Accepted Manuscript

Ceratocystidaceae exhibit high levels of recombination at the mating-type (*MAT*) locus

Melissa C. Simpson, Martin P.A. Coetzee, Magriet A. van der Nest, Michael J. Wingfield, Brenda D. Wingfield

PII: S1878-6146(18)30293-9

DOI: 10.1016/j.funbio.2018.09.003

Reference: FUNBIO 959

To appear in: *Fungal Biology*

Received Date: 10 November 2017

Revised Date: 11 July 2018

Accepted Date: 12 September 2018

Please cite this article as: Simpson, M.C., Coetzee, M.P.A., van der Nest, M.A., Wingfield, M.J., Wingfield, B.D., *Ceratocystidaceae* exhibit high levels of recombination at the mating-type (*MAT*) locus, *Fungal Biology* (2018), doi: https://doi.org/10.1016/j.funbio.2018.09.003.

Title

Ceratocystidaceae exhibit high levels of recombination at the mating-type (*MAT*) locus

Authors

Melissa C. Simpson

Melissa.Simpson@fabi.up.ac.za

Martin P.A. Coetzee

Martin.Coetzee@fabi.up.ac.za

Magriet A. van der Nest

Magriet.vanderNest@fabi.up.ac.za

Michael J. Wingfield

Mike.Wingfield@fabi.up.ac.za

Brenda D. Wingfield

Brenda.Wingfield@fabi.up.ac.za

Author Affiliations

Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, 0002, South Africa

Corresponding Author

Brenda D. Wingfield

Telephone Number: +27 12 420 6288

Fax Number: +27 12 420 3960

Email Address: Brenda.Wingfield@fabi.up.ac.za

Postal Address: Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, 0002, South Africa

Abstract

Mating is central to many fungal life cycles and is controlled by genes at the mating-type (*MAT*) locus. Genes at the *MAT* locus determine whether the fungus will be self-sterile (heterothallic) or self-fertile (homothallic). Species in the ascomycete family *Ceratocystidaceae* have different mating strategies (e.g., heterothallic or homothallic) making them interesting to consider with regards to their *MAT* loci. The aim of this study was to compare the composition of the regions flanking the *MAT* locus in 11 species of *Ceratocystidaceae* representing four genera. Genome assemblies for each species were examined to identify the *MAT* locus and determine the structure of the flanking regions. Large contigs containing the *MAT* locus were then functionally annotated and analysed for the presence of transposable elements. Similar to previous studies, the genes typically flanking the *MAT* locus in the sordariomycetes related to the *Ceratocystidaceae* were found to be highly conserved. In contrast, species of *Ceratocystidaceae* had a much greater level of variation in gene order and presence around the *MAT* locus compared to other ascomycetes. The different genera in the *Ceratocystidaceae* displayed little synteny outside of the immediate *MAT* locus flanking genes. Recombination within this locus and the regions flanking it has been shown to be very low. Even though the three species of *Ceratocystis* did not show much synteny outside of the immediate *MAT* locus flanking genes, species of *Huntiella* and *Endoconidiophora* were comparatively syntenic. Because *Ceratocystis* species had a higher number of transposable elements in the *MAT* flanking regions when compared to their genomes and to the rest of the species studied here, we hypothesise that *Ceratocystis* species may have undergone recombination in this region due to the presence of these elements.

Keywords

Ceratocystidaceae; *MAT*; transposable elements; recombination.

Abbreviations

MAT: mating-type

1. Introduction

During sexual reproduction in fungi, exchange of genetic material (recombination) takes place in order to produce genetically diverse offspring and to ensure the survival of species (Lee *et al*., 2010). Due to the structure of each chromosome, recombination "hot spots" and "cold spots" exist (Yamada *et al*., 2004). "Cold spots" of recombination safeguard specific regions of the genome in order to retain particular linked genes, while "hot spots" ensure that other regions constantly change in order to produce the necessary diversity. In fungi, the flanking regions of the mating-type (*MAT*) locus, which consist of *MAT* genes that control and ensure completion of sexual reproduction, are regions where variable levels of recombination are observed (Coppin *et al*., 1997; Gallegos *et al*., 2000; Hsueh *et al*., 2006).

Although there is some "recombination" in the form of mating-type switching in the *MAT* locus of some ascomycetous yeast species (Barsoum *et al*., 2010; Maekawa and Kaneko, 2014; White and Haber, 1990) and filamentous ascomycetes (Mathieson, 1952; Uhm and Fujii, 1983; Wheeler, 1950), the region exhibits limited recombination (Grognet *et al*., 2014; Kubisiak and Milgroom, 2006; Lee *et al*., 1999). This has particularly been observed in the ascomycete *Neurospora tetrasperma* where approximately 75% of the chromosome on which the *MAT* locus is located, has suppressed recombination (Ellison *et al*., 2011; Gallegos *et al*., 2000; Menkis *et al*., 2008). Moreover, obligatory recombination is observed just outside this region and is thought to ensure correct pairing of chromosomes for segregation (Ellison *et al*., 2011; Menkis *et al*., 2008). In stark contrast, the basidiomycete *Cryptococcus neoformans* exhibits very high levels of recombination in the *MAT* locus flanking regions, but recombination is suppressed in its remarkably large *MAT* locus (Hsueh *et al*., 2006; Lengeler *et al*., 2002). It appears that non-recombining DNA promotes recombination in other regions, thus encouraging evolution of new species (Idnurm, 2011).

Ascomycetes can be homothallic (self-fertile) or heterothallic (outcrossing) and this is dictated by the presence of particular *MAT* genes at the *MAT* locus (Nelson, 1996). Heterothallic ascomycetes contain one of two dissimilar "alleles" known as idiomorphs, at their *MAT* locus (Metzenberg and Glass, 1990). The *MAT1-1* idiomorph is characterised by the essential α-domain-containing *MAT1-1-1* gene (Debuchy and Coppin, 1992; Glass *et al*., 1990), while the *MAT1-2* idiomorph typically contains *MAT1-2-1* which harbors an HMG-domain (Debuchy *et al*., 1993; Philley and Staben, 1994). Both idiomorphs can also contain other *MAT* genes that are involved in post-fertilisation events (Ferreira *et al*., 1998). Homothallic ascomycetes contain genes from both idiomorphs in their genomes, either at a single locus or in more than one location (Paoletti *et al*., 2007; Pöggeler *et al*., 1997). Typically the *MAT* locus is flanked upstream by cytoskeleton assembly control (*SLA2*), and

downstream by DNA lyase (*APN2*), cytochrome C oxidase subunit (*COX*), and anaphase promoting complex (*APC*) genes (Amselem *et al*., 2011; Aronstein *et al*., 2007; Jacobson, 2005; Kanematsu *et al*., 2007).

Genera and species in the *Ceratocystidaceae* (Microascales) as defined by de Beer *et al*. (2014) display a variety of mating strategies and lifestyles. Most species of *Ceratocystis* are homothallic and many are ecologically important pathogens of trees and root crops (Kile, 1993; Wingfield *et al.,* 2013). Similarly, *Endoconidiophora* also has homothallic species and these include pathogens that infect conifers as well as species that cause sap-stain in wood (Seifert, 1993; Wingfield *et al.,* 1997, Redfern *et al.,* 1987). In contrast, *Huntiella* is a genus that includes saprobes, some of which cause sap-stain of timber and most are heterothallic (Roux *et al.,* 2004; Seifert, 1993; Tarigan *et al.,* 2010; Wilson *et al.,* 2015). Species of *Davidsoniella* and *Thielaviopsis* are important pathogens of trees and monocoteledonous plants, respectively, some of which have known sexual states (Mbenoun *et al.*, 2014; Wingfield *et al.,* 2013). The two genera, *Ambrosiella* and *Chalaropsis,* are considered asexual because no sexual structures have been observed for them (de Beer *et al.,* 2014). *Ambrosiella* spp. are symbionts of ambrosia beetles, while *Chalaropsis* spp. are found on woody plants but nothing is known regarding their ecological importance (Harrington *et al.*, 2010; de Beer *et al.,* 2014).

MAT loci amongst members of the *Ceratocystidaceae* have different architecture, and the genes flanking these loci are not always similar (Wilken *et al*., 2014; Wilson *et al*., 2015; Witthuhn *et al*., 2000). The *MAT* locus of *Ceratocystis fimbriata sensu stricto*, the type species of the genus, consists of *MAT1-1-1, MAT1-2-1* and *MAT1-1-2,* and is flanked upstream by *SLA2, APC, APN2* and *COX*, with a hypothetical protein and a putative importin-beta domain-containing protein downstream (Wilken *et al*., 2014). Mating-type switching through deletion of *MAT1-2-1* has been observed in this fungus, along with other species, such as *C. albifundus, Davidsoniella virescens, Endoconidiophora coerulescens* and *E. pinicola* (Harrington and McNew, 1997; Lee *et al*., 2015; Wilken et al., 2014; Witthuhn *et al*., 2000). The type species of *Huntiella, H. moniliformis,* contains *MAT1-2-1* and *MAT1-2-7* at its *MAT* locus and is still able to mate with itself despite only carrying genes from a single *MAT* idiomorph (Wilson *et al*., 2015). This is known as unisexuality and is a form of homothallism. *Huntiella omanensis* is heterothallic and contains either *MAT1-1-1* and *MAT1-1-2,* or *MAT1-2-1* and *MAT1-2-7* at its *MAT* locus (Wilson *et al*., 2015). In both of these species, *SLA2* and *APC* flank the *MAT* locus upstream, with two genes of unknown function found downstream of the locus (Wilson *et al*., 2015). The wide range of mating

strategies in genera and species of the *Ceratocystidaceae* makes this group of fungi particularly interesting to study with regards to mating and the structure of the *MAT* locus.

The aim of this study was to compare the composition of the *MAT* locus flanking regions and to identify trends between genera and species of the *Ceratocystidaceae* by detecting the genes and transposable elements contained within the *MAT* locus flanking regions. These comparisons were carried out on 11 species residing in four genera including *Ceratocystis* (four species), *Chalaropsis* (one species), *Endoconidiophora* (two species), and *Huntiella* (four species) species. Among the current species included in *Ceratocystis*, *C. adiposa* is no longer considered part of this genus (de Beer *et al*., 2014) but has not yet been renamed, and is treated as a separate genus in this study. Species of *Ceratocystis* and *Endoconidiophora* included in this study are plant pathogens, while *Huntiella* and *Chalaropsis* are saprobes (de Beer *et al.,* 2014). Other than *Ch. thielaviopsis*, sexual structures have been observed for all species included in this study, they are therefore capable of sexual reproduction (Halsted, 1890; Kamgan *et al.,* 2008; Nkuekam *et al.,* 2013; Redfern *et al.,* 1987; Sartoris, 1927*;* Solheim, 1986; van Wyk *et al.*, 1991; van Wyk *et al.*, 2004; van Wyk *et al.,* 2007; Wingfield *et al.,* 1996).

2. Methods and Materials

2.1. Genome assemblies included in this study

At the time of this study, genome assemblies for 11 species residing in the *Ceratocystidaceae* as defined by de Beer *et al*. (2014) were available on NCBI. These included *C. albifundus, C. fimbriata, C. manginecans, C. adiposa, Ch. thielavioides, E. laricicola, E. polonica, H. bhutanensis, H. decipiens, H. moniliformis,* and *H. savannae* (Table 1). All 11 genome sequences were generated using Illumina Hiseq or GAIIx sequencing and were assembled using one or a combination of assembly software, including ABySS (Simpson *et al.* 2009), ALLPATHS-LG (Gnerre *et al.,* 2011), CLC Genomics Workbench (CLC Bio, Aarhus, Denmark), GapFiller (Boetzer and Pirovano, 2012), SSPACE (Boetzer *et al.,* 2011), and Velvet (Zerbino and Birney, 2008). These genomes were used to create a local database for downstream analyses.

2.2. Phylogenetic analysis and ancestral state reconstruction

A phylogenetic tree was constructed to obtain a phylogenetic framework for the species included in this study. The phylogenetic tree was generated based on nine genes, including the small ribosomal subunt (18S), large ribosomal subunit (28S), beta tubulin (BT), internal transcribed spacer region (ITS), minichromosome maintenance complex component 7 (MCM7), DNA-directed RNA polymerase II core sub-units RPB1 and RBP2, translation

elongation factor 1α (TEF1α), and translation elongation factor 3α (TEF3α). Each genome was queried using genes obtained from NCBI for each species in local BLAST analyses using CLC Genomics Workbench v8.5.1 (CLC Bio, Aarhus, Denmark). The nine genes identified in each species were then concatenated using FASconCAT-G v1.02 (Kück and Longo, 2014). A neighbour-joining phylogenetic tree was produced in CLC Main Workbench v7.7.1 (CLC Bio, Aarhus, Denmark) using the "Create Tree" function with the Kimura 80 model and 1000 bootstrap replicates, after a test to determine the best fit nucleotide substitution model was performed. *Fusarium circinatum* was chosen as an outgroup to all the species selected for the *MAT* region synteny comparison and was therefore also used to root the phylogenetic tree. This species was selected because its *MAT* region showed synteny with that of the *Ceratocistidaceae* included in this study.

The ancestral mating strategy was determined using the ancestral state reconstruction function in Mesquite v3.5 (Maddison and Maddison, 2018). For this purpose a data matrix was created with the mating character states defined as self-fertile, asexual and unisexual. A user tree, based on the phylogeny obtained from the neigbour-joining analyses, was used as the phylogenetic framework to trace the character state changes over the tree. Inference of ancestral character states were done based on parsimony and maximum likelihood. For the maximim likelihood analysis, a one-parameter Markov k-state model was used and the probability of each character state expressed as the proportional likehood.

Identification of MAT locus and flanking regions

MAT genes and their immediate flanking genes, including *SLA2, APN2, COX* and *APC,* from *C. fimbriata* (Wilken *et al*., 2014) or *H. moniliformis* (Wilson *et al*., 2015) were used to query each of the *Ceratocystidaceae* genomes in local BLAST analyses using CLC Genomics Workbench v8.5.1 (CLC Bio, Aarhus, Denmark). The contigs on which the genes were identified were selected for further analyses. Two contigs each were joined at the *MAT* locus for *C. manginecans* and *E. laricicola* using the known *C. fimbriata MAT* locus (Wilken *et al.* 2014) and the *E. polonica MAT* locus identified during this study, respectively. AUGUSTUS v3.2.1 (Stanke and Morgenstern, 2005) was employed to predict genes on each contig and BLAST2GO 4.1.9 (Conesa *et al*., 2005) was utilised to determine the function of each gene.

2.3. Comparison of gene order and function

Once the contigs with the *MAT* genes were annotated, comparisons of gene order and function were made between genera and species of *Ceratocystidaceae*, in particular just outside of the *MAT* locus and the immediate flanking genes. This was achieved by manually

drawing the genes contained in approximately 30kb on either side of the *MAT* locus for each sequence. Homologous genes were identified through local BLAST searches and alignments were performed in CLC Genomics Workbench v8.5.1. A synteny map was drawn using GenoplotR v0.8.4 (Guy *et al*., 2010) in order to make overall comparisons. Transposable elements were identified in each genome sequence and in the region 60 kb on either of the *MAT* locus of each species using REPET v2.5 (Flutre *et al.,* 2011). The number of transposable elements was calculated per 10kb in order to compare the density of transposable elements in the whole genomes and *MAT* locus flanking regions of *Ceratocystidaceae*. A linear regression analysis was performed in Microsoft Excel to determine if there was a correlation between the number of contigs and the number of transposable elements identified in each genome.

3. Results

The phylogenetic tree (Figure 1) placed the 11 species studied here in monophylogenetic groups representing the different genera to which they belong. The exception was *Ceratocystis adiposa* that grouped paraphyletic to the other *Ceratocystis* species, confirming that this species should no longer be accommodated in the genus *Ceratocystis* (de Beer *et al*., 2014). Analysis of the ancestral state mating strategy revealed that the common ancestor of all the species studied here was most likely self-fertile (homothallic; Figure 1 and Table 2). The ancestral mating strategy at nodes one to six were most likely self-fertile, while those at nodes 9 and 10 were most likely self-sterile. Both the parsimony and maximum likelihood analyses indicated that the common ancestor at node 8 of the four *Huntiella* spp. was either self-fertile, self-sterile or unisexual.

In three genera of *Ceratocystidaceae; Ceratocystis, Chalaropsis* and *Endoconidiophora,* the *MAT* locus was flanked upstream by *COX* (A), *APN2* (B), *APC* (C) and *SLA2* (D; Figure 2). In all four *Huntiella* species, *COX* (A) and *APN2* (B) were found next to each other but residing on a contig different from the *MAT* locus, while *APC* (C) and *SLA2* (D) were found upstream of the *MAT* locus. Genes upstream from *COX* (A) in *Huntiella* were similar in function and order to those upstream of *COX* (A) in both species of *Endoconidiophora,* and *C. adiposa.*

Genes downstream of the *MAT* locus differed between the different genera in *Ceratocystidaceae.* In *Ceratocystis* and *Chalaropsis* the region downstream from the *MAT* locus was flanked by a hypothetical protein gene and a putative importin-beta domain-containing gene (now referred to as importin; E). *Endoconidiophora* contained two homologous flanking genes of unknown function (27, 28). *Huntiella* also contained two

flanking genes of unknown function (28, 40), but only one (28) was homologous to those present in *Endoconidiophora*. *Ceratocystis adiposa* differed from other genera in that the *MAT* locus was flanked downstream by a feruloyl esterase B gene (52) and a different gene of unknown function (53).

Synteny around the *MAT* locus differed between genera and within genera. Overall synteny between genera was low outside of the *MAT* locus and its immediate flanking genes (Figure 3). *Ceratocystis adiposa,* however, shared genes with *Endoconidiophora* and *Huntiella* outside of the *MAT* locus. In contrast to low synteny between genera, synteny of genes outside the *MAT* between species within a genus was generally high. This was clearly seen in *Endoconidiophora* and *Huntiella* (Figure 3). Species of *Huntiella* showed some differences in gene content but overall were highly syntenic. However, this was not the case for *Ceratocystis* species. Although *C. fimbriata* and *C. mangincecans* were somewhat syntenic, *C. albifundus* was quite different. Once outside of the *MAT* locus and the immediate flanking genes, differences between these species more closely resembled those found between genera.

The number of transposable elements identified in each genome differed between the species of *Ceratocystidaceae*, however, *Ceratocystis* spp. had the highest densities. The linear regression analysis showed a weak correlation between the number of contigs and number of transposable elements in each genome (Supplementary Figure 1). This indicates that assemblies with fewer contigs do not always have a higher number of transposable elements present. Most species of *Ceratocystidaceae* included in this study contained a higher number of transposable elements in the *MAT* locus flanking regions in comparison to the rest of their respective genomes.

A high density of transposable elements was present in the genomes of *Ceratocystis* (Table 3), particularly in *C. fimbriata* (2.38 TEs/10 kb) and *C. manginecans* (2.67 TEs/10 kb), and to a lesser extent in *C. albifundus* (1.79 TEs/10 kb). Their *MAT* locus flanking regions contained on average 7.17 TEs/10 kb, 6.09 TEs/10 kb and 3.00 TEs/10 kb respectively. A moderate density of transposable elements was present in the genomes of *E. laricicola* (1.09 TEs/10 kb), *E. polonica* (1.52 TEs/10kb), *H. bhutanensis* (1.48 TEs/10 kb), *H. decipiens* (1.30 TEs/10 kb), and *H. savannae* (1.75 TEs/10 kb). The density of transposable elements in the *MAT* locus flanking regions was similarly moderate in most of these species, with the exception of *H. bhutanensis* which had a higher density (3.67 TEs/10 kb), and *E. polonica* which had a lower density (0.84 TEs/10 kb). Far lower densities of transposable elements were present in the genomes of *C. adiposa* (0.71 TEs/10 kb), *Ch. thielaviopsis* (0.15 TEs/10

kb), and *H. moniliformis* (0.88 Tes/10 kb). Densities of transposable elements in the *MAT* locus flanking regions of *C. adiposa* and *Ch. thielaviopsis* were similarly low. The *MAT* locus flanking regions of *C. albifundus, C. fimbriata, C. manginecans, H. bhutanensis* and *H. moniliformis* contained between 1.5x and 3.5x more transposable elements when compared to the rest of their genomes.

4. Discussion

Results of this study showed that the organisation and gene content of the regions immediately flanking the *MAT* locus were less syntenic amongst genera of *Ceratocystidaceae*, especially in the genus *Ceratocystis*. Overall little synteny was observed across the *MAT* locus flanking regions amongst all genera studied here. Within a genus, there was more conservation present in the *MAT* locus flanking regions between species, however, *Ceratocystis* species are rather different in this region compared to other genera in the *Ceratocystidaceae*. In addition, *Ceratocystis* species contained a higher number of transposable elements in their genomes and *MAT* locus flanking regions when compared to other genera in the *Ceratocystidaceae*.

Some species in the *Ceratocystidaceae* contained *COX, APN2, APC* and *SLA2* upstream from the *MAT* locus. In *Huntiella*, two of these genes were found on different contigs. Ascomycete *MAT* loci are usually flanked upstream by *SLA2* and downstream by *APN2, COX* and *APC* (Amselem *et al*., 2011; Aronstein *et al*., 2007; Jacobson, 2005; Kanematsu *et al*., 2007). Although unusual, it is not without precedent that these flanking genes are sometimes found in a different organisation surrounding the *MAT* locus and in some cases other genes have been inserted in the region around this locus. *Diplodia sapinea* is one such example where *APN2, COX* and *APC* are found downstream of the *MAT* locus with the upstream side containing two hypothetical genes, and *SLA2* was not identified (Bihon *et al*., 2014). In another ascomycete, *Grosmannia clavigera* has *SLA2* upstream of the *MAT* locus with only *COX* and *APN2* present downstream (Tsui *et al*., 2013). However, a number of hypothetical genes have been inserted between these genes. *Microsporum gypseum* also has an unusual organisation of genes with *SLA2, COX, APN2* and a hypothetical gene present upstream of the *MAT* locus, and another hypothetical gene present downstream (Li *et al*., 2010).

Genes upstream of the *MAT* locus differ between most genera in the *Ceratocystidaceae,* and thus little synteny is seen in this region. The importin gene and a hypothetical gene were present downstream from the *MAT* locus in *C. albifundus, C. manginecans* and *Ch. thielaviopsis* considered in this study, which were previously identified in *C. fimbriata* (Wilken

et al., 2014). The importin gene is also present near the *MAT* locus in *Neurospora crassa, Podospora anserina* and *Sordaria macrospora*, though not directly flanking the *MAT* locus as we have seen (data not shown). Wilson *et al*. (2015) identified two genes of unknown function upstream of the *MAT* locus in *H. moniliformis and H. omanensis*, which were also seen in this study in *H. decipiens* and *H. savannae.* Only one of these two genes of unknown function is homologous with *Endoconidiophora*. *Ceratocystis adiposa* was different to all species at this position but it is interesting that genes more distant to the *MAT* locus appear to be a combination of some genes present in *Huntiella* and *Endoconidiophora.* This might suggest gene loss that occurred in *Huntiella* and *Endoconidiophora* after diverging from a common ancestor shared with *C. adiposa.*

Ceratocystis species do not show much synteny, even mesosynteny (Hane *et al*., 2011), outside of the *MAT* locus and surrounding few genes. *Ceratocystis albifundus* is phylogenetically distantly related to *C. fimbriata* and *C. manginecans* (de Beer *et al*., 2014), and as such, some differences surrounding the *MAT* locus were not unexpected. Synteny is usually expected between species within a genus (Galagan *et al*., 2005; Kubicek *et al*., 2011), which was observed in *Endoconidiophora* and *Huntiella*. However, the extent of the differences between *C. albifundus* and the two other *Ceratocystis* species was quite large. *Aspergillus* species show regions of little synteny across their genomes but also share many regions of conserved synteny (Galagan *et al*., 2005). This could also be the case for *Ceratocystis* but further investigation is required to understand this situation.

Ceratocystis species contained more transposable elements in their genomes and *MAT* locus flanking regions than other genera and species that we considered. Transposable elements have previously been identified within and upstream of the *MAT* locus in other fungi and may have resulted in recombination within the *MAT* locus itself (Lengeler *et al*., 2002; Pöggeler *et al.,* 2011). In some cases, it is thought that they are responsible for the relocation of the entire *MAT* locus (Gioti *et al*., 2012; Li *et al*., 2010; Rydholm *et al*., 2007; Zaffarano *et al*., 2010). A recent study in *C. cacaofunesta* revealed that this species contains five-fold more transposable elements than *C. fimbriata* and this could mediate adaptation and evolution of *C. cacaofunesta* (Molano *et al.,* 2018). It should be noted that Molano *et al.* (2018) used a version of the *C. fimbriata* genome that was published in 2013 (Wilken *et al.*). In this study we have generated a more complete version of the *C. fimbriata* genome using different whole genome sequencing methods and assembling software than was used for the original published genome. We found that the the number of transposable elements increased, as would have been expected as a consequence of the different sequencing methodologies. It is, therefore, possible that the transposable elements

identified in *Ceratocystis* have contributed to the differences in gene content and organisation in the *MAT* flanking regions, and could be drivers of evolution in this region for these species. The fact that *Ch. thielavioides* is rather different to *Ceratocystis* outside of the immediate flanking genes also points to transposable elements having been introduced into *Ceratocystis* after they split from a common ancestor and giving rise to the very different flanking regions that we have observed. However, it is possible that the different software used to assemble the genomes could influence the number of transposable elements identified in each of the species studied here.

5. Conclusions

The typical genes usually found immediately flanking the *MAT* locus in other ascomycetes are in a different organisation in species of the *Ceratocystidaceae*s, with *COX* and *APN2* sometimes found at a different location in the genome. Overall little synteny was present in the *MAT* locus flanking regions between the different genera. Synteny was observed in the *MAT* locus flanking regions between species in the genera *Huntiella* and *Endoconidiophora*. There was a higher density of transposable elements present in these regions in comparison to the rest of the genome across all species studied here. The three *Ceratocystis* species included in this study, however, showed varied *MAT* locus flanking regions and the transposable elements present in these regions may have contributed to the variation observed in this study.

6. Acknowledgements

We thank the National Research Foundation (NRF), members of the Tree Protection Co-Operative Programme (TPCP), the Department of Science and Technology (DST)/NRF Centre of Excellence in Tree Health Biotechnology (CTHB) and SARChI chair in Fungal Genomics, South Africa for funding this research.

7. References

Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, Couloux A, Coutinho PM, de Vries RP, Dyer PS, Fillinger S, Fournier E, Gout L, Hahn M, Kohn L, Lapalu N, Plummer KM, Pradier J-M, Quévillon E, Sharon A, Simon A, ten Have A, Tudzynski B, Tudzynski P, Wincker P, Andrew M, Anthouard V, Beever RE, Beffa R, Benoit I, Bouzid O, Brault B, Chen Z, Choquer M, Collémare J, Cotton P, Danchin EG, Da Silva C, Gautier A, Giraud C, Giraud T, Gonzalez C, Grossetete S, Güldener U, Henrissat B, Howlett BJ, Kodira C, Kretschmer M, Lappartient A, Leroch M, Levis C, Mauceli E, Neuvéglise C, Oeser B, Pearson M, Poulain J, Poussereau N, Quesneville H, Rascle C, Schumacher J, Ségurens B, Sexton A, Silva E, Sirven C, Soanes DM, Talbot NJ,

Templeton M, Yandava C, Yarden O, Zeng Q, Rollins JA, Lebrun M-H, Dickman M, 2011. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics* **7**: e1002230.

Aronstein KA, Murray KD, de León JH, Qin X, Weinstock GM, 2007. High mobility group (HMG-box) genes in the honeybee fungal pathogen *Ascosphaera apis*. *Mycologia* **99**: 553-561.

Barsoum E, Martinez P, Åström SU, 2010. α3, a transposable element that promotes host sexual reproduction. *Genes and Development* **24**: 33-44.

Bihon W, Wingfield MJ, Slippers B, Duong TA, Wingfield BD, 2014. *MAT* gene idiomorphs suggest a heterothallic sexual cycle in a predominantly asexual and important pine pathogen. *Fungal Genetics and Biology* **62**: 55-61.

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.

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

Coppin E, Debuchy R, Arnaise S, Picard M, 1997. Mating types and sexual development in filamentous ascomycetes. *Microbiology and Molecular Biology Reviews* **61**: 411-428.

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

Debuchy R, Arnaise S, Lecellier G, 1993. The *mat-* allele of *Podospora anserina* contains three regulatory genes required for the development of fertilized female organs. *Molecular and General Genetics* **241**: 667-673.

Debuchy R, Coppin E, 1992. The mating types of *Podospora anserina*: Functional analysis and sequence of the fertilization domains. *Molecular and General Genetics* **233**: 113-121.

Ellison CE, Stajich JE, Jacobson DJ, Natvig DO, Lapidus A, Foster B, Aerts A, Riley R, Lindquist EA, Grigoriev IV, 2011. Massive changes in genome architecture accompany the transition to self-fertility in the filamentous fungus *Neurospora tetrasperma*. *Genetics* **189**: 55-69.

Ferreira AVB, An Z, Metzenberg RL, Glass NL, 1998. Characterization of *matA-2, matA-3* and *matA* mating-type mutants of *Neurospora crassa*. *Genetics* **148**: 1069-1079.

Flutre T, Duprat E, Feuillet C, Quesneville H, 2011. Considering transposable element diversification in de novo annotation approaches. *PLOS ONE* **6**: e16526.

Fourie A, Simpson MC, Duong TA, Barnes I, Coetzee MPA, van der Nest MA, Wingfield MJ, Wingfield BD, 2018. Genome annotation for *Ceratocystis fimbriata*: an aggressive fungal pathogen of root crops. *Standards in Genomic Science***. Submission in progress**.

Galagan JE, Calvo SE, Cuomo C, Ma L-J, Wortman JR, Batzoglou S, Lee S-I, Basturkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scazzocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Penalva MA, Oakley BR, Momany M, Tanaka T, Kumagai T, Asai K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paoletti M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA, Birren BW, 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**: 1105-1115.

Gallegos A, Jacobson DJ, Raju NB, Skupski MP, Natvig DO, 2000. Suppressed recombination and a pairing anomaly on the mating-type chromosome of *Neurospora tetrasperma*. *Genetics* **154**: 623-633.

Gioti A, Mushegian AA, Strandberg R, Stajich JE, Johannesson H, 2012. Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. *Molecular Biology and Evolution* **29**: 3215-3226.

Glass NL, Grotelueschen J, Metzenberg RL, 1990. *Neurospora crassa A* mating-type region. *Proceedings of the National Academy of Science of the United States of America* **87**: 4912-4916.

Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB, 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences* **108**: 1513-1518.

Grognet P, Bidard F, Kuchly C, Tong LCH, Coppin E, Benkhali JA, Couloux A, Wincker P, Debuchy R, Silar P, 2014. Maintaining two mating types: Structure of the mating type locus and its role in heterokaryosis in *Podospora anserina*. *Genetics* **197**: 421-432.

Guy L, Roat Kultima J, Andersson SG, 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**: 2334-2335.

Halsted BD, 1890. Some fungous diseases of the sweet potato. The black rot. *Bulletin of the New Jersey Agricultural College Experiment Station* **76**: 7-14.

Hane JK, Rouxel T, Howlett BJ, Kema GH, Goodwin SB, Oliver RP, 2011. A novel mode of chromosomal evolution peculiar to filamentous ascomycete fungi. *Genome biology* **12**: R45.

462 Harrington TC, Aghayeva DN, Fraedrich SW, 2010. New combinations in *Raffaelea,*
463 *Ambrosiella*, and *Hyalorhinocladiella*, and four new species from the redbay ambrosia
464 beetle, *Xyleborus glabratus*. *Mycotaxon* **111**: 337-361.
465 Harrington TC, McNew DL, 1997. Self-fertility and uni-directional mating-type switching in
466 *Ceratocystis coerulescens*, a filamentous ascomycete. *Current Genetics* **32**: 52-59.
467 Hsueh Y-P, Idnurm A, Heitman J, 2006. Recombination hotspots flank the *Cryptococcus*
468 mating-type locus: Implications for the evolution of a fungal sex chromosome. *PLoS*
469 *Genetics* **2**: e184.
470 Idnurm A, 2011. Sex and speciation: the paradox that non-recombining DNA promotes
471 recombination. *Fungal Biology Reviews* **25**: 121-127.
472 Jacobson DJ, 2005. Blocked recombination along the mating-type chromosomes of
473 *Neurospora tetrasperma* involves both structural heterozygosity and autosomal genes.
474 *Genetics* **171**: 839-843.
475 Kamgan N, Jacobs K, De Beer Z, Wingfield M, Roux J, 2008. *Ceratocystis* and *Ophiostoma*
476 species including three new taxa, associated with wounds on native South African
477 trees. *Fungal Diversity* **29**: 37-59.
478 Kanematsu S, Adachi Y, Ito T, 2007. Mating-type loci of heterothallic *Diaporthe* spp.:
479 homologous genes are present in opposite mating-types. *Current Genetics* **52**: 11-22.
480 Kile GA, 1993. Plant diseases caused by species of *Ceratocystis sensu stricto* and *Chalara*,
481 In: Wingfield, M., Seifert, K., Webber, F. (eds.), *Ceratocystis and Ophiostoma:*
482 *Taxonomy, ecology and pathogenicity*. APS Press, St. Paul, Minnesota, USA, pp. 173-
483 183.
484 Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M,
485 Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, 2011. Comparative genome
486 sequence analysis underscores mycoparasitism as the ancestral life style of
487 *Trichoderma*. *Genome Biology* **12**: R40.
488 Kubisiak TL, Milgroom MG, 2006. Markers linked to vegetative incompatibility (vic) genes
489 and a region of high heterogeneity and reduced recombination near the mating type
490 locus (*MAT*) in *Cryphonectria parasitica*. *Fungal Genetics and Biology* **43**: 453-463.
491 Kück P, Longo GC, 2014. FASconCAT-G: extensive functions for multiple sequence
492 alignment preparations concerning phylogenetic studies. *Frontiers in zoology* **11**: 81.
493 Lee DH, Roux J, Wingfield BD, Wingfield MJ, 2015. Variation in growth rates and
494 aggressiveness of naturally occurring self-fertile and self-sterile isolates of the wilt
495 pathogen *Ceratocystis albifundus*. *Plant Pathology* **64**: 1103-1109.
496 Lee N, Bakkeren G, Wong K, Sherwood JE, Kronstad JW, 1999. The mating-type and
497 pathogenicity locus of the fungus *Ustilago hordei* spans a 500-kb region. *Proceedings*
498 *of the National Academy of Sciences* **96**: 15026-15031.

Lee SC, Ni M, Li W, Shertz C, Heitman J, 2010. The evolution of sex: a perspective from the fungal kingdom. *Microbiology and Molecular Biology Reviews* **74**: 298-340.

Lengeler KB, Fox DS, Fraser JA, Allen A, Forrester K, Dietrich FS, Heitman J, 2002. Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes. *Eukaryotic Cell* **1**: 704-718.

Li W, Metin B, White TC, Heitman J, 2010. Organization and evolutionary trajectory of the mating type (*MAT*) locus in dermatophyte and dimorphic fungal pathogens. *Eukaryotic Cell* **9**: 46-58.

Maddison WP, Maddison DR, 2018. Mesquite: a modular system for evolutionary analysis. Available via: http://mesquiteproject.org.

Maekawa H, Kaneko Y, 2014. Inversion of the chromosomal region between two mating type loci switches the mating type in *Hansenula polymorpha*. *PLoS Genetics* **10**: e1004796.

Mathieson MJ, 1952. Ascospore dimorphism and mating type in *Chromocrea spinulosa* (Fuckel) Petch n. comb. *Annals of Botany* **16**: 449-468.

Mbenoun M, De Beer ZW, Wingfield MJ, Wingfield BD, Roux J, 2014. Reconsidering species boundaries in the *Ceratocystis paradoxa* complex, including a new species from oil palm and cacao in Cameroon. *Mycologia* **106**: 757-784.

Menkis A, Jacobson DJ, Gustafsson T, Johannesson H, 2008. The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *PLoS Genetics* **4**: e1000030.

Metzenberg RL, Glass NL, 1990. Mating type and mating strategies in *Neurospora*. *Bioessays* **12**: 53-59.

Molano EPL, Cabrera OG, Jose J, do Nascimento LC, Carazzolle MF, Teixeira PJPL, Alvarez JC, Tiburcio RA, Tokimatu Filho PM, de Lima GMA, Guido RVC, Corrêa TLR, Leme AFP, Mieczkowski P, Pereira GAG, 2018. *Ceratocystis cacaofunesta* genome analysis reveals a large expansion of extracellular phosphatidylinositol-specific phospholipase-C genes (PI-PLC). *BMC Genomics* **19**: 58.

Nelson MA, 1996. Mating systems in ascomycetes: A romp in the sac. *Trends in Genetics* **12**: 69-74.

Nkuekam GK, Wingfield MJ, Roux J, 2013. *Ceratocystis* species, including two new taxa, from Eucalyptus trees in South Africa. *Australasian Plant Pathology* **42**: 283-311.

Paoletti MA, Seymour FA, Alcocer MJC, Kaur N, Calvo AM, Archer DA, Dyer PS, 2007. Mating type and the genetic basis of self-fertility in the model fungus *Aspergillus nidulans*. *Current Biology* **17**: 1384-1389.

Philley ML, Staben C, 1994. Functional analyses of the *Neurospora crassa* MT a-1 mating-type polypeptide. *Genetics* **137**: 715-722.

Pöggeler S, O'Gorman CM, Hoff B, Kück U, 2011. Molecular organization of the mating-type loci in the homothallic Ascomycete Eupenicillium crustaceum. *Fungal biology* **115**: 615-624.

Pöggeler S, Risch S, Kück U, Osiewacz HED, 1997. Mating-type genes from the homothallic fungus *Sordaria macrospora* are functionally expressed in a heterothallic ascomycetes. *Genetics* **147**: 567-580.

Redfern DB, Stoakley JT, Steele H, Minter DW, 1987. Dieback and death of larch caused by *Ceratocystis laricicola* sp. nov. following attack by *Ips cembrae*. *Plant Pathology* **36**: 467-480.

Roux J, Van Wyk M, Hatting H, Wingfield MJ, 2004. *Ceratocystis* species infecting stem wounds on *Eucalyptus grandis* in South Africa. *Plant Pathology* **53**: 414-421.

Rydholm C, Dyer P, Lutzoni F, 2007. DNA sequence characterization and molecular evolution of *MAT1* and *MAT2* mating-type loci of the self-compatible ascomycete mold *Neosartorya fischeri*. *Eukaryotic Cell* **6**: 868-874.

Sartoris, 1927. A cytological study of *Ceratostomella adiposum* (Butl.) comb. nov., the black-rot fungus of sugar cane. *Journal of Agricultural Research* **35**: 577-585.

Seifert K, 1993. Sapstain of commercial lumber by species of Ophiostoma and Ceratocystis, In: Wingfield, M.J., Seifert, K.A., Webber, F.J. (eds.), *Ceratocystis and Ophiostoma: Taxonomy, ecology and pathogenicity*. APS Press, St. Paul, Minnesota, USA, pp. 141-151.

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

Solheim H, 1986. Species of Ophiostomataceae isolated from *Picea abies* infested by the bark beetle *Ips typographic*. *Nordic Journal of Botany* **6**: 199-207.

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

Tarigan M, Van Wyk M, Roux J, Tjahjono B, Wingfield MJ, 2010. Three new *Ceratocystis* spp. in the *Ceratocystis moniliformis* complex from wounds on *Acacia mangium* and *A. crassicarpa*. *Mycoscience* **51**: 53-67.

Tsui CK-M, DiGuistini S, Wang Y, Feau N, Dhillon B, Bohlmann J, Hamelin RC, 2013. Unequal recombination and evolution of the mating-type (*MAT*) loci in the pathogenic fungus *Grosmannia clavigera* and relatives. *G3: Genes Genomes Genetics* **3**: 465-480.

Uhm JY, Fujii H, 1983. Heterothallism and mating type mutation in *Sclerotinia trifoliorum*. *Phytopathology* **73**: 569-572.

van der Nest MA, Beirn LA, Crouch JA, Demers JE, de Beer ZW, De Vos L, Gordon TR, Moncalvo J-M, Naidoo K, Sanchez-Ramirez S, 2014a. IMA Genome-F 3: Draft genomes of *Amanita jacksonii, Ceratocystis albifundus, Fusarium circinatum, Huntiella omanensis, Leptographium procerum, Rutstroemia sydowiana,* and *Sclerotinia echinophila*. *IMA Fungus* **5**: 472-485.

van der Nest MA, Bihon W, De Vos L, Naidoo K, Roodt D, Rubagotti E, Slippers B, Steenkamp ET, Wilken PM, Wilson A, 2014b. Draft genome sequences of *Diplodia sapinea, Ceratocystis manginecans,* and *Ceratocystis moniliformis*. *IMA Fungus* **5**: 135-140.

van der Nest MA, Steenkamp ET, McTaggart AR, Trollip C, Godlonton T, Sauerman E, Roodt D, Naidoo K, Coetzee M, Wilken P, 2015. Saprophytic and pathogenic fungi in the *Ceratocystidaceae* differ in their ability to metabolize plant-derived sucrose. *BMC Evolutionary Biology* **15**: 273.

van Wyk M, Al Adawi AO, Khan IA, Deadman ML, Al Jahwari AA, Wingfield BD, Ploetz R, Wingfield MJ, 2007. *Ceratocystis manginecans* sp. nov., causal agent of a destructive mango wilt disease in Oman and Pakistan. *Fungal Divers* **27**: 213-230.

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

van Wyk PWJ, Wingfield MJ, Van Wyk PS, 1991. Ascospore development in *Ceratocystis moniliformis*. *Mycological Research* **95**: 96-103.

Wheeler HE, 1950. Genetics of *Glomerella*. VIII. A genetic basis for the occurrence of minus mutants. *American Journal of Botany* **37**: 304-312.

White CI, Haber JE, 1990. Intermediates of recombination during mating type switching in *Saccharomyces cerevisiae*. *The EMBO Journal* **9**: 663-673.

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.

Wilken PM, Steenkamp ET, Wingfield MJ, de Beer ZW, Wingfield BD, 2014. DNA loss at the *Ceratocystis fimbriata* mating locus results in self-sterility. *PLoS One* **9**: e92180.

Wilson A, Godlonton T, Van der Nest MA, Wilken PM, Wingfield MJ, Wingfield BD, 2015. Unisexual reproduction in *Huntiella moniliformis*. *Fungal Genetics and Biology* **80**: 1-9.

Wingfield BD, Ambler JM, Coetzee M, de Beer ZW, Duong TA, Joubert F, Hammerbacher A, McTaggart AR, Naidoo K, Nguyen HD, 2016a. Draft genome sequences of *Armillaria fuscipes, Ceratocystiopsis minuta, Ceratocystis adiposa, Endoconidiophora laricicola, E. polonica* and *Penicillium freii* DAOMC 242723. *IMA Fungus* **7**: 217-227.

Wingfield BD, Berger DK, Steenkamp ET, Lim H-J, Duong TA, Bluhm BH, de Beer ZW, De Vos L, Fourie G, Naidoo K, Olivier N, Lin Y-C, Van de Peer Y, Joubert F, Crampton BG, Swart V, Soal N, Tatham C, van der Nest MA, van der Merwe NA, Van Wyk S, Wilken PM, Wingfield MJ, 2017. Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus, Huntiella decipiens* and *Ophiostoma ips*. *IMA Fungus* **8**: 385-396.

Wingfield BD, Duong TA, Hammerbacher A, van der Nest MA, Wilson A, Chang R, Wilhelm de Beer Z, Steenkamp ET, Markus Wilken P, Naidoo K, 2016b. IMA Genome-F 7 Draft genome sequences for *Ceratocystis fagacearum, C. harringtonii, Grosmannia penicillata*, and *Huntiella bhutanensis*. *IMA Fungus* **7**: 317-323.

Wingfield BD, van Wyk M, Roos H, Wingfield MJ, 2013. *Ceratocystis*: Emerging evidence for discrete genetic boundaries, In: Seifert, K., de Beer, Z.W., Wingfield, M.J. (eds.), *The ophiostomatoid fundi: Expanding frontiers*. CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands, pp. 57-64.

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, Harrington TC, Solheim H, 1997. Two species in the *Ceratocystis coerulescens* complex from conifers in western North America. *Canadian Journal of Botany* **75**: 827-834.

Witthuhn RC, Harrington TC, Wingfield BD, Steimel JP, Wingfield MJ, 2000. Deletion of the *MAT-2* mating-type gene during uni-directional mating-type switching in *Ceratocystis*. *Current Genetics* **38**: 48-52.

Yamada T, Mizuno Ki, Hirota K, Kon N, Wahls WP, Hartsuiker E, Murofushi H, Shibata T, Ohta K, 2004. Roles of histone acetylation and chromatin remodeling factor in a meiotic recombination hotspot. *The EMBO Journal* **23**: 1792-1803.

Zaffarano PL, Duò A, Grünig CR, 2010. Characterization of the mating type (*MAT*) locus in the *Phialocephala fortinii* sl–*Acephala applanata* species complex. *Fungal Genetics and Biology* **47**: 761-772.

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

Table 1 Information pertaining to the genome sequences of each species used in this study.

Species	Strain / Culture collection numbers	Genome size (bp)	Contigs	Coverage	Sequencing technology	Assembly method	Genbank Accession	Publication
Ceratocystis adiposa	CMW2573, CBS136.34	28 574 449	714	93x	Illumina GAIIx	CLC Genomics Workbench, SSPACE, GapFiller	LXGU00000000	Wingfield *et al.,* 2016a
Ceratocystis albifundus	CMW4068, CBS128992	30 652 727	186	24x	Illumina GAIIx	Velvet	JSSU00000000	van der Nest *et al.*, 2014a
Ceratocystis fimbriata	CMW14799, CBS141723	30 159 987	399	630x	Illumina HiSeq	Velvet, SSPACE, GapFiller	APWK00000000	Fourie *et al.*, 2018
Ceratocystis manginecans	CMW17570, CBS138185	31 869 883	231	22x	Illumina	Velvet	JJRZ00000000	van der Nest *et al*., 2014b
Chalaropsis thielavioides	JCM1933	23 333 049	12	387x	Illumina HiSeq	ALLPATHS-LG	BCGU00000000	-
Endoconidiophora laricicola	CMW 20928, CBS100207	33 339 713	1070	93x	Illumina GAIIx	CLC Genomics Workbench, SSPACE, GapFiller	LXGT00000000	Wingfield *et al*., 2016a
Endoconidiophora polonica	CMW20930, CBS100205	33 192 964	1010	82x	Illumina HiSeq	CLC Genomics Workbench, SSPACE, GapFiller	LXKZ00000000	Wingfield *et al*., 2016a
Huntiella bhutanensis	CMW8217, CBS114289	27310290	1183	127x	Illumina GAIIx	CLC Genomics Workbench, SSPACE, GapFiller	MJMS00000000	Wingfield *et al*., 2016b
Huntiella decipiens	CMW30855, CBS129736	26 880 851	638	120x	Illumina GAIIx	CLC Genomics Workbench, SSPACE, GapFiller	NETU00000000	Wingfield *et al.* 2017
Huntiella moniliformis	CMW10134, CBS118127	25 429 610	365	38x	Illumina GAIIx	ABySS	JMSH00000000	van der Nest *et al*., 2014b
Huntiella savannae	CMW17300, CBS121151	28 599 174	2013	22x	Illumina HiSeq	Velvet, SSPACE, GapFiller	LCZG00000000	van der Nest *et al*., 2015

CMW = the Culture Collection of the Forestry and Agricultural Biotechnology Institute, University of Pretoria, South Africa. CBS = the Culture Collection (CBS) of Westerdijk Fungal Biodiversity Institute, Utrecht, the Netherlands. JCM = Japan Collection of Microorganisms.

Table 2 Proportional maximum likelihood values for mating strategies in the *Ceratocystidaceae* species studied here from ancestral state reconstruction. Starred values indicate most probable ancestral state(s).

Node	Self-fertile	Self-sterile	Asexual	Unisexual
1	0,9075*	0,0335	0,0238	0,0352
2	0,9638*	0,0071	0,0216	0,0074
3	0,8944*	0,0072	0,0911	0,0073
4	0,9909*	0,0013	0,0066	0,0013
5	0,9984*	0,0004	0,0008	0,0004
6	0,9964*	0,0009	0,0009	0,0018
7	0,8818*	0,0505	0,0136	0,0541
8	0,2955*	0,3261*	0,0253	0,3531*
9	0,0221	0,9473*	0,0049	0,0257
10	0,0018	0,9953*	0,0021	0,0008

Table 3 Transposable elements identified by REPET and their density in the whole genome and in the *MAT* locus flanking regions of the species used in this study.

Species	Genome			60 kb upstream of *MAT* locus			60 kb downstream of *MAT* locus		
	Size (bp)	TEs	Density (TEs per 10kb)	Size (bp)	TEs	Density (TEs per 10kb)	Size (bp)	TEs	Density (TEs per 10kb)
C. adiposa	28 574 449	2 036	0.71	60 000	2	0.33	60 000	7	1.17
C. albifundus	30 652 727	5 478	1.79	60 000	20	3.33	60 000	16	2.67
C. fimbriata	30 159 987	7 169	2.38	60 000	36	6.00	60 000	50	8.33
C. manginecans	31 869 883	8 507	2.67	60 000	30	5.00	60 000	43	7.17
Ch. thielaviopsis	23 333 049	352	0.15	60 000	5	0.83	60 000	1	0.17
E. laricicola	33 339 713	3 643	1.09	60 000	8	1.33	60 000	7	1.17
E. polonica	33 192 964	5 050	1.52	60 000	6	1.00	60 000	4	0.67
H. bhutanensis	27 310 290	4 054	1.48	60 000	31	5.17	60 000	13	2.17
H. decipiens	26 880 851	3 497	1.30	60 000	12	2.00	60 000	7	1.17
H. moniliformis	25 429 610	2 237	0.88	60 000	12	2.00	60 000	10	1.67
H. savannae	28 599 174	5 000	1.75	60 000	13	2.17	60 000	7	1.17

Figure 1 Phylogenetic tree of the 11 species used in this study, and the outgroup *F. circinatum*. The neighbour-joining tree was constructed in CLC Main Workbench v7.7.1 using the Kimura 80 model and 1000 bootstrap replicates (significant values indicated at the nodes) from nine gene regions; 18S, 28S, BT, ITS, MCM7, RPB1, RBP2, TEF1α, and TEF3α. The most parsimonious mating strategy from the ancestral state reconstruction analysis were supported by the maximum likelihood analysis, the most parsimonious ancestral state(s) are shown at the nodes of the tree. More than one state indicates that they are equally most parsimonious. Circled numbers at the nodes correspond with the node numbers in Table 2 in which the ancestral state likelihood values for each node are presented.

Figure 2 Comparison of the gene organisation around the *MAT* locus in the *Ceratocystidaceae* species. The yellow bar indicates where the *MAT* locus is located. The four typical *MAT* locus flanking genes are represented by differently coloured arrows; the green arrow (A) is *COX*, the purple arrow (B) is *APN2*, the orange arrow (C) is *APC*, the red arrow (D) is *SLA2*, and the dark blue arrow (E) is importin. The other genes are all represented by light blue arrows and homologous genes are numbered. The neighbour-joining phylogenetic tree was produced in CLC Main Workbench v7.7.1 (CLC Bio, Aarhus, Denmark) using the "Create Tree" function with the Kimura 80 model and 1000 bootstrap replicates. Nine genes were selected based on the results from de Beer *et al.* (2014) which included the 18S, 28S, BT, ITS, MCM7, RPB1, RBP2, TEF1α, and TEF3α.

The other genes include impB mucB samB family (1), FAD-dependent oxidoreductase superfamily (2), hypothetical (3), hypothetical (4), reverse transcriptase (5), HET-E-1 (6), ATPase (7), HET-E-1 (8), copia protease (9), Inheritance of peroxisomes 1 (10), transcription factor (11), glycosyl hydrolase family 18 (12), unknown (13), tetracycline transporter (14), DNA pantothenate metabolism flavo (15), acetyl-coenzyme A transporter 1 (16), CENP-O kinetochore centromere component (17), ABC1-domain containing(18), WW domain binding 11 (19), kinase-like domain (20), serine threonine kinase SID2 (21), RRM domain-containing (22), transporter Sec61 subunit gamma (23), coenzyme Q biosynthesis COQ4 (24), programmed cell death 5 (25), WD domain-containing (26), unknown (27), unknown (28), endonuclease reverse transcriptase (29), RNA-directed DNA polymerase from mobile element jockey (30), pre-mRNA splicing factor CWC-24 (31), assembly factor CBP4 (32), nascent polypeptide-associated complex subunit beta (33), tRNA-specific adenosine deaminase (34), 60S ribosomal L24 (35), oxidoreductase (36), regulator of G signalling superfamily (37), isovaleryl-dehydrogenase (38), carboxyltransferase (39), unknown (40), pre-mRNA-splicing factor CWC-24 (41), anucleate primary sterigmata A (42), heterokaryon incompatibility (43), unknown (44), DNA polymerase zeta catalytic subunit (45), autophagy3 (46), 60S ribosomal L37 (47), transmembrane PFT27 (48), unknown (49), unknown (50), DNA-binding TFAR19-related (51), feruloyl esterase B (52), unknown (53), unknown (54), unknown (55), methionyl-tRNA mitochondrial (56), and unknown (57).

Figure 3 Comparison of *MAT* flanking regions across the 12 species of *Ceratocystidaceae* used in this study. Yellow arrows indicate *MAT* genes, red arrows indicate genes that contain transposable elements, and // indicates separation between contigs as they could not be assembled together.

Supplementary Figure 1 Linear regression analysis to determine correlation between the number of contigs and the number of transposable elements in each genome assembly.