IMA Genome-F 3

Draft genomes of *Amanita jacksonii*, *Ceratocystis albifundus*, *Fusarium circinatum*, *Huntiella omanensis*, *Leptographium procerum*, *Rutstroemia sydowiana*, and *Sclerotinia echinophila*

Magriet A. van der Nest¹, Lisa A. Beirn², Jo Anne Crouch³, Jill E. Demers³, Z. Wilhelm de Beer⁴, Lieschen De Vos¹, Thomas R. Gordon⁵, Jean-Marc Moncalvo⁶,⁷,⁸, Kershney Naidoo¹, Santiago Sanchez-Ramirez⁴, Danielle Roodt¹, Quentin C. Santana¹, Stephanie L. Slinski⁹, Matt Stata⁸, Stephen J. Taerum¹, P. Markus Wilken¹, Andrea M. Wilson¹, Michael J. Wingfield⁴, and Brenda D. Wingfield¹

¹Department 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
²Department of Plant Biology and Pathology, Rutgers University, 59 Dudley Road, New Brunswick, NJ 08901, USA
³Systematic Mycology and Microbiology Laboratory, US Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705, USA
⁴Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, P. Bag X20, Pretoria 0028, South Africa
⁵Department of Plant Pathology, University of California, Davis, CA 95616, USA
⁶Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada
⁷Department of Natural History, Royal Ontario Museum, 100 Queen’s Park, Toronto, ON M5S 2C6, Canada
⁸Centre for the Analysis of Genome Evolution and Function, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada

Abstract: The genomes of fungi provide an important resource to resolve issues pertaining to their taxonomy, biology, and evolution. The genomes of *Amanita jacksonii*, *Ceratocystis albifundus*, a *Fusarium circinatum* variant, *Huntiella omanensis*, *Leptographium procerum*, *Sclerotinia echinophila*, and *Rutstroemia sydowiana* are presented in this genome announcement. These seven genomes are from a number of fungal pathogens and economically important species. The genome sizes range from 27 Mb in the case of *Ceratocystis albifundus* to 51.9 Mb for *Rutstroemia sydowiana*. The latter also encodes for a predicted 17,350 genes, more than double that of *Ceratocystis albifundus*. These genomes will add to the growing body of knowledge of these fungi and provide a valuable resource to researchers studying these fungi.

Key words: Dispensable chromosome

IMA Genome-F 3A

Draft genome of the edible ectomycorrhizal basidiomycete *Amanita jacksonii* TRTC168611 from Awenda Provincial Park, Ontario, Canada

The genus *Amanita* (*Agaricales, Basidiomycota*) is primarily known for species that produce deadly toxic compounds such as phallotoxins and amatoxins (Vetter 1998). However, a few taxa from this genus – such as *Amanita jacksonii*, *A. hemibapha*, *A. caesarea* and allies – are also traditionally and culturally known to be excellent wild edible mushrooms in many regions of the world (Pegler 2002, Boa 2004). Most species of the genus are known to be ectomycorrhizal (EM), living in a mutualistic symbiosis with many members of woody tree families (Taylor & Alexander 2005), while few others are saprophytic. The latter condition is known to be ancestral, meaning that the EM habit probably evolved once within the genus (Wolfe et al. 2011). Currently, only two *Amanita* genomes are publicly available from the Joint Genome Institute (JGI) (http://genome.jgi.doe.gov/agaricomycotina/agaricomycotina.info.html) and the Myco-rrhizal Genomics Initiative (http://mycor.nancy.inra.fr/IMGC/Mycogenomes) as components of sequencing the Fungal Tree of Life (Martin et al. 2011): a strain from the saprophytic *A. thiersii*, and a strain from the EM *A. muscaria* var. *guessowii* (from here on referred to as *A. muscaria*). In addition, the partial genome of *A. bisporigera* has been sequenced with the purpose of isolating genes producing toxic compounds (Hallen et al. 2007). Furthermore, three...
other Amanita genomes have been recently sequenced, A. brunnescens, A. polypyramis, and A. inopinata, with the aim of assessing the dynamics of transposable elements in EM and asymbiotic species within the genus (Hess et al. 2014). Amanita muscaria has also been used as a model for understanding the biosynthetic pathways of betalain pigments, which are commercially used to dye food and shown to have antioxidant properties (Hinz et al. 1997, Strack et al. 2003).

Amanita jacksonii is a non-toxic EM member of the genus (Fig. 1), which also produces betalains. The genomic data here presented should facilitate further comparative genomic analyses between members of the genus Amanita. It will also shed light on the evolution of toxicity, the EM habit, and betalain biosynthetic pathways.

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The draft genome sequence of Amanita jacksonii (TRTC168611) has been deposited in EMBL/DDBJ/GenBank under the accession no. AYNK00000000. This submission represents the first draft version.

METHODS

Genomic DNA was isolated from the context of the stipe of a fresh specimen by removing the surface tissue with a clean razor blade. Pieces (~200 mg) of the stipe context were then frozen at -20 °C until the extraction step, for which we used a 2% CTAB protocol (modified from Zolan & Pukkila (1986)). This protocol included a proteinase-K digestion step followed by a chloroform:isoamylalcohol (1:24) extraction, RNA denaturation, and isopropanol precipitation (a document with the details can be found at https://sites.google.com/site/santiagosnchezrmirez/home/amanita-jacksonii-genomics/genomic-dna-extraction). A whole-genome shotgun approach was used to produce one library for Roche 454 pyrosequencing (standard single-ended) and one TruSeq library for Illumina HiSeq 2000 (paired-ended, insert size: 350–500 bp) (conducted at the Duke Genome Sequencing & Analysis Core Resource; http://www.genome.duke.edu/cores/sequencing/). The libraries
were run in half of a 454 pico-titre plate (PTP) and half of an Illumina lane, respectively. RAY v. 1.7 (Boisvert et al. 2010) was used to assemble all reads combined in multi-threaded mode. Gene prediction was conducted on contigs > 1000 bp using AUGUSTUS v. 2.5.5 (Stanke et al. 2004) and the hidden-markov-model (HMM) profile of Laccaria bicolor. BLAST2GO v. 2.6.6 (Conesa et al. 2005) was used for protein annotation. We used the CEGMA (Core Eukaryotic Genes Mapping Approach) pipeline to assess the level of genome completeness based on the qualitative and quantitative conditions of eukaryotic clusters of orthologous groups and core eukaryotic genes (CEGs) (Parra et al. 2007, 2008). Gene orthology and comparative genomic analyses were performed using custom Python scripts and BLAST, based on reciprocal best hits (Moreno-Hagelsieb & Latimer 2008).

**RESULTS AND DISCUSSION**

The 454 run yielded of ca. 1.4 million reads ranging from ~100 to ~1000 bp, whereas the Illumina platform produced ca. 157 million reads after quality control filtering. Both runs had read yields within their platform standards (Buermans & den Dunnen 2014). The combined read assembly produced a 30 285 912 bp draft genome with 2 988 contigs (>1000 bp), of which the largest was 504 181 bp. The average contig length was 10 139 bp, and N50 and N90 stats were 26 643 and 3 566 bp, respectively. According to CEGMA, the genome completeness based on 248 CEGs resulted in 93.15 % and 95.97 % for complete and partial genes, respectively. Similar genome statistics have been found for other recent Amanita genome sequencing projects (Hess et al. 2014; http://genome.jgi-psf.org/). For instance, the genome size of A. muscaria (61), which represent 60 % of the genome (48 % exons and 12 % introns). The hypothetical proteome ranged from 47 to 5 455 amino acids in length. In contrast, far more genes are predicted in the genomes of A. muscaria and A. thiersii, which include 18 153 and 10 354 structural genes, respectively (http://genome.jgi-psf.org/). As expected, A. thiersii has a higher number of genes encoding a glycoside hydrolase (EC:3.2.1) domain (128), compared to A. jacksonii (80) and A. muscaria (61). However, the number of cellular genes (GH5) was comparable with 12 (A. jacksonii), eight (A. muscaria), and 10 (A. thiersii) genes each. Furthermore, results from BLAST2GO protein annotation suggest that the genome of A. jacksonii is enriched with proteins related to metabolic and cellular processes including: oxido-reduction, biosynthetic, nitrogen compound processes, as well as primary, cellular and macromolecule metabolic processes (carbohydrate, lipid, phosphorus and DNA metabolism, gene expression). Finally, the reciprocal best-hits method described in Moreno-Hagelsieb & Latimer (2008) indicate that A. jacksonii shares 4 408 to 5 178 putative orthologs with the other Agaricales species presently available in the JGI database, sharing the most with A. muscaria and the least with Agaricus bisporus. Interestingly, this method suggests that more putative orthologs are shared with other non-congeneric ectomycorrhizal species, such as Laccaria bicolor, Hebeloma cylindrosporum, and Tricholoma matsutake, than with the congeneric Amanita thiersii, which is saprobic.

**Draft genome sequence of Ceratocystis albifundus**

The genus Ceratocystis (Ascomycota, Microascales) includes important pathogens of woody and herbaceous plants (Wingfield et al. 2013a, b, de Beer et al. 2014). Ceratocystis albifundus is thought to be native to southern Africa where it causes an important canker and wilt disease on non-native Acacia mearnsii propagated in intensively managed plantations (Roux & Wingfield 2013). The fungus has also been isolated from the wounds of many native South African trees and woody plants (Roux et al. 2007). Symptoms of infection include streaked discoloration of the vascular tissue, stem cankers, gum exudation, wilt and tree death, which can result in substantial economic losses for plantation owners (Morris et al. 1993, Roux et al. 1999, Barnes et al. 2005).

Like other Ceratocystis spp., C. albifundus produces a sweet odour that attracts insects such as nitidulid beetles that act as vectors of the fungus (Heath et al. 2009). This particular species can easily be distinguished from other morphologically similar Ceratocystis species by the presence of light coloured ascomatal bases, and substantial sequence differences in multiple gene regions (Wingfield et al. 1996).

The aim of this study was to sequence the genome of C. albifundus and thus to enable comparative studies with other Ceratocystis spp. In this regard, the genomes of two other species of Ceratocystis are publically available. These include the sweet potato pathogen C. fimбриata (Wilken et al. 2013) and the mango wilt pathogen, C. manginecans (van der Nest et al. 2014). The genome sequences for two species in the related genus Huntiella, that includes species formerly accommodated in the C. moniliformis complex (de Beer et al. 2014), are also publically available. These species are the saprophytes Huntiella omanensis (this issue) and H. moniliformis (van der Nest et al. 2014).

**SEQUENCED STRAIN**

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The Whole Genome Shotgun project of this *Ceratocystis albifundus* isolate has been deposited at D8J/EMBL/GenBank with the accession number JSSU00000000. Here we describe version JSSU01000000.

METHODS

Sequencing of the *Ceratocystis albifundus* isolate was performed on the Genome Analyzer Ix next-generation sequencing platform (Illumina) (Metzker 2009) at the Genome Centre, University of California, Davis (CA, USA). Paired-end libraries with respective insert sizes of 300 bp and 600 bp were used to generate read lengths of 100 bases. The CLC Genomics Workbench v. 6.0.1 (CLCBio, Aarhus, Denmark) was subsequently used to trim reads of bases. The CLC Genomics Workbench v. 6.0.1 (CLCBio, Aarhus, Denmark) was subsequently used to trim reads of poor quality (limit of 0.05) as well as terminal nucleotides. A tenth chromosome appears to be dispensable (Xu et al. 2014). The genome sequence information for *C. albifundus* will aid in investigations of the significance of these genome differences as well as other aspects of the biology of *Ceratocystis* spp. in general.

Authors: D. Roodt, M.A. van der Nest, P.M. Wilken, K. Naidoo, M.J. Wingfield, and B.D. Wingfield

IMA Genome-F 3C

**Draft genome sequence of *Fusarium circinatum***

*Fusarium circinatum* is an important pathogen of susceptible *Pinus* spp. causing a disease commonly known as pitch canker, a name describing the copious amount of resin that accumulates at the site of infection (Hepting & Roth 1946). This fungus is a member of the *F. fujikuroi* complex that includes many important pathogens of cultivated plants (Kvas et al. 2009, Geiser et al. 2013). Due to their importance as plant pathogens, the genomes of several *Fusarium* spp. have been published (*Fusarium* Comparative Sequencing Project; Jeong et al. 2013, Wiemann et al. 2013), including that of *F. circinatum* (Wingfield et al. 2012).

Members of the *F. fujikuroi* complex are known to have twelve chromosomes (Xu et al. 1995, Wiemann et al. 2013). The twelfth chromosome appears to be dispensable (Xu et al. 1995, Jurgenson et al. 2002, Ma et al. 2010) and can be strain-specific in members of the *F. fujikuroi* complex (Wiemann et al. 2013). A laboratory strain of *F. circinatum* (GL1327) has recently been found not to possess the twelfth chromosome when visualised using pulsed-field gel electrophoresis (PFGE) (Slinski et al., unpubl.). The aim of this study was to conduct whole genome shotgun sequencing of this strain and thus to allow comparisons with the genome of the already sequenced *F. circinatum* strain (Wingfield et al. 2012), as well as to other sequenced members of the *F. fujikuroi* complex. This formed part of a larger objective to expand our knowledge of dispensable chromosomes and their roles in the biological processes of an important group of plant pathogenic *Fusarium* spp.

**SEQUENCED STRAIN**

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The *Fusarium circinatum* genomic sequence has been deposited at DDBJ/EMBL/GenBank under the accession JRV311000000. The version described in this paper is the first version, JRV311000000.

METHODS

Genomic DNA was isolated (Iturritxa et al. 2011) from *Fusarium circinatum* isolate CBS 138821 and subjected to sequencing. Two mate-pair libraries (1 kb insert size) were constructed and sequenced using SOLID™ V4 technology (Applied Biosystems) at SEQOMICS (Hungary). Also, a single-read library was sequenced using the Illumina HiSeq 2500 at the Genome Centre, University of California (Davis, USA). All sequences had an average read length of 50 bp. Poor quality and duplicate reads were removed using CLC Genomics Workbench v. 6.5 (CLCbio, Aarhus, Denmark). Assembly and scaffolding was done using ABySS v. 1.3.7 (Simpson et al. 2009). Closing of gapped regions was performed using GapFiller v. 1.11 (Boetzer & Pirovano 2012). The completeness of the genome was evaluated using CEGMA (Parra et al. 2008). Putative open reading frames (ORFs) were predicted using AUGUSTUS (Hoff & Stanke 2013) with the *F. graminearum* frames (ORFs) were predicted using AUGUSTUS (Hoff & McDonald 2012), with chromosome 12 fitting the description of a dispensable chromosome that would form part of the accessory genome for members of the *F. fujikuroi* complex (Ma et al. 2013). The discovery of this laboratory strain of *F. circinatum* in which chromosome 12 is absent will enable studies of the dispensable chromosomes in this species.


IMA Genome-F 3D

Draft genome sequence of the fungus, *Huntiella omanensis*

Species in the genus *Huntiella* (De Beer et al. 2014) include a group of generally saprobic fungi commonly found on freshly cut timber or wounds on trees. Only one species, *H. bhutanensis*, is known to be associated with bark beetles on conifers (van Wyk et al. 2004). These fungi were previously accommodated in the Ceratocystis moniliformis complex (Wingfield et al. 2013a, b, de Beer et al. 2014). Members of the genus *Huntiella* are interesting due to their morphological and ecological similarities to species of *Ceratocystis*, which includes some important pathogens of trees (Wingfield et al. 2013a, b).

*Huntiella omanensis* was first described from diseased mango trees in Oman. However, a second fungus, *Ceratocystis manginecans*, was found to be the causal agent of this disease (Al-Subhi et al. 2006, van Wyk et al. 2007), while *H. omanensis* is weakly pathogenic. The fungus produces hat-shaped ascospores from relatively short necked ascomata with dark, globose and spiny bases (Al-Subhi et al. 2006). As in other species of *Ceratocystidaceae*, the ascospores exude from the ascomatal necks in slimy masses that are picked up by insects attracted to the fruity aromas produced by these fungi (Al-Subhi et al. 2006).

The aim of this study was to produce a draft nuclear genome assembly for an isolate of *H. omanensis*. This was intended to enable genome level comparisons with other species of *Huntiella* (van der Nest et al. 2014) and the family *Ceratocystidaceae* (Wilkens et al. 2013, van der Nest et al. 2014). For example, it would make possible comparisons of related fungi that differ in their pathogenicity levels, mating strategies and other important ecological and/or biological aspects.
The Whole Genome Shotgun project of the Huntiella omanensis genome has been deposited at DDBJ/EMBL/GenBank under the accession no. JSUI00000000.

**METHODS**

Genomic DNA was isolated and sequenced on the Genomics Analyzer Iix platform (Illumina) at the Genome Centre, University of California at Davis (CA, USA). Paired-end libraries with insert sizes of approximately 350 and 600 bases were used to produce reads with an average length of 97 bases. Poor-quality reads and terminal nucleotides were discarded and trimmed using the software package CLC Genomics Workbench v. 6.0.1 (CLCBio, Aarhus, Denmark). The remaining reads were assembled using the Velvet de novo assembler (Zerbino & Birney 2008), with an optimized k-mer size of 83. These assemblies were subsequently scaffolded using SSPACE v. 2.0 (Boetzer et al 2011) and gaps were filled using GapFiller v. 2.2.1 (Boetzer & Pirovano 2012). Whole genome completeness was measured using the Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra et al. 2007). Finally, open reading frames (ORFs) were predicted using AUGUSTUS (Stanke et al. 2004) based on the gene models for Fusarium graminearum.

**RESULTS AND DISCUSSION**

The Huntiella omanensis draft genome had an estimated size of 31 502 652 DNA bases, a 9x coverage, 550 contig size of 41 324 bases and a mean GC content of 47.6 %. The assembly resulted in 8 127 contigs, with 1 638 being discarded and trimmed using the software package CLC Genomics Workbench v. 6.0.1 (CLCBio, Aarhus, Denmark). The remaining reads were assembled using the Velvet de novo assembler (Zerbino & Birney 2008), with an optimized k-mer size of 83. These assemblies were subsequently scaffolded using SSPACE v. 2.0 (Boetzer et al 2011) and gaps were filled using GapFiller v. 2.2.1 (Boetzer & Pirovano 2012). Whole genome completeness was measured using the Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra et al. 2007). Finally, open reading frames (ORFs) were predicted using AUGUSTUS (Stanke et al. 2004) based on the gene models for Fusarium graminearum.

**IMAGES**

**Draft genome sequence of Leptographium procerum**

*Leptographium procerum* is an ascomycetous fungus in Ophiostomatales (Jacobs et al. 2001). This fungus is typically vectored between pines, spruce and fir by a variety of arthropods, particularly root and root collar infesting bark beetles and weevils (Jacobs & Wingfield 2001). *Leptographium procerum* has been reported from eastern North America, several European countries, Japan, China, and New Zealand (Jacobs & Wingfield 2001, Masuya et al. 2013). The populations in China and New Zealand are introduced and suspected to be invasive (Reay et al. 2002, Lu et al. 2011, Taerum et al. 2013), although their relevance is not fully understood.

*Leptographium procerum* has been linked to the decline and mortality of pines in North America (Lackner & Alexander 1982, Alexander et al. 1988, Klepzig et al. 1991). However, it has been suggested that the presence of *L. procerum* in diseased trees is coincidental to the presence of its insect vectors and that it is not a primary pathogen on North American pines (Wingfield et al. 1988). More recently, *L. procerum* was discovered to be the most common associate of the red turpentine beetle *Dendroctonus valens* in the invasive range of the insect in China (Lu et al. 2009a, b). The beetle was introduced from North America, where *L. procerum* is a common associate in part of the range of *D. valens* (Taerum et al. 2013). In China, *L. procerum* has been reported only as an associate of *D. valens*, suggesting that the fungus may have coinvaded China with *D. valens*. The association between *D. valens* and *L. procerum* has been suggested to contribute towards the aggressive tree-kill ing behaviour of *D. valens* in China. This is because pine trees native to China may produce larger quantities of monoterprenes that attract *D. valens* when infected by *L. procerum* (Lu et al. 2010, 2011, Sun et al. 2013).

In this study we sequenced the genome of an American isolate of *L. procerum* and produced a draft genome sequence of the fungus. This was done in order to provide fundamental data to develop tools such as population markers (i.e. microsatellites, SNPs) to better understand the global diversity of the fungus including its origin in China and New Zealand. In addition, this is the first genome sequenced from the *Leptographium procerum*-species complex, that currently includes nine described species (Yin et al. 2015). The genome will also be useful for future comparative genomics studies within the *L. procerum*-species complex and among species complexes in the Ophiostomatales.
SEQUENCED STRAIN

USA: Maine: Massabesic Experimental Forest, isol. ex Pinus resinosa (red pine), Nov. 2009, M.J. Wingfield (CBS 138288, CMW 34542; dried culture: PREM 61058). This culture represents the ex-epitype for *L. procerum* (MBT 198257), designated by Yin et al. (2015), and has the mating-type gene MAT1-2-1 (Duong et al. 2013).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The DDBJ/EMBL/GenBank accession number for the *Leptographium procerum* Whole Genome Shotgun project is JRUC00000000.

METHODS

DNA was extracted from a single spore culture following Möller et al. (1992). We submitted the extracted DNA to Inqaba Biotec (Pretoria, South Africa) for Illumina sequencing, where a 2 x 250 bp paired-end library was generated using the MiSeq v. 2500 cycle kit (Illumina, San Diego, USA). The average insert size was ~500 bp. Pairing and trimming was done in CLC Genomics Workbench v. 5.0.1 to pair reads and discard those of poor quality (limit 0.05). The remaining reads were assembled using the de novo assembler, Velvet v. 1.1 (Zerbino and Birney 2008) using an optimised k-mer of 79. Scaffolding was conducted using SSPACE v. 1.1 (Boetzer et al. 2011), and gap-closing was conducted using GapFiller v. 1.11 (Boetzer & Pirovano 2012). AUGUSTUS gene predictor (http://bioinf.uni-greifswald.de/augustus/) was used to estimate the number of open reading frames (ORFs) present in the genome using the Fusarium graminearum gene models (Stanke et al. 2006). We used the Core Eukaryotic Genes Mapping Approach (CEGMA) to evaluate the completeness of the assembly (Parra et al. 2007).

RESULTS AND DISCUSSION

We assembled a draft genome of 3 226 contigs. Of these, 2 460 were retained after filtering out those contigs smaller than 500 bases. The draft genome had an estimated genome size of 28.6 Mb, an average coverage of 32x, a mean GC content of 54.77 % and an N50 contig length of 22 487 bases. The assembly had a CEGMA completeness score of 92.74% for the complete set of eukaryotic genes and was predicted to contain 9 263 ORFs resulting in a putative density of 324 ORFs/Mb.

The estimated genome size was similar to those of *Grossmannia clavigera* (~29.8 Mb; DiGuistini et al. 2011) and *Leptographium longiclavatum* (~28.9 Mb; Ojeda et al. 2014), fungal species that are close relatives of *L. procerum* (de Beer & Wingfield 2013). In addition, the number of estimated ORFs in *L. procerum* was comparable to the numbers of ORFs found in *G. clavigera* (8314, excluding the mitochondrial genome) and *L. longiclavatum* (9861; larger than 33 amino acids: 9052).

Future transcriptome analyses of the *L. procerum* genome will improve the accuracy of the predicted protein-coding genes. Genome analyses will allow for comparisons between *L. procerum* and other fungi in the Ophiostomatales and thus to better understand differences in associations between these fungi, their hosts, and their vectors. In addition, access to the genome will allow for the development of population markers to better understand the global diversity and movement of *L. procerum* and its relatives.

Authors: S.J. Taerum, D. Roodt, Z.W. de Beer, and M.J. Wingfield

IMA Genome-F 3F

Draft genome sequence of *Rutstroemia sydowiana*

*Rutstroemia* (Rutstroemiaceae, Helotiales, Ascomycota) is a genus of fungi with a largely unknown ecological role, though many live as saprobes (Holst-Jensen et al. 1997). *Rutstroemia* species have been reported in association with *Betula* sp., *Quercus* sp., and *Carex* sp. (Holst-Jensen et al. 1997, Carbone & Kohn 1993), as well as dead plant tissue (e.g. CBS 854.97, CBS 115.86) and non-plant substrates such as rabbit dung (e.g. CBS 465.73), where they produce the substratal stroma characteristic of the genus (Holst-Jensen et al. 1997). *Rutstroemia* species may be unique among the apothecia-forming fungi, with apothecium development occurring in the late summer or autumn, rather than the spring fruiting of other *Sclerotiniaceae* (Whetzel 1945) (Fig. 2).

Fungi in Rutstroemiaceae are very closely related to the economically important *Sclerotiniaceae*, a family of necrotrophic phytopathogens and saprobes (Carbone & Kohn 1993, Holst-Jensen et al. 1997). Much like the ecology of these fungi, evolutionary relationships and taxonomy within the family and genus are poorly defined. At present, *Rutstroemiaceae* is considered to be polyphyletic (Johnston et al. 2013), and extensive, wide-scale sampling and molecular phylogenetic analysis are needed before any conclusions can be drawn about relationships within this family. Generating genomic resources for *Rutstroemiaceae* would provide a basis for developing molecular markers to resolve the taxonomy in this family, and may give insight into shared biological pathways between this family and the closely related *Sclerotiniaceae*. The goal of this study was to produce a whole genome sequence for a member of the genus *Rutstroemia*, the type genus for the *Rutstroemiaceae* family. Here we report the draft genome of *Rutstroemia sydowiana* (Fig. 2).

SEQUENCED STRAIN

The Whole Genome Shotgun project of the Rutstroemia sydowiana (CBS 115975) genome has been deposited in NCBI GenBank under the accession no. JWJB00000000, version JWJB01000000.

METHODS

DNA extraction, generation and assembly of Illumina next-generation sequence reads, and downstream analyses were performed as described for Sclerotinia echinophila (CBS 111548) as outlined elsewhere in this paper.

RESULTS AND DISCUSSION

The 51.9 Mb genome of Rutstroemia sydowiana (CBS 115975) is contained in 11 591 scaffolds of 500 bp or greater in length (6 217 scaffolds > 1 kbp). A summary of the genome assembly is presented in Table 1. The total length of the coding sequence is 24.4 Mb, with 17 350 predicted genes covering 47.1 % of the 51.9 Mb genome assembly. Mean gene and protein lengths are 1 408 bp and 408 aa, respectively, with an average gene density of 334 genes per Mb. The assembly is estimated as 99 % complete, based on the presence of a core set of conserved eukaryotic genes. An average of 1.4 introns are found per gene, with an average intron length of 97.7 bp (maximum = 1 460 bp). No introns are predicted in 18.9 % of the putative genes.

This draft genome assembly of R. sydowiana (CBS 115975) is the first genome sequence data generated from the family Rutstroemiaceae. Currently, the most closely related organisms with sequenced genomes are members of the sister family, Sclerotiniaceae: Sclerotinia borealis, S. echinophila, S. sclerotiorum, and three isolates of Botrytis cinerea (Anselem et al. 2011, Blanco-Ulate et al. 2013, Mardanov et al. 2014; this paper). There are several notable differences between the R. sydowiana genome and those of the sequenced Sclerotiniaceae. Compared with the Sclerotiniaceae genomes, the 51.9 Mb R. sydowiana CBS 115975 assembly is on average 11.5 Mb larger (Sclerotiniaceae range 38.3–42.5 Mb). Along with the relatively larger overall genome size, R. sydowiana CBS 115975 also has a greater number of predicted gene models (17 350) than any of the Sclerotiniaceae, where predicted gene models range between 10 171 (S. borealis; Mardanov et al. 2014) to 16 448 (B. cinerea BcDW1 causing noble rot of grape; Blanco-Ulate et al. 2013).

A total of 1 865 predicted proteins from the R. sydowiana predicted proteome possessed a transmembrane domain signature. The assembly is predicted to contain 1 120 secreted proteins, 51 that include a transmembrane domain. Although the R. sydowiana predicted secretome makes up just 6.5 % of the total predicted proteome, this cohort of genes is 21–44 % larger than the secretomes predicted from B. cinerea, S. echinophila, and S. sclerotiorum (Anselem et al. 2011, this paper). Based on BLASTp searches using the predicted secretome proteins as queries against the proteomes of B. cinerea, S. sclerotiorum, and S. echinophila (CBS 111548), 125 of these predicted extracellular proteins (11.2 %) are unique to R. sydowiana.
**Table 1.** Summary of whole genome DNA sequence assemblies generated in the current study, *Rutstroemia sydowiana* CBS 115975 and *Sclerotinia echinophila* CBS 11548, and previously published genome sequences from the family Sclerotiniaceae. The genomes of *R. sydowiana* CBS 115975, *S. echinophila* CBS 11548, and *S. borealis* F-4157 were generated using next generation sequencing technology; *S. sclerotiorum* 1980 and *B. cinerea* T4 genomes were produced from a Sanger sequencing approach. Not all summary data were available for *S. borealis* F-4157.

<table>
<thead>
<tr>
<th></th>
<th>Rutstroemia sydowiana CBS 115975</th>
<th>Sclerotinia echinophila CBS 11548</th>
<th>Sclerotinia borealis F-4157</th>
<th>Sclerotinia sclerotiorum 1980</th>
<th>Botrytis cinerea T4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>121</td>
<td>101</td>
<td>23</td>
<td>9.1</td>
<td>10</td>
</tr>
<tr>
<td>CEGMA</td>
<td>97.20%</td>
<td>93.60%</td>
<td>99.20%</td>
<td>100.00%</td>
<td>98.40%</td>
</tr>
<tr>
<td>Total sequence length (Mb)</td>
<td>51.9</td>
<td>40.3</td>
<td>39.5</td>
<td>38.3</td>
<td>39.5</td>
</tr>
<tr>
<td>Number of scaffolds</td>
<td>11591</td>
<td>7348</td>
<td>1241</td>
<td>36</td>
<td>118</td>
</tr>
<tr>
<td>Scaffold N50</td>
<td>14947</td>
<td>9536</td>
<td>130819</td>
<td>1630000</td>
<td>562000</td>
</tr>
<tr>
<td>Number of contigs</td>
<td>14655</td>
<td>12842</td>
<td>1741</td>
<td>679</td>
<td>2281</td>
</tr>
<tr>
<td>Contig N50</td>
<td>8570</td>
<td>4918</td>
<td>79787</td>
<td>122550</td>
<td>35000</td>
</tr>
<tr>
<td>GC (%)</td>
<td>43.1</td>
<td>43.1</td>
<td>42</td>
<td>41.8</td>
<td>42.4</td>
</tr>
<tr>
<td>Predicted gene models</td>
<td>17350</td>
<td>12555</td>
<td>10171</td>
<td>14503</td>
<td>10391</td>
</tr>
<tr>
<td>Predicted secreted proteins</td>
<td>1120</td>
<td>880</td>
<td>-</td>
<td>879</td>
<td>630</td>
</tr>
<tr>
<td><strong>Predicted CAZymes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total CAZymes</td>
<td>789</td>
<td>641</td>
<td>-</td>
<td>578</td>
<td>657</td>
</tr>
<tr>
<td>Pectate lyases</td>
<td>10</td>
<td>4</td>
<td>-</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Glycosyltransferases</td>
<td>122</td>
<td>100</td>
<td>-</td>
<td>92</td>
<td>105</td>
</tr>
<tr>
<td>Glycoside hydrolases</td>
<td>316</td>
<td>248</td>
<td>-</td>
<td>227</td>
<td>245</td>
</tr>
<tr>
<td>Carbohydrate esterases</td>
<td>133</td>
<td>120</td>
<td>-</td>
<td>99</td>
<td>130</td>
</tr>
<tr>
<td>Carbohydrate binding motifs</td>
<td>66</td>
<td>64</td>
<td>-</td>
<td>62</td>
<td>71</td>
</tr>
<tr>
<td><strong>Predicted Secondary Metabolite (SM) Clusters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total SM clusters</td>
<td>74</td>
<td>58</td>
<td>-</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>Type I polyketide synthetases (PKSs)</td>
<td>27</td>
<td>20</td>
<td>-</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>Type II PKSs</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Nonribosomal peptide synthetases (NRPSs)</td>
<td>10</td>
<td>7</td>
<td>-</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Terpene clusters</td>
<td>14</td>
<td>6</td>
<td>-</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Hglks</td>
<td>2</td>
<td>4</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Rutstroemia sydowiana* has an abundance of CAZyme modules, relative to the *Sclerotiniaceae* genomes. The general trend of a decrease in CAZymes in saprobe genomes relative to phytopathogen genomes (Zhao et al. 2013) does not hold true for *R. sydowiana*. With 789 CAZymes detected in the *R. sydowiana* genome, it has an average of 26 % more CAZyme modules than *S. sclerotiorum* 1980, *S. echinophila*CBS 111548, and *B. cinerea* T4 genomes. The increase is almost entirely attributable to *R. sydowiana*’s abundance of glycoside hydrolases (GHs), the most diverse group of enzymes used by microbes in the degradation of biomass (Murphy et al. 2011). The high number of GH motifs was attributable to an overall increase across all GH families, not the enrichment of any single motif. The expansion of GH motifs observed from the *R. sydowiana* genome places it amongst the fungi with the largest repertoires of GH modules, including phytopathogens such as *Colletotrichum higginsianum*, *C. graminicola*, *F. oxysporum*, and *Verticillium dahliae* (301–394; Zhao et al. 2013) and saprobes such as *Aspergillus oryzae*, *Gymnopus luxurians*, and *Ganoderma* sp. (294–346; Zhao et al. 2013). Overall, the *R. sydowiana* CAZyme cohort is numerically similar to the cohorts of these motifs in the genomes of *B. cinerea* (T4), *S. sclerotiorum* (1980), and *S. echinophila* (CBS 111548). The generalized reduction of the CAZyme families CE5, GT1, PL1 and PL3 which has been previously detected for saprophytic fungi relative to plant pathogenic fungi (Zhao et al. 2013) was also observed for *R. sydowiana*.

Seventy-four gene clusters putatively involved in the biosynthesis of secondary metabolites (SM) were identified from the *R. sydowiana* genome assembly. *Rutstroemia sydowiana* possesses 22–50 % more SM clusters than *B. cinerea* T4, *S. echinophila* CBS 111548 and *S. borealis* F-4157 genomes. The increased number of SMs is primarily due a greater number of PKS and terpene clusters, relative to those found in *Sclerotoniiaceae* genomes. The expansion of SM clusters in the *R. sydowiana* genome is consistent with the ability of saprophytic fungi to produce a large number of diverse SMs (Collemare & Lebrun 2012), and may impact in the ecological role of this fungus.

Based on the organization of genes present at the mating type locus, *R. sydowiana* is homothallic: both the alpha-domain and high mobility group (HMG) encoding MAT1-
1 and MAT1-2 genes are found at the MAT1 locus, along with MAT1-1-5 and MAT1-2-4. This is the first identification of MAT1-1-5 and MAT1-2-4 orthologs outside of the family Sclerotiniaceae.

The draft genome of *R. sydowiana* presented in this study is the first genome-scale resource for a member of the family Ruststroemiaceae. Together with the genome of *S. echinophila* presented in this paper, it provides a useful resource for comparative analyses of apothecia- and sclerotia-forming saprophytes and phytopathogenic fungi in Helotiales.

Authors: L.A. Beirn, J.E. Demers, and J.A. Crouch

**IMA Genome-F 3G**

**Draft genome sequence of Sclerotinia echinophila**

The genus *Sclerotinia* (*Helotiales, Sclerotiniaceae, Ascomycota*) includes over 250 species of both plant pathogenic and non-pathogenic fungi that thrive in almost every environment (Kohn 1979). The genus is the type of the family Sclerotiniaceae, and includes the causal agents of numerous destructive and economically important plant diseases, such as *S. borealis, S. minor,* and *S. sclerotiorum* causing disease to hundreds of hosts worldwide (Kohn 1979, Amselem et al. 2011, Mardanov et al. 2014). Since *Sclerotinia* was initially described in 1870 (Fuckel 1870), the genus has undergone several major taxonomic revisions. Once broadly defined to include numerous apothecia-forming fungi, Whetzel (1945) restricted the genus to include only those species producing apothecia from tuberoid sclerotia. Attempts to delineate generic and species boundaries have been confounded by the limitation of just a few reliable characters for taxonomic recognition (Kohn 1979). Molecular phylogenetic work using ribosomal DNA sequences indicate that *Sclerotinia* is polyphyletic (Holst-Jensen et al. 1997, 1998), but the genus has not been evaluated by modern multi-locus sequence analysis. In addition, while advances in the biology, genetics, genomics and epidemiology have been made for several plant pathogenic *Sclerotinia*, knowledge of the saprophytic species in this genus is almost non-existent.

The objective of this study was to produce a draft genome sequence assembly and basic genome summary statistics for *S. echinophila* (Fig. 2), a saprophytic *Sclerotinia* species that is most commonly associated with dead cupules and burrs of plants in *Fagaceae.* Together with existing genome resources in *Sclerotiniaceae*, the *S. echinophila* assembly will increase our ability to resolve longstanding questions regarding the evolution, taxonomy and ecological associations exhibited by members of this important group of fungi.

**SEQUENCED STRAIN**


**NUCLEOTIDE SEQUENCE ACCESSION NUMBER**

The Whole Genome Shotgun project of the *Sclerotinia echinophila* (CBS 111548) genome has been deposited in NCBI GenBank under the accession no. JWJA0000000, version JWJA01000000.

**METHODS**

Genomic DNA was isolated using the Omni-Pure Genomic DNA Extraction Kit (G-Biosciences, St Louis, MO) and used to prepare a sequencing library using Illumina Nextera tagmentation chemistry (Illumina, San Diego, CA) for shearing and ligation of adapters and Nextera indices. Quantification and fragment size assessment was performed using a Qubit fluorometer (Life Technologies, Grand Island, NY) and QIAXcel capillary electrophoresis instrument (Qiagen, Germantown, MD). After normalization, the library was sequenced using a paired end cycle on an Illumina MiSeq instrument using a 600-cycle MiSeq sequencing kit (Illumina). Reads were processed and assembled using CLC Genomics Workbench v. 7.0.4 (CLCBio, Germantown, MD). Adapters and indices were removed, and reads were trimmed of low quality sequences (limit 0.05) and runs of ambiguous nucleotides longer than two. Reads <30 nt were discarded. After trimming, 80.1% of the 28.2 million reads remained in pairs. *De novo* assembly of trimmed reads was performed using kmer size n = 51 and automatic bubble size, with contigs <500 nt discarded. Resultant contigs were joined using the CLC Genome Finishing Module by aligning through BLAST searches (kmer = 20, minimum match = 50, maximum e-value = 0.0001). Summary statistics regarding the assembly were calculated using CLC Genomics Workbench and QUAST (Gurevich et al. 2013). *Ab initio* gene predictions were performed using AUGUSTUS v3.0.2 (Stanke et al. 2008) with *Botrytis cinerea* gene models. The completeness of the assembly was assessed using CEGMA v. 2.4 (Parra et al. 2007) through the iPLANT interface (https://de.iplantcollaborative.org/de/). Using the predicted proteins of *Ruststroemia sydowiana* (CBS 115975), *S. echinophila* (CBS 111548), *S. sclerotiorum* (1980) and *Botrytis cinerea* (T4) (Amselem et al. 2011; this study), secondary metabolite clusters were predicted using AntiSMASH (Blin et al. 2013) and carbohydrate-active enzyme (CAZyme) motifs were predicted using dbCAN, including the repertoire of auxiliary enzymes (Yin et al. 2012). Putative secreted proteins and transmembrane domains were predicted using SignalP v. 4.1 (Petersen et. al 2011), and BLASTp searches performed in CLC Genomics (e-value threshold 1E-3).

**RESULTS AND DISCUSSION**

Summary statistics from the draft genome of *Sclerotinia echinophila* (CBS 111548) are presented in Table 1. The assembly of *S. echinophila* (CBS 111548) is contained in 7 348 scaffolds of 500 bp or greater in length for a total size of 40.3 Mb, consistent with genome size of previously sequenced
Sclerotiniaceae. The largest scaffold measures 56.4 Kb, with an average scaffold length of 6571. The GC content is 43.1 %. The total length of the coding sequence is 19.2 Mb, with 12 555 genes covering 47.6% of the 40.3 Mb genome assembly. The predicted number of *S. echinophila* genes is consistent with gene cohorts predicted for other members of Sclerotiniaceae, *B. cinerea*, *B. borealis*, and *S. sclerotiorum* (between 10 171 and 14 503; Anselem et al. 2011, Blanco-Ulate et al. 2013, Mardanov et al. 2014). Mean gene and protein lengths are 1 525 bp and 445 aa, respectively, with an average gene density of 312 genes per Mb. The assembly is estimated as 98.4 % complete, based on the presence of a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes. An average of two introns are found per gene, with an average intron length of 94.7 bp (maximum = 1 616 bp). No introns are predicted in a core set of conserved eukaryotic genes.

At the mating type locus, MAT1, the *S. echinophila* (CBS 111548) genome has an organization typical of a homothallic ascomycete, with the alpha-domain and high mobility group (HMG) encoding MAT1-1 and MAT1-2 genes present at the same locus. Also found at the MAT1 locus are the MAT1-1-5 and MAT1-2-4 genes, which are only known from members of Sclerotiniaceae (Amselem et al. 2011).

From the *S. echinophila* (CBS 111548) predicted proteome of 12 555 genes, 1 359 are predicted to possess transmembrane domains. The genome assembly possesses 880 predicted secreted proteins (31 with transmembrane domains), making up 7 % of the predicted proteome. Using the predicted secretome as BLASTp queries of the combined *B. cinerea*, *S. sclerotiorum*, and *R. sydowiana* (CBS 115975) proteomes, 5.6 % of the predicted secreted proteins were found to be unique *S. echinophila*.

Fifty-eight gene clusters putatively involved in the biosynthesis of secondary metabolites are found in the *S. echinophila* (CBS 111548) genome, and 641 CAZyme motifs. Observed CAZyme profiles are consistent with those reported for *S. sclerotiorum* and *B. cinerea* (Amselem et al. 2011). A reduction of CES, GT1, PL1 and PL3 CAZyme families was evident in the *S. echinophila* genome, a pattern that has also been documented from other saprophytic fungi when compared against plant pathogens (Zhao et al. 2013).

In this study we present a draft genome sequence from a member of *S. echinophila*, the first genome sequence of a saprophytic member of this genus. This sequence provides a unique resource that will facilitate further research into the biology, ecology, and evolution of different lifestyles by fungi in the genus *Sclerotinia*.

**Authors:** L.A. Beirn, J.E. Demers, and J.A. Crouch

**ACKNOWLEDGEMENTS**

The genome sequencing of *Sclerotinia echinophila* and *Rutstroemia sydowiana* was funded by the US Department of Agriculture (USDA) Agricultural Research Service. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. USDA is an equal opportunity provider and employer.

Sequencing the genome of *Amanita jacksonii* was made possible from grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Royal Ontario Museum to J.M.M.; Graduate Scholarships from the Consejo Nacional de Ciencia y Tecnología (Mexico) and the University of Toronto to SSR; and a Undergraduate Student Research Award from NSERC to M.S. We thank Hillary Hatzipetarakos and Tim Tully at Awenda Provincial Park for fieldwork assistance, and Lisa Bukovnik at Duke University for facilitating the genome sequencing. Renée Lebeuf, from the Cercle des Mycologues de Montréal, kindly provided the picture of *Amanita jacksonii*.

For the genomes of *Leptographium procerum*, *Ceratocystis albifundus*, *Fusarium circinatum*, and *Huntiella omanensis* financial support was provided by members of the Tree Protection Cooperative Program (TPCP), the Department of Science and Technology (DST)/National Research Foundation (NRF) Centre of Excellence in Tree Health Biotechnology, and the Genomics Research Institute of the University of Pretoria. This project was supported by multiple grants from the NRF, South Africa, including the grant specific unique reference number (UID) 83924. The grant holders acknowledge that opinions, findings and conclusions or recommendations expressed in publications generated by NRF supported research are that of the authors, and the NRF accepts no liability whatsoever in this regard.

For *Leptographium procerum*, we acknowledge the assistance of Tuan A. Duong (FABI, University of Pretoria) with the preparation of the sample for Illumina sequencing, and Kevin J. Dodds and Garret D. Dubois (USDA Forest Service, USA) for assistance with collections. For *Ceratocystis albifundus* we are grateful to Fourie Joubert for his assistance with the MAKER genome annotation pipeline tool.

**REFERENCES**


Yin M, Duong TA, Wingfield MJ, Zhou XD, de Beer ZW (2015) Taxonomy and phylogeny of the *Leptographium procerum* complex, including *L. sinense* sp. nov. and *L. longiconidiophorum* sp. nov. *Antoni van Leeuwenhoek*.


