IMA Genome-F 7

Draft genome sequences for Ceratocystis fagacearum, C. harringtonii, Grosmannia penicillata, and Huntiella bhutanensis

Brenda D. Wingfield1, Tuan A. Duong1, Almuth Hammerbacher2, Magriet A. van der Nest1, Andi Wilson1, Runlei Chang2, Z. Wilhelm de Beer2, Emma T. Steenkamp2, P. Markus Wilken1, Kershney Naidoo1, and Michael J. Wingfield2

1Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa; corresponding author e-mail: Brenda.Wingfield@up.ac.za
2Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa

Abstract: Draft genomes for the fungi Ceratocystis fagacearum, C. harringtonii, Grosmannia penicillata, and Huntiella bhutanensis are presented. Ceratocystis fagacearum is a major causal agent of vascular wilt of oaks and other trees in the family Fagaceae. Ceratocystis harringtonii, previously known as C. populicola, causes disease in Populus species in the USA and Canada. Grosmannia penicillata is the causal agent of bluestain of sapwood on various conifers, including Picea spp. and Pinus spp. in Europe. Huntiella bhutanensis is a fungus in Ceratocystidaceae and known only in association with the bark beetle Ips schmutzenhorferi that infests Picea spinulosa in Bhutan. The availability of these genomes will facilitate further studies on these fungi.

Key words: bluestain fungi entomogenous fungi Ips Populus tree diseases vascular wilt

IMA Genome-F 7A

Draft genome sequence for the oak pathogen Ceratocystis fagacearum

Ceratocystis fagacearum (Microascales; Ceratocystidaceae) is a wilt pathogen of oak (Quercus spp.) and other Fagaceae in the eastern and north-central US (Henry et al. 1944, Billings 2000, Juzwik et al. 2004). Based on ecological and phylogenetic data, however, C. fagacearum falls outside of the genus Ceratocystis sensu stricto as defined by De Beer et al. (2014). Its taxonomic position as a discrete genus is currently being established (De Beer, unpublished).

Ceratocystis fagacearum causes a damaging and important vascular wilt disease known as oak wilt (Appel 1995). Infection occurs in spring through wounds commonly made during pruning operations. Trees die rapidly and the pathogen can pass from one tree to another via root grafts resulting in rows or patches of dying trees. As the trees die, pressure pads develop under the bark to expose spore-bearing mats where C. fagacearum produces a fruity odour attractive to insects (Lin & Phelan 1992, Cease & Juzwik 2001). These include sap-feeding nitidulid beetles that can transfer the pathogen to freshly made wounds on trees, thereby resulting in new infections (Cease & Juzwik 2001, Juzwik et al. 2004).

The aim of this study was to sequence the genome of C. fagacearum. The data are thus intended to compliment the previously established genome resources for species in the Ceratocystidaceae (Wilken et al. 2013, van der Nest et al. 2014a, b, Wingfield et al. 2015a, b, Belbahri 2015, Wingfield et al. 2016), particularly in order to draw comparisons between them.

SEQUENCED STRAIN


NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The Ceratocystis fagacearum isolate CMW 2656 Whole Genome Shotgun project has been deposited in GenBank under the accession number MKGJ00000000.

MATERIALS AND METHODS

Ceratocystis fagacearum isolate CMW 2656 is obtainable from the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa and the CBS-KNAW Fungal
**RESULTS AND DISCUSSION**

The draft nuclear genome of *C. fagacearum* had an estimated size of 26,736,264 bases with 123x coverage. The assembly included 1257 contigs larger than 500 bases with an N50 value of 42,305 bases and an average GC content of 46.9%. AUGUSTUS predicted 6703 ORFs, which correlated with an average gene density of 251 ORFs/Mb. This assembly had a high degree of completeness with a BUSCO score of 96%, of which 1321 were Complete Single-Copy BUSCOs, 61 were Complete Duplicated BUSCOs, 47 were Fragmented BUSCOs, and only nine were missing BUSCO orthologs out of the 1438 BUSCO groups searched (Simao et al. 2015).

Comparison of these genome statistics revealed no striking differences between the genome of *C. fagacearum* and those determined in previous studies (Wilken et al. 2013, van der Nest et al. 2014a, b, Wingfield et al. 2015a, b, 2016, Belbahri 2015). This is despite its unique ecology and taxonomic position (de Beer 2014). Future genome-based analyses will undoubtedly improve our understanding of the molecular and evolutionary processes underlying the biology of this pathogen.

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

*Contact:* Magriet.vanderNest@fabi.up.ac.za

**IMA Genome-F 7B**

**Draft genome sequence of the poplar pathogen *Ceratocystis harringtonii***

The genome of black cottonwood (*Populus trichocarpa*) has been sequenced and is available for study, which has resulted in it becoming an important model tree for woody plant research (Tuskan et al. 2006). This tree species is well known for its resistance to fungal infection, and to date only a few pathogens are known to infect it (Royle & Ostry 1995). Fungal species currently causing severe disease on plantation-grown poplar include rust diseases caused by species of *Melampsora* (e.g. Newcombe et al. 2000) and leaf spots caused by *Marssonia* spp. (e.g. Erickson et al. 2004). In order to gain increased understanding of the adaptations allowing pathogens to infect poplar, the genome of the European poplar rust fungus, *Melampsora larici-populina*, was sequenced. This provided a powerful tool for studies in molecular plant-pathogen interactions in the phyllosphere of a perennial host (Duplessis et al. 2011). In contrast, little is known regarding stem-infecting pathogens in poplar, and particularly which resistance mechanisms exist to provide such extensive protection against fungal infection of its woody tissues. Furthermore, knowledge pertaining to strategies that pathogens employ to overcome this resistance has not been well documented.

*Ceratocystis harringtonii* resides in *Ceratocystidaceae* (de Beer et al. 2013a, b), and was previously treated as *C. populicola* (Johnson et al. 2005). It causes target-shaped cankers on infected trunks and rooted cuttings (Wood & French 1963, Johnson et al. 2005). Furthermore, it is one of the few pathogens known to breach the effective defence barriers in poplar stems. This fungus occurs throughout the natural range of *Populus* species in the USA and Canada, as well as in hybrid poplar plantations in Poland (Gremmen & de Kam 1977). The pathogen is most aggressive on the North American poplar species *P. trichocarpa*, *P. balsamifera*, and *P. tremuloides*, while the European *P. nigra* is known to be almost entirely resistant to infection (Johnson et al. 2005). One mechanism by which *C. harringtonii* elicits host defences is by the production of cerato-populin, a pathogen-associated molecular pattern protein. This toxin is similar to the well-characterized ceratoplatinan and cerato-ulmin, which are known in *C. platani* and *Ophiostoma novo-ulmi*, respectively (Comparini et al. 2009, Lombardi et al. 2013, Martellini et al. 2013).

In order to extend our knowledge on plant-pathogen interactions in woody hosts and to understand the basis of resistance against fungal stem infection in poplar, the genome of *C. harringtonii* was sequenced. This genome sequence will provide a powerful tool to interrogate the mechanisms by which plant pathogens infect woody stems and cause canker disease in a highly resistant tree species.

**SEQUENCED STRAIN**

**Poland:** isol. ex hybrid poplar *Populus maximowiczii* x *P. laurifolia* x *P. nigra* ‘Italica’ (*P. xberolinensis*), 1977, J.
NUCLEOTIDE SEQUENCE ASSESSMENT NUMBER

This Whole Genome Shotgun project of the Ceratocystis harringtonii genome has been deposited at DDBJ/EMBL/GenBank under accession number MKGM00000000; this is the first version described here.

MATERIALS AND METHODS

Total genomic DNA was isolated from the mycelium of a single-spore culture of isolate C. harringtonii CMW 14789 grown on 2 % malt extract agar (MEA; 2 % w/v, Biolab, Midrand, South Africa) supplemented with 100 µg/L thiamine for 10 d using the method of Barnes et al. (2001). The Genomics analyzer Ilx platform (Illumina) at the Genome Centre (University of California at Davis, CA) was used for sequencing the genome. Two 350-bp and three 600-bp paired-end libraries were made using standard Illumina protocols. Sequence data was assessed for quality, trimmed and assembled with the software package CLC Genomics (CLCBio, Aarhus, Denmark) using default settings. Poor quality reads (limit of 0.05) and/or terminal nucleotides were discarded. The contigs were assembled into scaffolds using SSPACE v.2.0 (Boetzer et al. 2011) and gaps were filled using GapFiller v.2.2.1 (Boetzer & Pirovano 2012). The genome assembly was verified and completeness assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) software (Simão et al. 2015). Contigs smaller than 500 bp were removed from the final dataset.

RESULTS AND DISCUSSION

Sequencing of five C. harringtonii DNA libraries yielded a total of 31 057 034 paired-end reads with an average length of 101 bases. After trimming, 30 972 782 reads were recovered with an average length of 96 bases. The estimated size of the assembled draft genome was 26 Mb with a coverage of 110x. The genome had a mean GC content of 48.8 % and a N50 contig size of 66 kb. A total of 920 contigs were assembled of which 813 were larger than 500 bases (excluding the mitochondrial genome sequence data). BUSCO analysis revealed that out of 1438 possible BUSCO groups searched 1327 were single-copy, while 67 were duplicated and six were fragmented or missing. Analysis using AUGUSTUS (Stanke et al. 2006) revealed 6627 putative open reading frames.

Genome sizes differ widely among Sordariomycetes, for example the Fusarium graminearum and Trichoderma reesei genomes have approximate sizes of 36 Mb and 34 Mb, respectively (Martinez et al. 2008, King et al. 2015), while the genome of Ceratocystis minuta is only 21 Mb in size (Wingfield et al. 2016). The estimated genome size of C. harringtonii (26 Mb) is marginally smaller than the genomes reported for other closely related members of Ceratocystidaceae (28 – 32 Mb; Wingfield et al. 2016, Wilken et al. 2013, van der Nest et al. 2014). Despite the smaller genome size, the C. harringtonii genome contains similar numbers of predicted ORF’s as other sequenced species of Ceratocystis (Wilken et al. 2013, van der Nest et al. 2014, Wingfield et al. 2016).

The C. harringtonii draft genome will be an important resource in studies of plant/pathogen interactions in woody tissues. This is especially true because this fungus is one of only a few reported pathogens that can overcome the defence mechanisms in the stem tissues of the model tree, P. trichocarpa. Furthermore, with growing numbers of genomes available for species in Ceratocystidaceae, the C. harringtonii genome will also be used for comparative genomic studies and those considering the mechanisms of host specialization in this fascinating group of plant pathogens.

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

*IContact: almuth.hammerbacher@fabi.up.ac.za

IMA Genome-F 7C

Draft genome sequence of Grosmannia penicillata

The asexual morph of Grosmannia penicillata was first described as Leptographium penicillatum (Grosman 1931), the causal agent of blue stain of sapwood surrounding the galleries of the spruce bark beetle, Ips typographus. The sexual morph was discovered soon afterwards and described as Ceratostomella penicillata (Grosman 1932), but Goidánich (1936) considered the species to be distinct from other Ceratostomella spp. and introduced a new generic name, Grosmannia, with G. penicillata as type species. Subsequently, Grosmannia was treated as a synonym of Ophiostoma (Siemaszko 1939, Arx 1952, Jacobs & Wingfield 2001) and Ceratocystis (Bakshi 1951, Hunt 1956, Upadhyay 1981) until Zipfel et al. (2006) re-instated the name to accommodate the sexual morphs of Leptographium spp. After the implementation of the one fungus one name principles (Hawksworth 2011), Leptographium, as the older generic name, currently has preference over Grosmannia (de Beer & Wingfield 2013). However, the type species of the two genera, L. lundbergii and G. penicillata, group in distinct lineages of which the generic status needs reconsideration (de Beer & Wingfield 2013). For the interim, the lineage that includes G. penicillata and 17 other Leptographium and Grosmannia species, are referred to as the G. penicillata species complex in Leptographium s. lat. (Six et al., 2011, Linnakoski et al. 2012, de Beer & Wingfield 2013). Grosmannia penicillata occurs on various conifers including Picea and Pinus spp. in Europe (Jacobs & Wingfield 2001). It is vectored by various scolytine bark beetle species (Coleoptera; Curculionidae; Scolytinae) but most importantly, by the aggressive tree-killing bark beetle Ips typographus (Jacobs & Wingfield 2001, Linnakoski et al. 2012). Inoculation
studies have indicated that *G. penicillata* is not pathogenic to its hosts (Jankowiak et al. 2009, Repe et al. 2015). This is unlike *Endoconidiophora polonica*, as defined by de Beer et al. (2014), that is a common associate of *I. typographus* and is able to kill trees in inoculation tests (Horntvedt et al. 1983, Repe et al. 2015).

In this study, we determined the genome sequence of *G. penicillata*. The primary intention was to provide basal genomic data to enable further studies on the taxonomy and evolutionary relationships of this species and other genera and species in the Ophiostomatales.

**SEQUENCED STRAIN**


**NUCLEOTIDE SEQUENCE ACCESSION NUMBER**

The genomic sequence of *Grosmannia penicillata* (CMW 2644, CBS 116008) has been deposited at DDBJ/EMBL/GenBank under the accession number MLJ0000000. The version described in this paper is version MLJ0V01000000.

**MATERIALS AND METHODS**

*Grosmannia penicillata* isolate CMW 2644 was obtained from the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI), the University of Pretoria, South Africa. Genomic DNA was extracted from the freeze-dried mycelium using the method described by Duong et al. (2013). Two pair-end libraries (350 bp and 530 bp average insert size) were prepared and sequenced using the Illumina HiSeq 2000 platform with 100 bp read length. Obtained reads were subjected to quality and adapters trimming using Trimomatic v. 0.36 (Bolger et al. 2014). The genome was assembled using the program SPAdes v. 3.9 (Bankevich et al. 2012). The scaffolds obtained from SPAdes were subjected to further scaffolding with SSPACE-standard v. 3.0 (Boetzer et al. 2011). Assembly gaps were filled with GapFiller V1-10 (Boetzer & Pirovano 2012). The completeness of the assembly was assessed with the Benchmarking Universal Single-Copy Orthologs (BUSCO v. 1.1b1) program using the fungal dataset (Simão et al. 2015). The number of protein coding genes was determined using Augustus v 3.2.2 (Stanke et al. 2006).

**RESULTS AND DISCUSSION**

Over 11 million read pairs and 1.4 million single reads were obtained after the quality trimming. *De-novo* assembly using SPAdes resulted in 293 scaffolds which were larger than 500 bp. The number of final scaffolds was reduced to 199 after scaffolding with SSPACE and filling gaps with GapFiller. The current assembly has an N50 of 316 Kb and size of 26.33 Mb, with an overall GC content of 58.73 %. When compared with other species of *Leptographium s. lat.* for which whole genome data are available, *G. penicillata* has a similar genome size to that of *Leptographium lundbergii* (26.6 Mb) (Wingfield et al. 2015), and slightly smaller than those of *G. clavigera* (29.8 Mb) (DiGuistini et al. 2011) and *L. procerum* (28.6 Mb) (van der Nest et al. 2014). The assembly included 94.6 % complete, 4.2 % fragmented, and 1.2 % missing, BUSCOs. Augustus predicted a total of 8713 protein coding genes, of which 6718 are multi-exon and 1995 are single-exon genes. As the type species of *Grosmannia*, the genome sequence of *G. penicillata* will be a valuable resource to study the taxonomic relationships between species in the genus. The data will also be useful in comparisons between genera in Ophiostomatales where questions relating to their ecology and evolutionary biology are of particular interest.

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

*Contact*: Tuan.Duong@fabi.up.ac.za

**IMA Genome-F 7D**

Draft genome sequence for the bark beetle-associated fungus *Huntiella bhutanensis*.

*Huntiella bhutanensis* (*Ascomycota: Microascales*) is a filamentous fungus in *Ceratocystidaceae*. The species was previously accommodated in *Ceratocystis* as *C. bhutanensis* (van Wyk et al. 2004). However, a recent taxonomic review of this and related genera led to the re-circumscription of these fungi into distinct genera based on morphological, ecological and molecular characteristics (de Beer et al. 2014). Thus, *H. bhutanensis*, along with other cosmopolitan saprobes such as *Huntiella omanensis* and *H. moniliformis*, were assigned to the genus *Huntiella* (de Beer et al. 2014).

*Huntiella bhutanensis* was first isolated from adults of the bark beetle *Ips schmutzenhoferi* (*Coleoptora; Curculionidae; Scolytinae*) or their galleries found on *Picea spinulosa* in Bhutan (van Wyk et al. 2004). This was only the second report of any ophiostomatoid fungi from this country and represented the first species of *Ceratocystidaceae* to be reported in this locality (van Wyk et al. 2004). Despite its close relationship with the bark beetle, and unlike many other bark beetle-associates in the *Ceratocystidaceae*, *H. bhutanensis* is not considered a primary pathogen and in inoculation studies, it gave rise to only small necrotic lesions on *P. spinulosa* trees (Kirisits et al. 2012).

The non-pathogenic nature of *H. bhutanensis* is consistent with other *Huntiella* species that are weak pathogens or saprobes (de Beer et al. 2014). It also shares many morphological characteristics with other *Huntiella* species, such as globose ascomatal bases, extended ascomatal necks, and hat-shaped ascospores (van Wyk et al. 2004). It is unusual amongst species in the genus in that it is able to grow at 4°C (van Wyk et al. 2004) and it gives off a distinctive odor.
putrid odour unlike the sweet smelling odours typical of other *Huntiella* species (van Wyk et al. 2004).

The aim of this study was to produce a good quality draft genome assembly for *H. bhutanensis* for use in future comparative genomics studies, both between species and genera in Ceratocystidaceae. Genome sequences are already publically available for *H. moniliformis* (van der Nest et al. 2014a) and *H. omanensis* (van der Nest et al. 2014b) and thus the availability of this genome assembly will aid in the elucidation and comparison of ecological strategies, sexual cycles, and other key life-style aspects of and between these species.

**SEQUENCED STRAINS**

*H. bhutanensis* isolate CMW 8217 (CMW 8217, CBS 114289, PREM 57807-dried culture).

**NUCLEOTIDE SEQUENCE ACCESSION NUMBER**

The *Huntiella bhutanensis* isolate CMW 8217 Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the MJMS00000000. The version described in this paper is version MJMS01000000.

**MATERIALS AND METHODS**

*Huntiella bhutanensis* isolate CMW 8217 is preserved in the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI) at the University of Pretoria, South Africa. It can also be requested from the CBS-KNAW Fungal Biodiversity Centre (CBS), Utrecht, The Netherlands. Genomic DNA was isolated using the method of Barnes et al. (2001) and sequenced using the Genomics Analyzer IIX, Illumina platform at the UC Davis Genome Centre, University of California, Davis (CA). Paired-end libraries of both 350-bp and 600-bp insert sizes were prepared and sequenced following Illumina protocol. Quality trimming of the reads obtained was conducted using CLC Genomics Workbench v. 7.5.1 (CLCBio, Aarhus). This same program was used to assemble a draft genome sequence using the *de novo* assembly function, with default settings. Thereafter, SSPACE v. 2.0 (Boetzer et al. 2011) was used to scaffold the assemblies, and gaps created during this process were filled using GapFiller v. 2.2.1 (Boetzer & Pirovano 2012). The web-based *de novo* gene prediction software AUGUSTUS was used to predict the number of putative open reading frames (ORFs) in this assembly using the gene models of *Fusarium graminearum* (Stanke et al. 2004). Genome completeness was assessed with the Benchmarking Universal Single-Copy Orthologs (BUSCO v. 1.2.2) tool (Simão et al. 2015) using the fungal dataset.

**RESULTS AND DISCUSSION**

Sequencing of the *Huntiella bhutanensis* libraries yielded 35,886,298 raw reads with an average length of 101 bases. Trimming and quality control left 35,822,293 reads with an average length of 96 bases. The assembled nuclear genome of *H. bhutanensis* had a size of 26.77 Mb with an approximate GC content of 47.9 %. Web AUGUSTUS predicted 7,261 ORFs. This corresponded to an average gene density of 279 ORFs/Mb. The BUSCO analysis for this assembly also indicated a high level of completeness. Out of the 1438 fungal BUSCO groups searched, the genome contained 1315 (91.4 %) complete single copy BUSCOs, 62 (4.3 %) complete duplicated BUSCOs, 52 (3.6 %) fragmented BUSCOs and only nine (< 1 %) missing BUSCO orthologs. The genome of *H. bhutanensis* is about 5 Mb smaller than that of its phylogenetically-close relative *H. omanensis* and possesses 1134 fewer genes (van der Nest et al. 2014b). However, the genome assemblies of *H. bhutanensis* and *H. moniliformis* are more similar, at 25.4 Mb and 26.7 Mb with 6832 and 7261 ORFs, respectively (van der Nest et al. 2014a). The genome sequence of *H. bhutanensis* will serve as an essential resource in future genome comparisons between species of this genus and will aid in the characterization of their lifestyles, sexual cycles, and other key aspects of their biology.

**ACKNOWLEDGEMENTS**

This work was co-funded by the Genomics Research Institute (University of Pretoria), the University of Pretoria Research Development Programme, the DST/NRF Center of Excellence in Tree Health Biotechnology (FABI, University of Pretoria), and the National Research Foundation (NRF) (Grant number 87332). Sequencing the genomes of *Ceratocystis fagacearum*, *C. harringtonii*, *Huntiella bhutanensis* was made possible through the Improving Academics Qualifications (IAQ) grant from the NRF.

**REFERENCES**


Bakshi, BK (1951) Studies on four species of *Ceratocystis*, with a discussion on fungi causing sap-stain in Britain. Mycological Papers 35: 1–16.


Mayers CG, McNew DL, Harrington TC, Roepar RA, Fraedrich SW, Biedermann PH, Casstrillo LA, Reed SE (2015) Three genera in the *Ceratocystidaceae* are the respective symbionts of three independent lineages of ambrosia beetles with large, complex


