doi:10.5598/imafungus.2014.05.02.11

IMA FUNGUS · 5(2): 473–486 (2014)

ARTICLE

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 Nest[1], Lisa A. Beirn[2], Jo Anne Crouch[3], Jill E. Demers[3], Z. Wilhelm de Beer[4], Lieschen De Vos[1], Thomas R. Gordon[5], Jean-Marc Moncalvo[6,7,8], Kershney Naidoo[1], Santiago Sanchez-Ramirez[6], Danielle Roodt[1], Quentin C. Santana[1], Stephanie L. Slinski[1,5], Matt Stata[6], Stephen J. Taerum[1], P. Markus Wilken[1], Andrea M. Wilson[1], Michael J. Wingfield[4], and Brenda D. Wingfield[1]

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

[3]Systematic Mycology and Microbiology Laboratory, US Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705, USA

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

[5]Department of Plant Pathology, University of California, Davis, CA 95616, USA

[6]Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada

[7]Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, ON M5S 2C6, Canada

[8]Centre 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 value resource to researchers studying these fungi.

Key words:
Dispensable chromosome

Article info: Submitted: 3 August 2014; Accepted: 4 December 2014; Published: 16 December 2014.

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

Fig. 1. *Amanita jacksonii* from a population in the province of Québec (Canada). Photo: Renée Lebeuf.

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.

SEQUENCED STRAIN

Canada: *Ontario*: Awenda Provincial Park, N 44.84620, W 079.97507, alt. 222m, on soil in a mixed conifer (*Picea*) broadleaf (*Fagus*, *Acer*, *Betula*, *Quercus*) forest, 27 Aug. 2011, *S. Sánchez-Ramírez & J.-M. Moncalvo* (TRTC168611 – dried basidiome).

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 (pair-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.doe.gov/). For instance, the genome size of *A. muscaria*, *A. thiersii*, *A. brunnescens*, *A. inopinata,* and *A. polypyramis* is 40.7, 33.7, 57.6, 22.1, 23.5 Mbp, respectively (Hess *et al*. 2014). The HMM-based gene prediction found 8 511 structural protein-coding genes, 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 cellulases (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.

Authors: **S. Sánchez-Ramírez, Matt Stata, and J.-M. Moncalvo**

IMA Genome-F 3B

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. fimbriata* (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

South Africa: *Limpopo*: Kruger National Park, isol. ex *Terminalia sericea*, March 2005, *J. Roux* (CMW17620, CBS 138876, CBS-H 61112 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The Whole Genome Shotgun project of this *Ceratocystis albifundus* isolate has been deposited at DBJ/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 IIx 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 poor quality (limit of 0.05) as well as terminal nucleotides. The remaining reads were assembled using the *de novo* genome assembler Velvet (Zerbino & Birney 2008) with an optimized k-mer value of 75. Thereafter, scaffolding was completed using SSPACE v. 2.0 and gaps reduced with the use of GapFiller v. 2.2.1 (Boetzer *et al.* 2011, Boetzer & Pirovano 2012). The completeness of the assembly was evaluated using the Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra *et al.* 2007). The automated genome annotation pipeline tool, MAKER, was trained and used to structurally annotate the assembly (Cantarel *et al.* 2008). This tool includes steps for the masking of repetitive elements, *ab-initio* gene predictions using SNAP, AUGUSTUS and GeneMark, protein information from related organisms with the use of BLASTx and Protein2Genome and further refinement of intron-exon boundaries with the use of Exonerate (Smit *et al.* 1996, Stanke *et al.* 2006). Manual curation on a subset (1 458) of genes predicted by MAKER by incorporating all the above mentioned elements was performed for manual verification of start and stop codons, intron-exon boundaries and overall gene structure.

RESULTS AND DISCUSSION

The genome of *Ceratocystis albifundus* had an estimated size of 27 149 029 bases with an average average coverage of 24×. The N50 size was 58 335 bases, and the assembly had a mean GC content of 48.6 %. The total number of contigs generated was 1 958, with 939 contigs larger than 1 000 nucleotides in size. The assembly had a CEGMA completeness score of 96.4 %, indicating that most of the core eukaryotic genes were present. MAKER predicted and structurally annotated a total of 6 967 genes after training, at a gene density of 257 genes/Mb.

The draft genome of *C. albifundus* is smaller than that of the type species of the genus, *C. fimbriata,* and also of *C. manginecans,* that are 29.4 Mb and 31.7 Mb, respectively (Wilken *et al.* 2013, van der Nest *et al.*, 2014). The genome is closer in size to that of the saprobic *Huntiella moniliformis*, which has a genome size of 25.4 Mb (van der Nest *et al.* 2014). *Ceratocystis albifundus* also has a similar number of putative genes to that of *H. moniliformis* (6 832 predicted ORFs) than to the more closely related *C. fimbriata* (7 266 predicted ORFs) and *C. manginecans* (7 404 predicted ORFs). This could indicate that the additional predicted genes in *C. fimbriata* and *C. manginecans* may not be associated with pathogenicity as might have been expected prior to the assembly of this genome (van der Nest *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 *Fusarum* spp.

SEQUENCED STRAIN

USA: *California*: laboratory strain, Aug. 2009*, S.L. Slinski* (CMW41611, CBS138821, GL1327, PREM61154 - dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The *Fusarium circinatum* genomic sequence has been deposited at DDBJ/EMBL/GenBank under the accession JRVE00000000. The version described in this paper is the first version, JRVE01000000.

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* gene models and cDNA data from the *F. circinatum* genome (Wingfield *et al.* 2012).

RESULTS AND DISCUSSION

Assembly of the draft genome for the laboratory strain (GL1327) of *Fusarium circinatum* yielded a genome size of 42 540 497 bp with an average coverage of 408×. The assembly generated 909 contigs of size greater than 200 bp, has an N50 of 372 559 bp and an average scaffold size of 46 799 bp. The largest scaffold was 1 475 703 bp in size. The GC content is 48.2 %. Based on the occurrence of a core set of conserved eukaryotic genes, the assembly is 97.99 % complete (Parra *et al.* 2008). The assembly was predicted to contain 14 314 putative ORFs with an average length of 1 455 bp and an average density of 336 ORFS/Mb. In comparison, the *F. circinatum* strain Fsp34 sequenced by Wingfield *et al*. (2012) has a larger genome (44.3 Mb) with 708 more putative ORFs and a comparable average density of 339 ORFS/Mb.

Members of the *F. fujikuroi* complex are known to possess 12 chromosomes (Xu *et al.* 1995). Sequence comparisons confirmed that chromosome 12 has been lost in the strain of *F. circinatum* sequenced in this study. This was evident when BLAST analyses done against the *F. fujikuroi* chromosome 12 failed to identify similar sequences (Wiemann *et al*. 2013). This confirmed PFGE results (Slinski *et al*., unpubl.) showing that this chromosome has been lost in the laboratory strain.

Chromosome 12 has been shown to be the smallest of the chromosomes found in species within the *F. fujikuroi* complex (Xu *et al.* 1995). These vary significantly in size intra- and interspecifically, displaying chromosome length polymorphism, in comparison to the other chromosomes (Xu *et al*. 1995). They have been found to be strain-specific in members of the *F. fujikuroi* complex (Wiemann *et al*. 2013). Chromosome 12 also has the lowest sequence similarity between species (Xu *et al.* 1995). .Furthermore, these chromosomes can be lost (Xu *et al.* 1995, Jurgenson *et al*. 2002, Ma *et al.* 2010). The presence of accessory chromosomes in the genus *Fusarium* has been well-documented (Coleman *et al.* 2009, Ma *et al*. 2010, Croll & 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.

Authors: **L. De Vos, S.L. Slinski, Q.C. Santana, M.J. Wingfield, T.R. Gordon, and B.D. Wingfield**

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. bhutanenesis*, 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* (Wilken *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.

SEQUENCED STRAIN

Oman: Al-Batinah region isol. ex *Mangifera indica,* May 2003, *A.O. Al Adawi*. (CMW 11056 = CBS 118113; CBS H-61113 – dried culture).

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

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, N50 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 retained once contigs of less than 1 Kb were filtered out. The draft assembly had a CEGMA completeness score of 90.83 % for complete genes and 95.97 % for partial genes. The final assembly was predicted to encode 8 395 putative ORFs at a density of 266 ORFs/Mb.

The *H. omanensis* draft genome was larger than that of its close relative, *H. moniliformis*. The genome of the latter species was reported to be 25 Mb in size and encodes less than 7000 ORFs (van der Nest *et al.* 2014). The draft genome of *H. omanensis* was, however, similar in size to two pathogenic species of *Ceratocystis*, namely *C. manginecans* (31.7 Mb, 7 494 ORFs; van der Nest *et al.* 2014) and *C. fimbriata* (29.4 Mb, 7 266 ORFs; Wilken *et al.* 2013), but *H. omanensis* encodes a larger number of putative ORFs. The availability of this genome sequence will be invaluable in increasing our knowledge and understanding the biology of this saprobic fungus. The genome will allow for future comparative genomic studies within this group of fungi, and with species in the greater *Ceratocystidaceae* family.

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

IMA Genome-F 3E

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-killing behaviour of *D. valens* in China. This is because pine trees native to China may produce larger quantities of monoterpenes 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 × 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 est-imate 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 32×, 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 *Grosmannia 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 Netherlands: *Prov. Utrecht*: Soest, De Stompert, Oct. 2002, *G. Verkley* (CBS 115975; dried culture: BPI 892981).

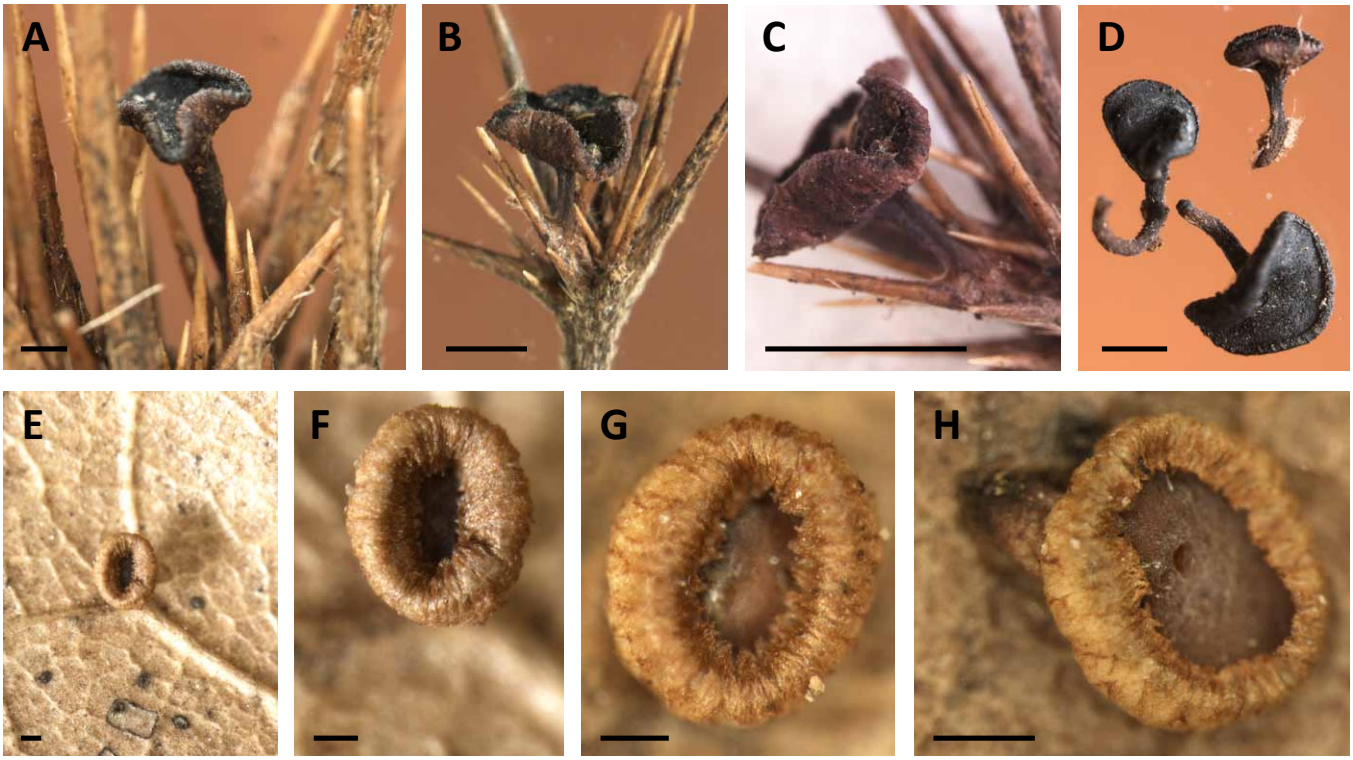

Fig. 2. Apothecia of *Sclerotinia echinophila* and *Rutstroemia sydowiana*. **A–D.** *Sclerotinia echinophila* on burrs of *Castanea sativa* (BPI 737886). **E–H.** *Rutstroemia sydowiana* on leaves of *Quercus rubra* (BPI 653856). Bars = 1 mm. Photo: Jo Anne Crouch.

NUCLEOTIDE SEQUENCE ACCESSION NUMBER

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 97.7bp (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. sclerotium* (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 *Sclerotineaceae*. 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.

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

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. sclerotiorum* 1980 genomes. The increased number of SMs is primarily due a greater number of PKS and terpene clusters, relative to those found in *Sclerotiniaceae* genomes. The expansion of SM clusters in the *R. sydowiana* 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 *Sclerotiniacea*.

The draft genome of *R. sydowiana* presented in this study is the first genome-scale resource for a member of the family *Rutstroemiaceae*. 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 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. echiniphila* 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

The Netherlands: *Prov. Limburg*: Bunderbos, isol. ex dead cupules of *Castanea sativa*, 2002, *G. Verkley* (CBS 111548; dried culture: BPI 892982).

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. JWJA00000000, 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 *Rutstroemia sydowiana* (CBS 115975), *S. echinophila* (CBS 111548), *S. sclerotiorium* (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*, *S. 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 19.1 % of the 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* (Anselem *et al*. 2011). A reduction of CE5, 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 Tecnologia (Mexico) and the University of Toronto to SSR; and a Undergraduate Student Research Award from NSERC to M.S. We thank Hillary Hatzipetrakos 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

Alexander SA, Horner WE, Lewis KJ (1988) *Leptographium procerum* as a pathogen of pines. In: Leptographium *Root Diseases on Conifers* (TC Harrington, FW Cobb, eds.): 97–112. St Paul, MN: American Phytopathological Society Press.

Al-Subhi, AM, Al-Adawi AO, van Wyk M, Deadman ML, Wingfield (2006) *Ceratocystis omanensis*, a new species from diseased mango trees in Oman. *Mycological Research* **110**: 237–245.

Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, *et al.* (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics* **7**: e1002230.

Barnes I, Nakabonge G, Roux J, Wingfield BD, Wingfield MJ (2005) Comparison of populations of the wilt pathogen *Ceratocystis albifundus* in South Africa and Uganda. *Plant Pathology* **54**: 189–195.

Blanco-Ulate B, Allen G, Powell ALT, Cantu D (2013) Draft genome sequence of *Botrytis cinerea* BcDW1, inoculum for noble rot of grape berries. *Genome Announcements* **1**: e00252-13.

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

Boa ER (2004) *Wild Edible Fungi: a global overview of their use and importance to people*. [Non-wood Forest Products, vol. 17.] Rome: Food & Agriculture Organization.

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.

Boisvert S, Laviolette F, Corbeil J (2010) Ray: simultaneous assembly of reads from a mix of high– throughput sequencing technologies. *Journal of Computational Biology* **17**: 1519–1533.

Buermans HPJ, den Dunnen JT (2014) Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta* **1842**: 1932–1941.

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.

Carbone, I, Kohn LM (1993) Ribosomal DNA sequencing divergence within internal transcribed spacer 1 of the *Sclerotiniaceae*. *Mycologia* **85**: 415–427.

Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, *et al.* (2009) The genome of *Nectria haematococca:* Contribution of supernumerary chromosomes to gene expansion. *PLoS Genetics* **5**: e10000618.

Collemare J, Lebrun M-H (2012) Fungal secondary metabolites: ancient toxins and novel effectors in plant-microbe interactions. In: *Effectors in Plant-Microbe Interactions* (Martin F, Kamoun S, eds): 377–399. UK: John Wiley & Sons.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Croll D, McDonald BA (2012) The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens* **8**: e1002608.

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* (KA Seifert, ZW de Beer, MJ Wingfield, eds): 21–46. [CBS Biodiversity Series no. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

DiGuistini S, Wang Y, Liao NY, Taylor G, Tanguay P, *et al* (2011) Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen. *Proceeding 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.

Fuckel L (1870) ["1869"] Symbolae mycological: Beiträge zur Kenntnis der Rheinischen Pilze. *Jahrbücher des Nassauischen Vereins für Naturkunde* **23–24**: 1–459.

Geiser DM, Aoki T, Bacon CW, Baker SE, Bhattacharayya MB, *et al.* (2013) One Fungus, One Name: defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology* **103**: 400–408.

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

Hallen HE, Luo H, Scott-Craig JS, Walton JD (2007) Gene family encoding the major toxins of lethal *Amanita* mushrooms. *Proceeding of the National Academy of Sciences, USA* **104**: 19097–19101.

Heath RN, Wingfield MJ, van Wyk M, Roux J (2009) Insect associates of *Ceratocystis albifundus* and patterns of association in a native savanna ecosystem in South Africa. *Environmental Entomology* **38**: 356–364.

Hepting GH, Roth ER (1946) Pitch canker, a new disease of some southern pines. *Journal of Forestry* **44**: 742–744.

Hess J, Skrede I, Wolfe BE, LaButti K, Ohm RA, *et al.* (2014) Transposable element dynamics among asymbiotic and ectomycorrhizal *Amanita* fungi. *Genome Biology and Evolution* **6**: 1564–1578.

Hinz UG, Fivaz J, Girod PA, Zyrd JP (1997) The gene coding for the DOPA dioxygenase involved in betalain biosynthesis in *Amanita muscaria* and its regulation. *Molecular and General Genetics* **256**: 1–6.

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

Holst-Jensen A, Kohn, LM, Schumacher T (1997) Nuclear rDNA phylogeny of the *Sclerotiniaceae*. *Mycologia* **89**: 885–899.

Holst-Jensen A, Vaage, M, Schumacher T (1998) Approximation to the phylogeny of *Sclerotinia* and related genera. *Nordic Journal of Botany* **18**: 705–719.

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

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

Jacobs K, Wingfield MJ, Wingfield BD (2001) Phylogenetic relationships in *Leptographium* based on morphological and molecular characters. *Canadian Journal of Botany* **79**: 719–732.

Jeong H, Lee S, Choi GJ, Lee T, Yun S-H (2013) Draft genome sequence of *Fusarium fujikuroi* B14, the causal agent of the bakanae disease of rice. *Genome Announcements* **1**: e00035-13.

Johnston PJ, Park D (2013) The phylogenetic position of *Lanzia berggrenii* and its sister species. *Mycostema* **32**: 366–385.

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

Klepzig KD, Raffa KF, Smalley EB (1991) Association of an insect-fungal complex with red pine decline in Wisconsin. *Forest Science* **37**: 1119–1139.

Kohn LM (1979) Delimitation of the economically important plant pathogenic *Sclerotinia* species. *Phytopathology* **69**: 881–886.

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.

Lackner AL, Alexander SA (1982) Occurrence and pathogenicity of *Verticicladiella procera* in Christmas tree plantations in Virginia. *Plant Disease* **66**: 211–212.

Lu M, Wingfield MJ, Gillette NE, Mori SR, Sun JH (2010) Complex interactions among host pines and fungi vectored by an invasive bark beetle. *New Phytologist* **187**: 859–866.

Lu M, Wingfield MJ, Gillette NE, Sun JH (2011) Do novel genotypes drive the success of an invasive bark beetle-fungus complex? Implications for potential reinvasion. *Ecology* **92**: 2013–2019.

Lu M, Zhou XD, de Beer ZW, Wingfield MJ, Sun JH (2009a) Ophiostomatoid fungi associated with the invasive pine-infesting bark beetle, *Dendroctonus valens*, in China. *Fungal Diversity* **38**: 133–145.

Lu Q, Decock C, Zhang XY, Maraite H (2009b) Ophiostomatoid fungi (*Ascomycota*) associated with *Pinus tabuliformis* infested by *Dendroctonus valens* (*Coleoptera*) in northern China and an assessment of their pathogenicity on mature trees. *Antonie Van Leeuwenhoek* **96**: 275–293.

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

Mardanov AV, Beletsky AV, Kadnikov VV, Ignatov AN, Ravin NV (2014) Draft genome sequence of *Sclerotinia borealis*, a psychrophilic plant pathogenic fungus. *Genome Announcements* **2**: e01175-13.

Martin F, Cullen D, Hibbett D, Pisabarro A, Spatafora JW, *et al.* (2011) Sequencing the fungal tree of life. *New Phytologist* **190**: 818–821.

Masuya H, Yamaoka Y, Wingfield MJ (2013) Ophiostomatoid fungi and their associations with bark beetles in Japan. In: *The Ophiostomatoid Fungi: expanding frontiers* (Seifert KA, de Beer ZW, Wingfield MJ, eds): 77–90. [CBS Biodiversity Series no. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

Metzker ML (2009) Sequencing technologies—the next generation. *Nature Reviews Genetics* **11**: 31–46.

Möller EM, Bahnweg G, Sandermann H, Geiger HH (1992) A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies and infected plant tissues. *Nucleic Acids Research* **20**: 6115–6116.

Moreno-Hagelsieb G, Latimer K (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **24**: 319–324.

Morris MJ, Wingfield MJ, de Beer C (1993) Gummosis and wilt of *Acacia mearnsii* in South Africa caused by *Ceratocystis fimbriata*. *Plant Pathology* **42**: 814–817.

Murphy C, Powlowski J, Wu M, Butler G, Tsang A (2011) Curation of characterized glycoside hydrolases of fungal origin. *Database* doi:10.1093/database/bar020.

Ojeda DI, Dhillon B, Tsui CKM, Hamelin RC (2014) Single-nucleotide polymorphism discovery in *Leptographium longiclavatum*, a mountain pine beetle-associated symbiotic fungus, using whole-genome resequencing. *Molecular Ecology Resources* **14**: 401–410.

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 (2008) Assessing the gene space in draft genomes. *Nucleic Acids Research* **37**: 289–297.

Pegler DN (2002) Useful fungi of the world: Caesar's mushroom and the Christmas mushroom. *Mycologist* **16**: 140–141.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785–786.

Reay SD, Walsh PJ, Ram A, Farrell RL (2002) The invasion of *Pinus radiata* seedlings by sapstain fungi, following attack by the Black Pine Bark Beetle, *Hylastes ater* (*Coleoptera*: *Scolytidae*). *Forest Ecology and Management* **165**: 47–56.

Roux J, Dunlop R, Wingfield MJ (1999) Susceptibility of elite *Acacia mearnsii* families to Ceratocystis wilt in south Africa. *Journal of Forest Research* **4**: 187–190.

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

Roux J, Wingfield MJ (2013) *Ceratocystis* species on the African continent, with particular reference to *C. albifundus*, an African species in the *C. fimbriata sensu lato* species complex, In: *Ophiostomatoid Fungi: expanding frontiers* (Seifert KA, de Beer ZW, Wingfield MJ, eds): 131–138. [CBS Biodiversity Series no. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

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

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

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

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

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

Strack D, Vogt T, Schliemann W (2003) Recent advances in betalain research. *Phytochemistry* **62**: 247–269.

Sun JH, Lu M, Gillette NE, Wingfield MJ (2013) Red turpentine beetle: innocuous native becomes invasive tree killer in China. *Annual Reviews of Entomology* **58**: 293–311.

Taerum SJ, Duong TA, de Beer ZW, Gillette NE, Sun JH, *et al.* (2013) Large shift in symbiont assemblage in the invasive red turpentine beetle. *PLoS ONE* **8**: e78126.

Taylor AFS, Alexander I (2005) The ectomycorrhizal symbiosis: life in the real world. *Mycologist* **19**: 102–112.

van der Nest MA, Bihon W, de Vos L, Naidoo K, Roodt D *et al.* (2014) Draft genome sequences of *Diplodia sapinea*, *Ceratocystis manginecans*, and *Ceratocystis moniliformis*. *IMA Fungus* **5**: 135–140.

van Wyk M, Al Adawi AO, Khan IA, Deadman ML, Al Jahwari AA, *et al* (2007) *Ceratocystis manginecans* sp. nov., causal agent of a destructive mango wilt disease in Oman and Pakistan. *Fungal Diversity* **27**: 213–230.

van Wyk, M, Roux J, Barnes I, Wingfield BD, Chhetri DB, *et al.* (2004) *Ceratocystis bhutanensis* sp. nov., associated with the bark beetle *Ips schmutzenhoferi* on *Picea spinulosa* in Bhutan. *Studies in Mycology* **50**: 365–379.

Vetter J (1998) Toxins of *Amanita phalloides*. *Toxicon* **36**: 13–24.

Whetzel HH (1945) A synopsis of the genera and species of the *Sclerotiniaceae*, a family of stromatic inoperculate discomycetes. *Mycologia* **37**: 648–714.

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

Wingfield BD, Steenkamp ET, Santana QC, Coetzee MPA, Bam S, *et al.* (2012) First fungal genome sequence from Africa: A

preliminary analysis. *South African Journal of Science* **108**: 1–2.

Wingfield BD, van Wyk M, Roos H, Wingfield MJ (2013b) *Ceratocystis*: emerging evidence for discrete generic boundaries In: *Ophiostomatoid Fungi: expanding frontiers* (Seifert KA, de Beer ZW, Wingfield MJ, eds): 57–64. [CBS Biodiversity Series no. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

Wingfield MJ, de Beer C, Visser C, Wingfield BD (1996) A new *Ceratocystis* species defined using morphological and ribosomal DNA sequence comparisons. *Systematic and Applied Microbiology* **19**: 191–202.

Wingfield MJ, Capretti P, MacKenzie M (1988) *Leptographium* spp. as root pathogens of conifers. An international perspective. In: *Leptographium root diseases on conifers* (Harrington TC, Cobb FW, eds.): 113–128. St Paul, MN: American Phytopathological Society Press.

Wingfield MJ, Roux J, Wingfield BD, Slippers B (2013a) *Ceratocystis* and *Ophiostoma*: international spread, new associations and plant health. In: *Ophiostomatoid Fungi: expanding frontiers* (Seifert KA, de Beer ZW, Wingfield MJ, eds): 191–200. [CBS Biodiversity Series no. 12.] Utrecht: CBS-KNAW Fungal Biodiversity Centre.

Wolfe BE, Tulloss RE, Pringle A (2012) The irreversible loss of a decomposition pathway marks the single origin of an ectomycorrhizal symbiosis. *PLoS ONE* **7**: e39597.

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

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. *Antonie van Leeuwenhoek*.

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

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

Zhao Z, Liu H, Wang C, Xu J-R (2013) Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **14**: 274.

Zolan ME, Pukkila PJ (1986) Inheritance of DNA methylation in *Coprinus cinereus*. *Molecular and Cellular Biology* **6**: 195–200.