* (40.8 Mb) is similar to the genome size of multiple *Dothideomycetes*, including *Pyrenophora teres* (41.95 Mb), *Zymoseptoria tritici* (39.7 Mb), and *Stagonospora nodorum* (37.2 Mb) (Hane *et al.* 2007, Ellwood *et al.* 2010, Goodwin *et al.* 2011). This was much larger than the genome of *Didymella exigua*, which has a predicted size of 34.39 Mb (http://genome.jgi-psf.org/Didex1/Didex1.info.html).

The draft genome of *S. tanaceti*, has been used for the detection of the mating-type locus (Vaghefi *et al.* 2015a) and development of a microsatellite library (Vaghefi *et al.* 2015b). To the best of our knowledge, this is the first reported draft genome of the genus *Stagonosporopsis* and provides a useful genomic resource for further genome-enabled studies into plant-microbe interactions.

Authors: **N. Vaghefi, P.K. Ades, S.J. Pethybridge, and P.W.J. Taylor**

IMA Genome-F 4F

Draft genome sequence of the fungus *Limonomyces culmigenus*

Basidiomycete fungi in the order *Corticiales* are an ecologically diverse group of organisms that derive nutrients through pathogenic associations with plants, lichenicolous symbioses, or saprophytically from dead organic matter (Lawrey *et al.* 2008). Two plant-associated genera in the *Corticiales, Limonomyces* and *Laetisaria,* are particularly well known for their vivid red and pink coloration (Stalpers & Loerakker 1982). Morphology is the primary method used to discriminate between members of *Limonomyces* and *Laetisaria* (Stalpers & Loerakker 1982). However, based on phylogenetic analyses of nuclear and mitochondrial

ribosomal DNA, both *Laetisaria* and *Limonomyces* appear to be polyphyletic, and are inadequately described from morphological data alone (Diederich *et al.* 2011).

Members of the genus *Limonomyces* are associated with grasses in several genera, including *Agrostis*, *Cynodon*, *Festuca*, *Lolium,* and *Poa* (Smiley *et al.* 2005). The genus includes two species: *L. roseipellis* and *L. culmigenus* (Stalpers & Loerakker 1982). *Limonomyces roseipellis* and *L. culmigenus* closely resemble one another, except *L. culmigenus* develops larger basidiospores and basidia, which have two sterigmata (Stalpers & Loerakker 1982). In natural environments, *L. roseipellis* is often found in association with *Laetisaria fuciformis* (Smiley *et al.* 2005). Both *L. roseipellis* and *Laetisaria fuciformis* are weakly pathogenic to grasses under conducive environmental conditions, causing diseases known as pink patch and red thread, respectively (Smiley *et al.* 2005). In managed turf grasses, red thread and pink patch diseases are unsightly, therefore turf grass breeders actively work to incorporate resistance to these pathogens in their cultivars (Bonos *et al.* 2006).

Unlike *L. roseipellis* and *Laetisaria fuciformis*, the lifestyle of *Limonomyces culmigenus* is poorly known, with just a few living samples of the fungus collected from grasses in England and Canada during the past five decades. The fungus has been described as a parasite of *Dactylis glomerata*, primarily growing within the host tissue and producing narrow pink sporophores (Stalpers & Loerakker 1982). References to *L. culmigenus* as a causal agent of turfgrass pink patch are sometimes found, but no credible evidence demonstrating pathogenicity or disease is available (Farr & Rossman 2015). To enable research on this enigmatic fungus, we generated a draft genome assembly from *L. culmigenus* CBS 661.85. This draft assembly provides the first genomic resource for a member of the family *Corticiaceae*.

SEQUENCED STRAIN

Canada: *British Columbia*: isolated from dead grass culm, 1985, *J.R. Bandoni* (CBS 661.85; BPI 893193 – dried culture).

NUCLEOTIDE ACCESSION NUMBERS

The Whole Genome Shotgun project of the CBS 661.85 genome has been deposited in NCBI GenBank under the accession LCTX00000000, with version LCTX00000000 described in this paper. Nucleotide sequences for homeodomain (HD1, HD2), pheromone receptor (*STE3*), and mitochondrial intermediate peptidase precursor (*MIP*) genes are deposited in NCBI GenBank under accession numbers KR779915−KR779922.

METHODS

Genomic DNA was extracted and purified as previously described (Malapi-Wight *et al.* 2015) from a natural isolate of *Limonomyces culmigenus*, CBS 661.85. An Illumina TruSeq PCR-Free DNA sequencing library (Illumina, San Diego, CA)

was constructed and quantified using the Qubit 2.0 fluorometer (Life Technologies, Grand Island, NY) and LabChipXT DNA 750 (Caliper Life Sciences, Hopkinton, MA), then sequenced on an Illumina MiSeq instrument using a 600-cycle sequencing cartridge (Illumina, Inc). Reads were processed and assembled using the CLC Genomics Workbench v. 7.6.1 software (CLC Bio, Boston, MA). After adapter removal, reads were trimmed to remove low quality sequences (Phred quality score <0.05) and reads shorter than 15 or longer than 600 bases. A *de novo* genome sequence was assembled based on de Bruijn graphs using a kmer=61, bubble size of 50, with contigs >500 bp retained. Assembly summary statistics were calculated using CLC Genomics Workbench, PRINSEQ v. 0.20.4 (Schmieder and Edwards 2011) and QUAST β (Gurevich *et al.* 2013). The number of core eukaryotic genes present in the assembly was assessed using CEGMA (Parra *et al.* 2007) through the iPLANT Discovery Environment interface (https://de.iplantcollaborative.org/de/) and gene predictions were obtained using AUGUSTUS v. 3.0.2 with *Coprinopsis cinerea* gene models (Stanke *et al.* 2008). Identification of protein ortholog families was performed using OrthoMCL (Li *et al.* 2003). Read depth of coverage and variant detection relative to the genome assembly were generated in CLC Genomics Workbench, with variant calls restricted to regions with a minimum of 10x read coverage and >90% probability, modeled as a diploid. Scaffolds were analyzed for large-scale similarities with other scaffolds in the assembly by performing a reciprocal BLASTn analysis of scaffolds >1000 bp, then determining the average depth of reads across: (a) 100 pairs of scaffolds where a matching scaffold was identified from the assembly; and (b) 100 individual scaffolds without a matching scaffold in the assembly. Putative secreted proteins lacking GPI-anchor sites and transmembrane domains were identified using SECRETOOL (Cortazar *et al.* 2014). Putative secondary metabolite clusters were identified using antiSMASH (Blin *et al.* 2013), and CAZymes were predicted using dbCAN (Yin *et al.* 2012) from the assemblies of *L. culmigenus* and three additional *Agaricomycotina* species downloaded from the Joint Genome Institute Mycocosm Genome Portal (http://genome.jgi. doe.gov/programs/fungi/index.jsf) for comparative purposes. The mitochondrial genome was identified by performing a tBLASTn search of the *L. culmigenus* CBS 661.85 assembly with the complete mitochondrial genome of *Rhizoctonia solani* AG8 (Scaffold 77; 140 kb; Hane *et al.* 2014). Genes were identified from the mitochondrial genome using MITOS (Bernt *et al.* 2013) under genetic code four. Transfer RNA was identified from the CBS 661.85 mitochondrial genome using tRNAscan-SE v. 1.21 (Lowe & Eddy 1997), with the source designated as mitochondrial and the genetic code as yeast mitochondrial (Schattner *et al.* 2005). The mitochondrial assembly was compared with *R. solani* Rhs1AP using the promer, showcoords, mapview, and mummerplot functions in MUMmer v. 3.23 (Kurtz *et al.* 2004). All BLAST searches in this work were performed with an E-value threshold of 1.0E-5.

RESULTS AND DISCUSSION

The nuclear genome assembly of *Limonomyces culmigenus* CBS 661.85 is 36.1 Mb (30 843 264 paired-end reads; N_{50}=9

365, L_{50}=706). The average read depth of coverage is 211-fold, with a GC content of 57.8 %. The assembly is organized into 8 889 scaffolds of 500 bp or greater in length, with the largest scaffold measuring 250 757 bp (av.= 4 065 bp). The assembly size of CBS 661.85 is consistent with the genomes sequences of other published *Agaricomycetes*, such as *Schizophyllum commune* (38.5 Mb; Ohm *et al.* 2010), *C. cinerea* (37 Mb; Stajich *et al.* 2010) and *Phanerochaete chrysosporium* (35.1 Mb; Martinez *et al.* 2004). The nuclear assembly is approximately 91.5 % complete, based on the presence of a portion of the ultra-conserved eukaryotic gene set (Parra *et al.* 2007). The high depth of read coverage and relatively fragmented assembly suggest that CBS 661.85 has a repetitive genome.

Although a comprehensive analysis of the genome characteristics of *L. culmigenus* CBS 661.85 is beyond the scope of the present report, the predicted dikaryotic nature of this fungus indicates that the 36.1 Mb assembly is likely an overestimate of the haploid genome size. Three features of the CBS 661.85 assembly are consistent with a dikaryotic organism, as proposed for *L. culmigenus* based on morphological observations (Stalpers & Loerakker 1982). Variant detection assessment identified 935 heterozygous variants (92.6 % single nucleotide polymorphisms; SNPs) across 674 scaffolds (7.6 % of the total scaffolds), suggesting a high level of similarity between the two nuclei in these regions. All but eight of the identified variants are present at a 1:1 ratio relative to the assembly, consistent with the presence of two nuclei. We also observed an almost two-fold difference in the average depth of read coverage between scaffolds identified as unique in the assembly (average read coverage 173.4x), *versus* the read coverage calculated from scaffolds where a matching scaffold is present in the assembly (average read coverage 83.4x). Duplicated copies of the genes associated with the mating type locus also support the presence of two unique nuclei in the assembly. The A mating-type locus is represented by two scaffolds containing pairs of divergently transcribed genes containing homeodomain transcription factors (HD1 and HD2) alongside the mitochondrial intermediate peptidase precursor (*MIP*) gene on scaffolds 1965 and 2914. Pairwise nucleotide identities for the HD genes are 83.5 % and 67.6 % (HD1 and HD2, respectively), and the two *MIP* genes share 95.8 % identity. Two highly divergent copies of the *STE3* pheromone receptor from the B mating type locus are present in the genome (scaffolds 5234 and joined_contig_68), with only 38.0% shared identity at the nucleotide level. The A and B mating type loci are located on different scaffolds, but due to the relatively small size of these scaffolds, we could not determine if the separation is due to fragmentation of the assembly, or if the loci occupy different locations in the genome. Therefore, it is still unknown whether *L. culmigenus* CBS 661.85 employs a bipolar or tetrapolar mating strategy.

A total of 16 394 putative coding regions were identified from the *L. culmigenus* CBS 661.85 genome assembly, with an average gene length of 1139 bp and a mean gene density of 454 genes/Mb. Proteins in the predicted proteome ranges from 16−6 262 amino acids in length (av.= 379). Comparison with available basidiomycete fungal genomes shows a general agreement between the predicted gene numbers

Table 1. Summary of the draft genome sequence of *Limonomyces culmigenus* (CBS 661.85) and published genome sequences for three basidiomycete fungi (Stajich *et al.* 2010, Floudas *et al.* 2012, Wibberg *et al.* 2013).

Species	*Limonomyces culmigenus* CBS 661.85	*Rhizoctonia solani* AG-1 1B	*Tremella mesenterica* Fries v1.0	*Coprinopsis cinerea* 7#130
Summary Data				
Total sequence length (Mb)	36.1	47.7	28.6	37.5
Number of scaffolds	8889	1600	45	13
GC (%)	57.8	48.1	45.7	51.7
Predicted gene models	16394	15157	8313	13393
Predicted secreted proteins	1027	876	213	776
Predicted CAZymes				
Total CAZymes	809	944	282	661
Auxiliary activity	86	132	95	105
Carbohydrate binding	125	130	12	122
Carbohydrate esterase	128	141	9	115
Glycoside hydrolase	355	348	78	212
Glycosyltransferase	106	101	84	89
Pectate lyase	9	92	4	18
Predicted Secondary Metabolite Clusters				
Total SM Clusters	23	13	4	14
Type I polyketide synthetases (PKSs)	3	0	0	1
Type III PKSs	2	0	0	0
Nonribosomal peptide synthetases	1	2	0	1
Terpene clusters	8	8	3	9
Hglks	0	0	0	0

of *L. culmigenus* and other *Agaricomycetes;* i.e. *Laccaria bicolor, Coprinopsis cinerea* and *Schizophyllun commune* with 20 614, 13 342, and 13 210 predicted coding regions respectively (Stajich *et al.* 2010). Ortholog families were identified for 75.6 % of the *L. culmigenus* proteome.

Putative effector genes that function in the host infection process are normally identified from genome datasets as proteins that contain secretion signals, are small and cysteine rich, and are lineage specific. There are 1027 predicted secreted proteins in the *L. culmigenus* CBS 661.85 genome assembly, making up 8 % of the predicted proteome. Predicted amino acid sequence lengths within the secretome range from 36−1833 aa (av. = 300.5 aa). Within the secretome, 601 proteins are classified as small (SSPs; ≤300 aa), and from among these, 125 are cysteine-rich (relative cysteine content of ≥ 5 %; SSCPs). Several of the multi-copy SSCPs are positioned adjacent to one another, and may have arisen due to tandem duplication events. Fifty-two of the SSCPs are predicted as members of 23 known ortholog families, and include enzymes that would facilitate breakdown of host tissue, such as cellulases, peptide hydrolases, exo-beta-1,3-glucanases, and aspartic and serine endopeptidases. The predicted effecterome consists of the remaining 87 SSCPs that do not have similarity to any known ortholog family and produce no BLASTp hits in NCBI GenBank. Twenty-nine of the predicted effectorome proteins are present as single copies in the genome assembly, and 58 of these proteins are members of multi-copy groups of 2–9 proteins.

As with the other three *Agaricomycotina* species surveyed, the plant pathogen *Rhizoctonia solani*, the saprophytic jelly fungus *Tremella mesenterica*, and the saprophytic fungus *Coprinopsis cinerea,* the *L. culmigenus* CBS 661.85 genome assembly contains a minimal cohort of gene clusters involved in the biosynthesis of secondary metabolites, including one non-ribosomal peptide synthase, five polyketide synthase clusters and eight terpene clusters (Table 1). There are 809 CAZyme domains in the genome of CBS 661.85. Comparison of the CAZyme profiles between *L. culmigenus* and three other *Agaricomycotina* species shows that *L. culmigenus* possesses a cohort of CAZymes that is most similar to the plant pathogenic *R. solani*. In particular, both *L. culmigenus* and *R. solani* have a greatly expanded cohort of glycoside hydrolases (GH), relative to the two saprophytic species (1.5x–1.6x; Table 1). The increase in GH motifs is due to an overall increase of all GH families, and not attributable to expansion of any one family. As expected for a fungus that utilizes grass as a nutritional substrate, there are only nine pectate lyase (PL) motifs present in the *L. culmigenus* genome. This reduction in PL motifs is in stark contrast with the genome of *R. solani* AG-1 1B, where 92 PL motifs were predicted. To our knowledge, the abundance of PL motifs in *R. solani* AG-1 1B is the largest repertoire reported for any fungus to date, and may play a role in *R. solani*'s wide host range and adaptability.

The mitochondrial genome of *L. culmigenus* CBS 661.85 was identified from scaffold 72, which measures 67.4 kb and has a G+C content of 27 % (average read coverage 3

505x). Based on tRNA and gene content, this scaffold likely represents the complete mitochondrial genome. Twenty-seven tRNAs were identified from the mitochondrial scaffold, coding for all 20 amino acids. Transfer RNAs for five amino acids were duplicated (Arg, Phe, Ser=2 tRNAs; Leu, Met=3 tRNAs). The standard set of 14 conserved mitochondrial protein coding genes (Bullerwell & Lang 2005) was identified, sometimes in multiple copies, including all three ATP synthase subunits (*ATP6*, *ATP8*, *ATP9*), apocytochrome b (*COB*), all three cytochrome oxidase subunits (*COX1*, *COX2*, *COX3*), and all seven NADH dehydrogenase subunits (*NAD1*, *NAD2*, *NAD3*, *NAD4*, *NAD4L*, *NAD5, NAD6*). As reported from *R. solani* Rhs1AP, most mitochondrial genes and tRNAs were found on the same strand (Losada *et al.* 2014); with the exception of *NAD1* and six of the tRNAs. Overall, 53 proteins were predicted, including the 14 conserved proteins, two predicted homing endonucleases with GIY_YIG domains, seven predicted homing endonucleases with LAGLIDADG domains, two rRNAs and 13 hypothetical proteins. Core mitochondrial proteins from *L. culmigenus* CBS 661.85 shared 79–100 % amino acid identity to their homologues in *R. solani* Rhs1AP (Losada *et al.* 2014). However, gene order comparisons showed almost no synteny between the mitochondria of these two species (data not shown). In general, the mitochondrial genome of *L. culmigenus* CBS 661.85 is average relative to the sequenced mitochondria observed from the other 32 basidiomycete fungi reported on NCBI GenBank (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/organelles.html, accessed 05/05/2015), with respect to size (av. = 72.2 kb), G+C content (av. = 28.3), tRNA content (av. = 24.5 tRNAs) and gene content (av. = 32.6 genes).

The draft genome sequence of *L. culmigenus* CBS 661.85 is the first *Corticiaceae* genomic resource publically available. This dataset will provide opportunities to resolve longstanding questions regarding taxonomy for organisms in this group, and may contribute to our understanding of the lifestyle of *L. culmigenus* of other *Corticiaceae* through comparative studies with closely related organisms.

Authors: **M. Malapi-Wight, L.A. Beirn, D. Veltri, and J.A. Crouch**

IMA Genome-F 4G

Draft genome sequence and annotation of *Thielaviopsis punctulata*, the causal agent of black scorch disease in date palm

Date palm (*Phoenix dactylifers*) is an important subsistence crop in arid regions due to its ability to grow under harsh environmental conditions such as high temperature and drought. Nevertheless, the ideal conditions for its growth and production are also favourable to fungal diseases, such as black scorch disease. *Thielaviopsis punctulata* (previously known as *Ceratocystis radicola*) is the causative agent of various plant-associated diseases, including black scorch (Suleman *et al.* 2001, Al-Naemi *et al.* 2014), rhizosis (Linde & Smit 1999), basal trunk root rot (Kile *et al.* 1993, Mohammadi 1992, Suleman *et al.* 2001), and sudden death disease in both the USA and South Africa (Bliss 1941, Linde & Smit 1999). Additionally, it infects other plant hosts causing leaf and root rot in pineapple (Kile 1993) and fruit rot in lemon (Mirzaee & Mohammadi 2005). This soil-borne pathogen (Wingfield *et al.* 1993) infects host plants through mechanical injury and growth cracks or during changes of moisture levels in the soil (Suleman *et al.* 2001).

An initial study identified *Thielaviopsis paradoxa* as the causal agent of black scorch in Qatar (Abbas & Abdulla 2003). In 2013, *T. punctulata* was isolated from black scorch-infected farms for the first time in Qatar (Al-Naemi *et al.* 2014). This fungus produces single ovate aleuroconidia with a smooth or slightly rough wall and hyaline to brown cylindrical phialoconidia in chains. *T. punctulata* differs from *T. paradoxa* in the means of aleuroconidial production. While *T. punctulata* produces single aleuroconidia from conidiophores, *T. paradoxa* forms chains of conidia (Abdullah *et al.* 2009).

The aim of this work was to sequence the whole genome of a virulent isolate (CRDP1) of *T. punctulata*, in addition to *de novo* assembly and functional annotation. Information from the fungal genome sequencing and annotation will allow us to identify fungal genes that are involved in pathogenicity and toxin production. The generated knowledge will help in designing an effective genetic control for black scorch disease.

SEQUENCED STRAIN

Qatar: Doha, isol. ex *Phoenix dactylifers* (date palm), May 2013, *F.A. Al-Naemi* (BPI 893173 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LAEV00000000. The version described in this paper is version LAEV01000000.

METHODS

CRDP1, a virulent isolate of *Thielaviopsis punctulata*, was grown in potato dextrose broth, and DNA was isolated using the CTAB method (Webb & Knapp 1990). TruSeq paired-end shotgun libraries and Nextera-based mate-pair libraries were generated and sequenced on a HiSeq2500, with a read length of 150 bp. The raw sequences were trimmed using Trimmomatic (Bolger *et al.* 2014); reads shorter than 40 bp were discarded. SPAdes (Bankevich *et al.* 2012) was used to perform *de novo* assembly of the genome. CEGMA (Parra *et al.* 2007) was used to identify ultra-conserved genes in *T. punctulata*. MAKER v. 2.31.7 (Cantarel *et al.* 2008) *ab initio* gene prediction pipeline was used for genome annotation. The MAKER v. 2.31.7 (Cantarel *et al.* 2008) *ab initio* gene

prediction pipeline was used for genome annotation and included the prediction models of CEGMA (Parra *et al*. 2007), SNAP (Johnson *et al*. 2008), AUGUSTUS (Stanke *et al*. 2004), GeneMark (Besemer & Borodovsky 2005), and RepeatMasker (Smit *et al*. 2010). For the first round through the MAKER pipeline, we used results from CEGMA and SNAP to train the MAKER gene prediction models. The set of proteins obtained from the first round were blasted against SwissProt and the best hits were chosen as the training set for the second round through the MAKER pipeline. For the second round through the MAKER pipeline, we first retrained SNAP and then ran AUGUSTUS with the SwissProt dataset; finally we used those two gene prediction models to rerun MAKER in order to refine its own gene prediction models. Subsequent iterations of the MAKER pipeline proceeded in the same fashion as before, i.e. results from the previous round of gene prediction models were used to retrain and refine the MAKER gene prediction models. We stopped at iteration four when no further improvement was observed. InterProScan v. 5.48 (Jones *et al*. 2014) and BLASTP v. 2.2.28+ (Camacho *et al*. 2009) against NCBI-RefSeq (release 69, 7 Jan. 2015) (Pruitt *et al*. 2012) were both performed to functionally annotate the proteins. MUMmer (Kurtz *et al*. 2004) and QUAST v. 2.3 (Gurevich *et al*. 2013) were used to perform a comparative analysis of the *T. punctulata* draft assembly with the publicly-available genomes of three other species of *Ceratocystis s. lat*.: *Huntiella moniliformis* (van der Nest *et al*. 2014), *C. fimbriata* (Wilken *et al*. 2013), and *C. manginecans* (van der Nest *et al*. 2014).

RESULTS AND DISCUSSION

Two libraries of shotgun (480 bp av. size) and mate-pair (5–7 kb) were prepared and sequenced using HiSeq2500 for the whole genome sequencing with total raw reads of 23 946 967 bp and 41 972 008 bp, respectively. The resulting draft assembly was 28.1 Mb in length (75x sequence coverage), and all contigs with a length of > 199 bp were submitted to NCBI's genome database. The draft genome comprised of 2379 scaffolds greater than 200 bp, with an N50 of 92 723 and an L50 of 87 scaffolds.

To assess the completeness of the *T. punctulata* genome, we ran CEGMA in the draft assembly of the *T. punctulata* genome and identified 244 out of 248 ultra-conserved genes in *T. punctulata* (98.38 % completeness). MAKER v. 2.31.7 predicted 5480 proteins using the *ab initio* gene prediction pipeline. InterProScan predicted functional annotations for 4841 proteins; with BLASTP reporting hits for 2749 proteins with significant sequence similarity (expect value less than 1.0×10^{-3}). We parsed GO terms from the InteproScan report and produced a list of unique entries containing 1636 GO categories (The Gene Ontology Consortium 2000). Additionally, we used ReviGO (Supek *et al*. 2011) to cluster these GO terms and identify major represented categories. These categories include ubiquitin-ubiquitin ligase activity, endopolyphosphatase activity, urate oxidase activity, cyanate hydratase activity, dynein binding, RNA ligase (ATP) activity, acetyl-CoA transporter activity and chitin binding. Of note, some of the annotated genes belong to toxin-related genes

such as necrosis inducing protein (NPP1) and cerato-platanin reflecting a potential functional role of the fungal toxin(s) in inducing disease symptoms.

We performed comparative genomics between *T. punctulata* draft assembly and the publicly available genomes of three other species of *Ceratocystis s. lat.*, using MUMmer (pairwise comparisons) and QUAST (four-way comparative analysis). The genomes were those of *Ceratocystis fimbriata* (NCBI WGS assembly APWK00000000.1) (Wilken *et al*. 2013), *C. manginecans* (WGS assembly JJRZ00000000.1) (de Beer *et al*. 2014), and *Huntiella moliniformis* (WGS assembly JMSH00000000.1) (de Beer *et al*. 2014). Though alignment results indicate significant similarity to the other genomes, they also suggest these species have less than 75 % in common between *T. punctulata* (CRDP1 isolate) and the other *Ceratocystis* and *Huntiella* genome sequences. Initial results suggest that these species may be related, but their genomes have undergone many changes since they split from a common ancestor. Altogether, our analyses confirm the degree of relatedness among the four *Ceratocystis* and *Huntiella* species previously reported (Wilken *et al*. 2013, de Beer *et al*. 2014).

The annotated genome of *T. punctulata* (CRDP1 isolate) will provide a basis for understanding fundamental biological mechanisms in this pathogen. This will also lead to further understanding of the molecular interactions between *T. punctulata* and date palm, and eventually help in designing effective genetic control measures for black scorch disease in date palm.

Authors: **O. Radwan, F.A. Al-Naemi, G. Rendon, and C.J. Fields**

REFERENCES

Abbas EH, Abdulla AS (2003) First report of neck bending disease on date palm in Qatar. *Plant Pathology* **52**: 790.

Abdullah S K, Asensio L, Monfort E, Gomez-Vidal S, Salinas J, *et al*. (2009) Incidence of the two date palm pathogens, *Thielaviopsis paradoxa* and *T. punctulata* in soil from date palm plantations in Elx, south-east Spain. *Journal of Plant Protection Research* **49**: 276–279.

Al-Naemi FA, Nishad R, Ahmed TA, Radwan O (2014) First report of *Thielaviopsis punctulata* causing black scorch disease on date palm in Qatar. *Plant Disease Journal* **98**: 1437.

Al-Raisi YM, B'Chir MM, Al-Mandhari AM, Deadman ML, Gowen SR (2011) First report of *Ceratocystis radicicola* associated with date palm disease in Oman. *New Disease Reports* **23**: 23.

Aveskamp M, De Gruyter J, Woudenberg J, Verkley G, Crous PW (2010) Highlights of the *Didymellaceae*: A polyphasic approach to characterise *Phoma* and related Pleosporalean genera. *Studies in Mycology* **65**: 1–60.

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.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, *et al.* (2013) MITOS: Improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**: 313–319

Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research* **33**: W451–W454.

Bihon W, Slippers B, Burgess T, Wingfield MJ, Wingfield BD (2011) *Diplodia scrobiculata* found in the Southern Hemisphere. *Forest Pathology* **41**: 175–181.

Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, *et al.* (2013) antiSMASH 2.0-a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research* **41**: W204–W212.

Bliss DE (1941) Relation of *Ceratostomella radicicola* to rhizosis of the date palm. *Phytopathology* **31**: 1123–1129.

Blodgett JT, Stanosz GR (1997) *Sphaeropsis sapinea* morphotypes differ in aggressiveness but both infect non-wounded red or jack pines. *Plant Disease* **81**: 143–147.

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*: 10.1093/bioinformatics/btu170.

Bonos SA, Clarke BB, Meyer WA (2006) Breeding for disease resistance in the major cool-season turfgrasses. *Annual Review of Phytopathology* **44**: 213–234.

Booth C, (1971). *The Genus Fusarium*. Kew: Commonwealth Mycological Institute.

Bullerwell CE, Lang BF (2005) Fungal evolution: the case of the vanishing mitochondrion. *Current Opinion in Microbiology* **8**: 362–369.

Burgess T, Gordon TR, Wingfield MJ, Wingfield BD (2004) Geographic isolation of *Diplodia scrobiculata* and its association with native *Pinus radiata*. *Mycological Research* **108**: 1399–1406.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **15**: 10:421.

Campbell, MS, Holt C, Moore B, Yandell M (2014) Genome annotation and curation using MAKER and MAKER-P. *Current Protocols in Bioinformatics* **48**: 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.

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.

Comeau AM, Dufour J, Bouvet GF, Jacobi V, Nigg M, *et al.* (2015) Functional annotation of the *Ophiostoma novo-ulmi* genome: insights into the phytopathogenicity of the fungal agent of Dutch elm disease. *Genome Biology and Evolution* **7**: 410–430.

Cortazar AR, Aransay AM, Alfaro M, Oguiza JA, Lavin JL (2014) SECRETOOL: integrated secretome analysis tool for fungi. *Amino Acids* **46**: 471–473.

Daub ME, Ehrenshaft M (2000) The photoactivated *Cercospora* toxin cercosporin: contributions to plant disease and fundamental biology. *Annual Review of Phytopathology* **38**: 461–490.

de Beer ZW, Duong TA, Barnes I, Wingfield BD, Wingfield MJ (2014) Redefining *Ceratocystis* and allied genera. *Studies in Mycology* **79**: 187–219.

de Beer ZW, 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.

de Gruyter J, Aveskamp MM, Woudenberg JHC, Verkley GJM, Groenewald JZ, *et al*. (2009) Molecular phylogeny of *Phoma* and allied anamorph genera: towards a reclassification of the *Phoma* complex. *Mycological Research* **113**: 508–519.

de Gruyter J, Boerema G, Van Der Aa H (2002) Contributions towards a monograph of *Phoma* (Coelomycetes) VI-2 Section *Phyllostictoides*: outline of its taxa. *Persoonia* **18**: 1–53.

de Gruyter J, van Gent-Pelzer MP, Woudenberg JH, van Rijswick PCJ, Meekes ETM, *et al.* (2012) The development of a validated real-time (TaqMan) PCR for detection of *Stagonosporopsis*

andigena and *S. crystalliniformis* in infected leaves of potato and tomato. *European Journal of Plant Pathology* **134**: 301–13.

de Wet J, Slippers B, Preisig O, Wingfield BD, Wingfield MJ (2003) Multiple gene genealogies and microsatellite markers reflect relationships between morphotypes of *Sphaeropsis sapinea* and distinguish a new species of *Diplodia*. *Mycological Research* **107**: 557–566.

de Wet J, Wingfield MJ, Coutinho TA, Wingfield BD (2002) Characterisation of the 'C' morphotype of the pine pathogen *Sphaeropsis sapinea*. *Forest Ecology and Management* **161**: 181–188.

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.

Diederich P, Lawrey JD, Sikaroodi M, Gillevet PM (2011) A new lichenicolous teleomorph is related to plant pathogens in *Laetisaria* and *Limonomyces* (*Basidiomycota, Corticiales*). *Mycologia* **103**: 525–533.

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.

Ellwood SR, Liu Z, Syme RA, Lai Z, Hane JK, *et al.* (2010) A first genome assembly of the barley fungal pathogen *Pyrenophora teres* f. *teres*. *Genome Biology* **11**: R109.

Espagneĺ E, Lespinetĺ O, Malagnacĺ F, Da C, Aury M, *et al.* (2008) The genome sequence of the model ascomycete fungus *Podospora anserina*. *Genome Biology* **9**: R77.

Farr DF, Rossman AY (2015) *Fungal Databases*. Beltsville: Systematic Mycology and Microbiology Laboratory, ARS, USDA; http://nt.ars-grin.gov/fungaldatabases/.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, *et al*. (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**: 1715–1719.

FRAC (2013) *Fungicide Resistance Action Committee code list: Fungicides sorted by mode of actions.* http://www.frac.info/publication/anhang/FRAC%20Code%20List%202013-update%20April-2013.pdf

Friesen TL, Zhang Z, Solomon PS, Oliver RP, Faris JD (2008) Characterization of the interaction of a novel *Stagonospora nodorum* host-selective toxin with a wheat susceptibility gene. *Plant Physiology* **146**: 682–693.

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.

Goodwin SB, M'Barek SB, Dhillon B, Wittenberg AHJ, Crane CF, *et al.* (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics* **7**: e1002070.

Groenewald S, van den Berg N, Marasas WFO, Viljoen A (2006) The application of high-throughput AFLPs in assessing genetic diversity in *Fusarium oxysporum* f.sp. *cubense*. *Mycological Research* **110**: 297–305.

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, Mena PJ, Clegg P, Wingfield BD, Wingfield MJ (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.

Hane JK, Anderson JP, Williams AH, Sperschneider J, Singh KB (2014) Genome sequencing and comparative genomics of the broad host-range pathogen *Rhizoctonia solani* AG8. *PLOS Genetics* **10**: e1004281.

Hane JK, Lowe RGT, Solomon PS, Tan K, Schoch CL, *et al.* (2007) *Dothideomycete*–plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum. Plant Cell* **19**: 3347–3368.

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 (1993) Diseases of conifers caused by species of *Ophiostoma* and *Leptographium*. In: *Ceratocystis and Ophiostoma: taxonomy, ecology and pathogenicity* (Wingfield MJ, Seifert KA, Webber JF, eds): 161–172. St Paul, MN: American Phytopathological Society Press.

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 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.

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

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

Jacobs, K, Wingfield MJ (2001) *Leptographium species: tree pathogens, insect associates and agents of blue-stain*. St Paul, MN: American Phytopathological Society Press.

Johnson AD, Handsaker RE, Pulit S, Nizzari MM, O'Donnell CJ, de Bakker PIW (2008) SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**: 2938–2939.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*: doi:10.1093/bioinformatics/btu031.

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): 173–180. St Paul, MN: American Phytopathological Society Press.

Klaasen JA, Nelson PE (1996) Identification of a mating population, *Gibberella nygamai* sp. nov., within the *Fusarium nygamai* anamorph. *Mycologia* **88**: 965–969.

Korf, I (2004) Gene finding in novel genomes. *BMC Bioinformatics* **5**: 59.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway 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.

Lagerberg, T, Lundberg G, Melin E (1927) Biological and practical researches into blueing in pine and spruce. *Svenska Skogsvårdsföreningens Tidskrift* **25**: 145–272.

Leslie JF, Summerell BA, (2006) *The Fusarium Laboratory Manual*. London: Blackwell Publishing.

Li I, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**: 713–714.

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**: 2178–2189.

Linde C, Smit WA (1999) First report of rhizosis caused by *Ceratocystis radicicola* on date palms in South Africa. *Plant Disease* **83**: 880.

Linnakoski, R, de Beer ZW, Duong TA, Niemelä P, Pappinen A, *et al.* (2012) *Grosmannia* and *Leptographium* spp. associated with conifer-infesting bark beetles in Finland and Russia, including *Leptographium taigense* sp. nov. *Antonie Van Leeuwenhoek* **102**: 375–399.

Losada L, Pakala SB, Fedorova ND, Joardar V, Shabalina SA, *et al.* (2014) Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen *Rhizoctonia solani* AG-3. *FEMS Microbiology Letters* **352**: 165–173.

Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**: 955–964.

Lukashin, A V, Borodovsky M (1998) GeneMark. hmm: new solutions for gene finding. *Nucleic Acids Research* **26**: 1107–1115.

Malapi-Wight M, Salgado-Salazar C, Demers J, Veltri D, Crouch JA (2015) Draft genome sequence of *Dactylonectria macrodidyma*, a plant pathogenic fungus in the *Nectriaceae*. *Genome Announcements* **3**: e00278-15.

Martinez D, Larrondo LF, Putnam N, Gelpke MDS, Huang K, *et al.* (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature Biotechnology* **22**: 695–700.

Mirzaee MR, Mohammadi M (2005) A new report of fruit rot of lemon caused by *Ceratocystis radicicola* from *Iranian Journal of Phytopathology* **153**: 11–14.

Mohammadi O (1992) Mycostop biofungicide – present status. In: *Biological Control of Plant Diseases* (Tjamos ES, Papavizas GC, Cook RJ, eds): 207–210. New York: Plenum Press.

Myburg H, Gryzenhout M, Heath R, Roux J, Wingfield BD, Wingfield MJ (2002) Cryphonectria canker on *Tibouchina* in South Africa. *Mycological Research* **106**: 1299–1306.

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.

Nakabonge G, Gryzenhout M, Wingfield BD, Wingfield MJ, Roux J (2007) Genetic diversity of *Chrysoporthe cubensis* in eastern and southern Africa. *South African Journal of Science* **103**: 261–264.

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.

O'Donnell K, Cigelnik E, Nirenberg HI (1998) Molecular systematic and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia* **90**: 465–493.

O'Donnell K, Nirenberg HI, Aoki T, Cigelnik E (2000) A multigene phylogeny of the *Gibberella fujikuroi* species complex: detection of additional phylogenetically distinct species. *Mycoscience* **41**: 61–78.

Ohm RA, de Jong JF, Lugones LG, Aerts A, Kothe E, *et al.* (2010) Genome sequence of the model mushroom *Schizophyllum commune*. *Nature Biotechnology* **28**: 957–910.

Palmer MA, Stewart EL, Wingfield MJ (1987) Variation among isolates of *Sphaeropsis sapinea* in the north central United States. *Phytopathology* **77**: 944–948.

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

Parra G, Bradnam K, Ning Z, Keane T, Korf I (2009) Assessing the gene space in draft genomes. *Nucleic Acids Research* **37**: 289–297.

Pethybridge S, Wilson C (1998) Confirmation of ray blight disease of pyrethrum in Australia. *Australasian Plant Pathology* **27**: 45–48.

Pethybridge SJ, Hay FS, Esker PD, Gent DH, Wilson CR, *et al.* (2008) Diseases of pyrethrum in Tasmania: challenges and prospects for management. *Plant Disease* **92**: 1260–1272.

Pethybridge SJ, Hay FS, Wilson CR, Groom T (2005) Development of a fungicide-based management strategy for foliar disease caused by *Phoma ligulicola* in Tasmanian pyrethrum fields. *Plant Disease* **89**: 1114–20.

Pethybridge SJ, Scott JB, Hay FS (2012) Lack of evidence for recombination or spatial structure in *Phoma ligulicola* var. *inoxydabilis* populations from Australian pyrethrum fields. *Plant Disease* **96**: 746–751.

Phillips AJL, Alves A, Abdollahzadeh J, Slippers B, Wingfield MJ, *et al.* (2013) The *Botryosphaeriaceae*: genera and species known from culture. *Studies in Mycology* **76**: 51–167.

Price AL, Jones NC, Pevzner PA (2005) *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**: i351–i358.

Pruitt KD, Tatusova T, Brown GR, Maglott DR (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Research* **40**: D130–5.

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.

Roux J, Meke G, Kanyi B, Mwangi L, Mbaga A *et al.* (2005) Diseases of plantation forestry trees in eastern and southern Africa. *South African Journal of Science* **101**: 1–5.

Sangalang AE, Burgess LW, Backhouse D, Duff J, Wurst M (1995) Mycogeography of *Fusarium* species in soils from tropical, arid and mediterranean regions of Australia. *Mycological Research* **99**: 523–528.

Schadler DL, Bateman DF (1974) *Ascochyta chrysanthemi* toxin: production and properties. *Phytopathology* **64**: 779–784.

Schadler DL, Bateman DF (1975) *Ascochyta chrysanthemi* toxin: purification and partial characterization. *Phytopathology* **65**: 912–917.

ARTICLE

ARTICLE

Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* **33**: W686–W689.

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27:** 863–864.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* doi:10.1093/bioinformatics/btv351.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol İ (2009) ABySS: A parallel assembler for short read sequence data. *Genome Research* **19**: 1117–1123.

Smit AFA, Hubley R, Green P (1996–2010) RepeatMasker Open-3.0. http://www.repeatmasker.org.

Smit AFA, Hubley R, Green P (2013–2015). RepeatMasker Open-4.0. http://www.repeatmasker.org.

Stajich JE, Wilke SK, Ahren D, Au CH, Birren BW, *et al.* (2010) Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proceedings of the National Academy of Sciences, USA* **107**: 11889–11894.

Stalpers JA, Loerakker WM (1982) *Laetisaria* and *Limonomyces* species (*Corticiaceae*) causing pink diseases in turf grasses. *Canadian Journal of Botany* **60**: 529–537.

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637–644.

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

Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**: 62.

Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research* **32**: W309–W312.

Stanke, M, S Waack (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**: ii215–ii225.

Stanosz GR Swart WJ, Smith DR (1999) RAPD marker and isozyme characterization of *Sphaeropsis sapinea* from diverse coniferous hosts and locations. *Mycological Research* **103**: 1193–1202.

Stevens FL (1907) The chrysanthemum ray blight. *Botanical Gazette* **44**: 241–258.

Suleman P, Al-Musallam A, Menezes CA (2001) The effect of solute potential and water stress on black scorch caused by *Chalara paradoxa* and *Chalara radicicola* on date palms. *Plant Disease* **85**: 80–83.

Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6:** e21800.

The Gene Ontology Consortium (2000) Gene ontology: tool for the unification of biology. *Nature Genetics* **25**: 25–29.

Trimboli DS, Burgess LW (1983) Fungi associated with basal stalk rot and root rot of drayland grain sorghum in new South Wales. *Plant Protection Quarterly* **1**: 3–9.

Tsui CKM, DiGuistini S, Wang Y, Feau N, Dhillon B *et al.* (2013) Unequal recombination and evolution of the mating-type (*MAT*) loci in the pathogenic fungus *Grosmannia clavigera* and relatives. G3: *Genes|Genomes|Genetics* **3**: 465–480.

Vaghefi N, Ades PK, Hay FS, Pethybridge SJ, Ford R, *et al.* (2015a) Identification of the *MAT1* locus in *Stagonosporopsis tanaceti*, and exploring its potential for sexual reproduction in Australian pyrethrum fields. *Fungal Biology* **119**: 408–419.

Vaghefi N, Hay FS, Ades P, Pethybridge S, Ford R, *et al.* (2015b) Rapid changes in the genetic composition of *Stagonosporopsis tanaceti* population in Australian pyrethrum fields. *Phytopathology* **105**: 358–369.

Vaghefi N, Pethybridge SJ, Ford R, Nicolas ME, Crous PW, *et al.* (2012) *Stagonosporopsis* spp. associated with ray blight disease of *Asteraceae*. *Australasian Plant Pathology* **41**: 675–686.

van der Aa HA, Noordeloos ME, de Gruyter J (1990) Species concepts in some large genera of the coelomycetes. *Studies in Mycology* **32**: 3–19.

van der Merwe N A, 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 M, Beirn LA, Crouch JA, Demers JE, De Beer ZW, *et al.* (2014b) 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) Draft genome sequences of *Diplodia sapinea, Ceratocystis manginecans,* and *Ceratocystis moniliformis*. *IMA Fungus* **5**: 135–140.

Vermeulen M, Gryzenhout M, Wingfield MJ, Roux J (2011) New records of the *Cryphonectriaceae* from southern Africa including *Latruncellus aurorae* gen. sp. nov. *Mycologia* **103**: 554–569.

Vermeulen M, Gryzenhout M, Wingfield MJ, Roux J (2013) Population structure of *Chrysoporthe austroafricana* in southern Africa determined using Vegetative Compatibility Groups (VCGs). *Forest Pathology* **43**: 124–131.

Webb DM, Knapp SJ (1990) DNA extraction from a previously recalcitrant plant genus. *Molecular Biology Reports* **8**: 180–185.

Wibberg D, Jelonek L, Rupp O, Hennig M, Eikmeyer F, *et al.* (2013) Establishment and interpretation of the genome sequence of the phytopathogenic fungus *Rhizoctonia solani* AG1-IB isolate. *Journal of Biotechnology* **167**: 142–155.

Wiemann P, Sieber CMK, von Bargen KW, Studt L, Niehaus E-M, *et al.* (2013) Deciphering the cryptic genome: genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *PloS Pathogens* **9**: e1003475.

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

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

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

ARTICLE

Yin Y, Mao X, Yang J, Chen X, Mao F, *et al.* (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **40**: W445–W451

Zdobnov EM, Apweiler R (2001) InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**: 847–848.