IMA Genome-F 6

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

Brenda D. Wingfield1, Jon M. Ambler1, Martin P.A. Coetzee1, Z. Wilhelm de Beer2, Tuan A. Duong1, Fourie Joubert1, Almuth Hammerbacher2, Alistair R. McTaggart2, Kershney Naidoo1, Hai D.T. Nguyen4,5, Ekaterina Ponomareva4, Quentin S. Santana1, Keith A. Seifert1, Emma T. Steenkamp2, Conrad Trollip1, Magriet A. van der Nest1, Cobus M. Visagie4,5, P. Markus Wilken1, Michael J. Wingfield1, and Neriman Yilmaz4,6

1Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, 0028 South Africa; corresponding author: brenda.wingfield@fabi.up.ac.za
2Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, 0028 South Africa
3Centre for Bioinformatics and Computational Biology, Department of Biochemistry and Genomics Research Institute, University of Pretoria, Private Bag X20, Pretoria 0028, South Africa
4Biodiversity (Mycology), Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, 960 Carling Ave., Ottawa, Ontario, K1A 0C6, Canada
5Department of Biology, University of Ottawa, 30 Marie-Curie, Ottawa, Ontario, K1N6N5, Canada
6Department of Chemistry, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6, Canada

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

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

IMA Genome-F 6A

Draft genome sequence of the root rot fungus Armillaria fuscipes

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

Armillaria root rot disease in the Southern hemisphere is caused by several species, including A. fuscipes (Gregory et al. 1991, Coetzee et al. 2000). The species is restricted to the African continent where it affects the health of agronomic and timber plantations. In South Africa, Armillaria root rot disease was first reported during the early 1900s (Bottomley 1937), but A. fuscipes was only identified as the causal agent of root...
rot disease on Pinus species in 2000 (Coetzee et al. 2000). Subsequent studies showed that this pathogen also affects the health of other tree and woody plant species elsewhere in Africa (Mohammed & Guillaumin 1993, Mwenje et al. 2003, Gezahgne et al. 2004). Although research has been done to characterise African Armillaria species at the molecular level in terms of enzymes (e.g. Agustian et al. 1994, Mwenje & Ride 1996, 1997, 1999) - and genomic regions used for taxonomic purposes (e.g. Chillali et al. 1997, Pérez-Sierra et al. 2004, Coetzee et al. 2005), nothing is yet known regarding the genome of A. fuscipes. The aim of this study was, therefore, to assemble a draft genome of this fungus with the objective of using the genome in future comparative genome studies.

SEQUENCED STRAIN


NUCLEOTIDE SEQUENCE ACCESSION NUMBER

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

METHODS

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

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

RESULTS AND DISCUSSION

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

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

The genome assembly of A. fuscipes, as well as those of the other Armillaria species, were compared to the publicly available assembled genomes of other members of the Physalacriaceae, namely Cylindrobasidium torrendii (Floudas et al. 2015) and Flammulina velutipes (Park et al. 2014) (Table 1). Cylindrobasidium torrendii and F. velutipes were characterised with smaller genomes than the Armillaria species, but the number of protein coding genes, average exon length, and average number of exons per gene ranges were comparable (Table 1). The genome of F. velutipes deviated from the other fungal genomes included in this study by having an average gene length that is much larger (2.294 in comparison to 1350–1655 bp) and an exceptional average intron length (180 in comparison to the 59–73.62 bp for the other species).

The genome sequence of A. fuscipes determined in this study is considered a high quality draft genome by virtue of the number of genes that were identified using the CEGMA pipeline and BUSCO analyses as well as the correlation in gene characteristics with other Armillaria species and members of
the family Physalacriaceae. Armillaria fuscipes forms a basal taxon in the phylogeny of Armillaria species (Coetzee et al. 2011) and its genome sequence therefore provides a crucial resource for future comparative genome studies.

**Authors:** M.P.A. Coetzee*, J.M. Ambler, F. Joubert, Q.S. Santana, and E.T. Steenkamp

*Contact: Martin.Coetzee@fabi.up.ac.za

IMA Genome-F 6B

**Draft genome sequence of Ceratocystisps minuta**

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

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

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

<table>
<thead>
<tr>
<th></th>
<th><em>A. fuscipes</em></th>
<th><em>A. mellea</em></th>
<th><em>A. solidipes</em></th>
<th><em>A. gallica</em></th>
<th><em>C. torrendii</em></th>
<th><em>F. velutipes</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome Assembly (Mb)</td>
<td>53.00</td>
<td>58.30</td>
<td>58.01</td>
<td>85.34</td>
<td>31.57</td>
<td>35.6</td>
</tr>
<tr>
<td>Number of protein-coding genes</td>
<td>14 515</td>
<td>14 473</td>
<td>20 811</td>
<td>25 704</td>
<td>13 940</td>
<td>12 218</td>
</tr>
<tr>
<td>Average gene length (bp)</td>
<td>1 350</td>
<td>1 575</td>
<td>1 580</td>
<td>1 459</td>
<td>1 655</td>
<td>2 294</td>
</tr>
<tr>
<td>Average exon length (bp)</td>
<td>211.75</td>
<td>217.52</td>
<td>227</td>
<td>217</td>
<td>274</td>
<td>245.7</td>
</tr>
<tr>
<td>Average intron length (bp)</td>
<td>68.94</td>
<td>73.62</td>
<td>64</td>
<td>68</td>
<td>59</td>
<td>180.4</td>
</tr>
<tr>
<td>Average number of exons per gene</td>
<td>6</td>
<td>6.02</td>
<td>5.69</td>
<td>5.40</td>
<td>5.17</td>
<td>5.81</td>
</tr>
</tbody>
</table>

**SEQUENCED STRAIN**


**NUCLEOTIDE SEQUENCE ACCESsION NUMBER**

The genomic sequence of *Ceratocystisps minuta* (CMW4352, CBS138717) has been deposited at DDBJ/EMBL/GenBank under the accession L2PB0000000. The version described in this paper is version L2PB001000000.

**METHODS**

A single hyphal-tip culture was grown in 2 % malt extract and 0.5 % yeast extract broth at 25 °C, shaking at 150 rpm for several days. Mycelium was harvested and subjected to lyophilization. Genomic DNA was extracted from lyophilized mycelium using the method described by Aljanabi & Martinez (1997) with additional modification steps (Duong et al. 2013). The methods of genome sequencing, assembly and annotation were similar to those used for *Leptographium lundbergii* (Wingfield et al. 2015a). Two pair-end libraries (350 bp and 530 bp average insert size) were prepared and sequenced using the Illumina HiSeq 2000 platform. Obtained
reads were subjected to quality filtering and trimming and assembled using CLC Genomics Workbench v. 8.0.1 (CLCBio, Aarhus). The number of protein coding genes were predicted using the MAKER genome annotation pipeline (Cantarel et al. 2008) following similar steps as for L. Lundbergii (Wingfield et al. 2015a). The quality and completeness of the assembly were estimated using BUSCO (Simão et al. 2015).

RESULTS AND DISCUSSION

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

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

*Contact:* tuan.duong@fabi.up.ac.za

**IMA Genome-F 6C**

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

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

The aim of this study was to produce a draft nuclear genome sequence for C. adiposa. Best known as the causal agent of root rot on sugarcane (Butler 1906), this species has also been reported from both Prunus (Paulin-Mahady et al. 2002) and Pinus (Talbot 1996) species. Since its first description as Ceratostomella adiposum (Sartoris 1927), this species has also been treated as Ophiostoma adiposum and E. adiposa, and in 1981 C. adiposa was also suggested to be synonymous with C. major (de Beer et al. 2014). This long and confused taxonomic history has yet to resolve the exact placement of C. adiposa within Ceratocystidaceae (de Beer et al. 2014). Therefore, apart from adding to the Ceratocystidaceae genome resource, the availability of a genome sequence for this species could also provide useful tools to address its taxonomy.

**SEQUENCED STRAIN**

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

**NUCLEOTIDE SEQUENCE ACCESSION NUMBER**

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

**METHODS**

*Ceratocystis adiposa* isolate CMW 2574 was obtained from the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute of the University of Pretoria and grown on 2 % malt extract agar (20 g ME and 20 g Agar, Biolab, South Africa) supplemented with 100 μg/L thiamine for 2 wk. The plates were subjected to a previously described DNA isolation procedure (Roux et al. 2004), before the purified nucleic acid was sequenced at the Genome Centre, University of California at Davis (CA) using a Genome Analyzer Ix platform (Illumina). For sequencing, paired-end libraries with approximately 350 and 600 base insert sizes were produced, and used to produce reads with an average length of 100 bases. Poor-quality reads and/or terminal
nucleotides were discarded using the software package CLC Genomics Workbench v. 8.5.1 (Qiagen, Aarhus). A draft genome sequence was produced using the de novo assembly option in CLC Genomics Workbench under default options. The contigs produced were subjected to scaffolding using SSPACE and to fill the resultant gaps using GapFiller (Boetzer et al. 2011, Boetzer & Pirovano 2012). Putative open reading frames (ORFs) were predicted using the online version of the de novo prediction software AUGUSTUS based on the gene models for Fusarium graminearum (Stanke et al. 2006), while genome completeness was assessed with the Benchmarking Universal Single-Copy Orthologs (BUSCO) software on all contigs greater than 500 bases in size (Simão et al. 2015).

RESULTS AND DISCUSSION

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

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

*Contact: Markus.Wilken@fabi.up.ac.za

IMA Genome-F 6D

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

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

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

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

SEQUENCED STRAIN


NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The draft genome sequence of Endoconidiophora laricicola (CMW 20928) and E. polonica (CMW 20930) have been deposited in DDBJ/ENA/GenBank under the accession numbers LXGT0000000 and LXXK0000000, respectively. The versions described in this paper are the first version of each genome, versions LXGT0100000 and LXXK0100000.
METHODS

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

RESULTS AND DISCUSSION

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

Based on size, gene number and GC content, the *E. laricicola* (33.3 Mb, 6 897 ORFs, 45 % GC) and *E. polonica* (33.5 Mb, 6 792 ORFs, 45 % GC) draft genomes resembled those of other species in the Ceratocystidaceae. These include *Davidsoniella virescens* (33.7 Mb, 6 953 ORFs, 44.5 % GC), *C. fimбриata* (29.4 Mb, 7 266 ORFs, 48 % GC), *C. eucalypticola* (31.3 Mb, 7 353 ORFs, 47.9 % GC), *C. manginecans* (31.7 Mb, 7 494 ORFs, 47.9 % GC), *C. platani* (29.2 Mb, 5 963 ORF, 48.2 % GC) and *C. albifundus* (27.2 Mb, 6 967 ORFs, 48.6 % GC) that infect eudicots. This similarity also extended to *Thielaviopsis musarum* (28.5 Mb, 6 963 ORF, 49.2 % GC) and *T. punctualis* (28.1 Mb, 5 480 ORFs, 48.3 % GC) that infect monocots, as well as the non-pathogens *H. moniliformis* (25 Mb, 6 832 ORF, 48 % GC) and *H. omanensis* (31.5 Mb, 8 359 ORFs eudicots, 47.6 % GC) (Wilken et al. 2013, van der Nest et al. 2014a, b, Wingfield et al. 2015a, b, Belbahri 2015).

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

Authors: M.A. van der Nest*, T.A. Duong, P.M. Wilken, K. Naidoo, A. Hammerbacher, E.T. Steenkamp, A.R. McTaggart, M.J. Wingfield, and B.D. Wingfield

*Contact: magriet.vandernest@fabi.up.ac.za

IMA Genome-F 6E

Draft genome sequence of *Penicillium freii* DAOMC 242723

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

SEQUENCED STRAIN

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

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

*Penicillium freii* DAOMC 242723 was grown on Blakeslee’s malt extract agar for 7 d at 25 °C (Visagie et al. 2014) after which a spore suspension was prepared by flooding the colony with 5 mL sterile dH₂O. One mL spore suspension was used to inoculate 100 mL Blakeslee’s malt extract broth in shaking flasks incubated at 300 rpm at 25 °C for 6 d. Fungal material was collected by filtration and DNA extracted with the OmniPrep kit for fungi (G-Biosciences) following the manufacturer’s protocol. Whole-genome sequencing (paired-end with 101 base pairs, bp) was performed on an Illumina HiSeq 2500 with TrueSeq V3 chemistry at the National Research Council Canada in Saskatoon, Saskatchewan, Canada.

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

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

**RESULTS AND DISCUSSION**

Illumina sequencing of 32 million reads of *Penicillium freii* DAOMC 242723 represented 3.2 Gbp of data, which were assembled into 1923 scaffolds. The whole assembly was 33.5 Mbp with a GC content of 47.4 %. The following statistics for the assembly were obtained: N50 was 61.8 Kbp; the longest scaffold was 325 Kbp; the median nucleotide coverage across the whole assembly was 78.5x. Assessment of the completeness of the genome using BUSCO groups for fungi resulted in values of C: 98%[D:6.9%], F:1.1%, M:0.0% ,n:1438 (C:complete [D:duplicated], F:fragmented, M:missed, n:genes) whereas scores of 95.6 % and 97.6 % were obtained from the complete and partial gene set respectively using CEGMA. After annotation and validation, 11 739 predicted gene models were obtained, with 11 418 complete (97.2 %) and 320 lacking a start codon, stop codon or both (2.7 %). Mean gene length was 1489 bp, mean exon length was 483 bp and mean intron length was 79 bp. This draft genome of *P. freii* represents a useful resource for comparative genomic studies and in particular should be useful for those interested in polyketide and other mycotoxin synthesis in *Penicillium* and related genera.

Authors: H.D.T. Nguyen*, C.M. Visagie, N. Yilmaz, E. Ponomareva, and K.A. Seifert

*Contact: hai.nguyen.1984@gmail.com*

**ACKNOWLEDGEMENTS**

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

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

**REFERENCES**


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


