Unexpected placement of the *MAT1-1-2* gene in the *MAT1-2* idiomorph of *Thielaviopsis*

P. Markus Wilken*, Emma T. Steenkamp, Magriet A. van der Nest, Michael J. Wingfield, Z. Wilhelm de Beer, Brenda D. Wingfield

Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), Faculty of Natural and Agricultural Sciences, University of Pretoria, Private Bag X20, Pretoria, 0028, South Africa

* corresponding author: markus.wilken@fabi.up.ac.za

Abstract

Sexual reproduction in the Ascomycota is controlled by genes encoded at the mating-type or *MAT1* locus. The two allelic versions of this locus in heterothallic species, referred to as idiomorphs, are defined by the *MAT1-1-1* (for the *MAT1-1* idiomorph) and *MAT1-2-1* (for the *MAT1-2* idiomorph) genes. Both idiomorphs can contain additional genes, although the contents of each is typically specific to and conserved within particular Pezizomycotina lineages. Using full genome sequences, complemented with conventional PCR and Sanger sequencing, we compared the mating-type idiomorphs in heterothallic species of *Thielaviopsis* (Ceratocystidaceae). The analyses showed that the *MAT1-1* idiomorph of *T. punctulata, T. paradoxa, T. euricoi, T. ethacetica* and *T. musarum* harboured only the expected *MAT1-1-1* gene. In contrast, the *MAT1-2* idiomorph of *T. punctulata*, *T. paradoxa* and *T. euricoi* encoded the *MAT1-2-1*, *MAT1-2-7* and *MAT1-1-2* genes. Of these, *MAT1-2-1* and *MAT1-2-7* are genes previously reported in this idiomorph, while *MAT1-1-2* is known only in the *MAT1-1* idiomorph. Phylogenetic analysis showed that the *Thielaviopsis MAT1-1-2* groups with the known homologues of this gene in other Microascales, thus confirming its annotation. Previous work suggests that *MAT1-1-2* is involved in fruiting body development, a role that would be unaffected by its idiomorphic position. This notion is supported by our findings for the *MAT1* locus structure in *Thielaviopsis* species. This also serves as the first example of a *MAT1-1*-specific gene restricted to only the *MAT1-2* idiomorph.

Keywords: *MAT1-1-2*, *Thielaviopsis*, sexual reproduction, Mating-type (*MAT*) locus, heterothallic

Introduction

Understanding the genetic basis of sexual reproduction in the Pezizomycotina (Ascomycota) has been the focus of numerous studies subsequent to the identification and sequencing of the mating-type idiomorphs in *Neurospora crassa* (Glass et al., 1990, 1988; Staben and Yanofsky, 1990). As is the case in many biological systems, most of this work has focused on either model taxa (e.g. *Neurospora* and *Sordaria*), or anthropocentric species such as medically important fungi in the genus *Aspergillus* (Dyer et al., 2016; Heitman et al., 2007; Ni et al., 2011). These previous studies have provided valuable insights into the reproductive genes and their roles in the mating systems of the target fungi (Gioti et al., 2012). Knowledge gained from these studies has also provided an important foundation for studies of the sexual systems in non-model species (Bihon et al., 2014; Gioti et al., 2012; Martin et al., 2011). Consequently, a robust framework of knowledge has emerged regarding fungal mating systems and molecular processes that govern sexual reproduction (Dyer et al., 2016; Lee et al., 2010; Wilson et al., 2015b).

Most sexually reproducing fungi in the Ascomycota can be classified as either heterothallic or homothallic (Billiard et al., 2011). This phenotypic designation is widely used to describe mating systems (Wilson et al., 2015b), where heterothallism refers to the requirement for two individuals of complementary mating specificity (or mating-types) for sexual reproduction to occur (Ni et al., 2011). In contrast, a single homothallic individual can complete the sexual cycle in the absence of any other individual, producing the characteristic ascomata and ascospores (Lin and Heitman, 2007). Despite this relatively simple definition of homothallism, the phenotype can encompass a multitude of mating strategies ranging from primary homothallism to unisexual mating and mating-type switching (Lin and Heitman, 2007; Wilson et al., 2015b). The nuances of these different systems can in part be addressed by studying the genes controlling sexual reproduction (Dyer et al., 2016; Wilken et al., 2017; Wilson et al., 2015b).

In the Ascomycota, mating-type is conferred by the genes encoded at a single locus; the mating-type or *MAT1* locus (Whitehouse, 1949). Two complementary versions of the locus are known in heterothallic species, and these are referred to as idiomorphs (Metzenberg and Glass, 1990). The *MAT1-1* idiomorph is characterized by the presence of a *MAT1-1-1* gene encoding a protein with an alpha-box DNA binding motif, although other genes may also be present (Butler, 2007; Dyer et al., 2016; Wilken et al., 2017). In comparison, the *MAT1-2-1* gene encoding an HMG-box domain protein characterizes the *MAT1-2* idiomorph, although it can also include additional genes depending on the species involved (Butler, 2007; Dyer et al., 2016; Wilken et al., 2017). In homothallic species, the genes in the two idiomorphs of

heterothallic species are generally present in a single haploid genome, thus conferring self-fertility. Currently, 20 *MAT*-specific gene names have been assigned to the mating-type genes of Pezizomycotina species, nine of which are present in the *MAT1-1* idiomorph (*MAT1-1-1* to *MAT1-1-9*) and eleven in the *MAT1-2* idiomorph (*MAT1-2-1* to *MAT1-2-11*) (Wilken et al., 2017). The role of many of the mating-type genes is not well-understood, although limited studies have shown that some of these genes may regulate downstream processes crucial to sexual reproduction. These functions include the production of peptide pheromones and their corresponding pheromone receptors, the regulation of internuclear recognition and the development of ascogenous hyphae and other functions related to fruiting body development (Debuchy et al., 2010; Dyer et al., 2016; Turgeon and Debuchy, 2007; Yun et al., 2017).

This study considered the mating-type genes of *Thielaviopsis*, a genus in the Ceratocystidaceae (order Microascales) that includes important pathogens of monocotyledonous plants (De Beer et al., 2014). Previous studies suggest that the diversity of reproductive strategies in the Ceratocystidaceae is reflected in the structure of their *MAT1* locus (Wilken et al., 2014; Wilson et al., 2015a; Witthuhn et al., 2000). For example, in *Ceratocystis fimbriata*, self-fertility is conferred by the presence of three mating-type genes (*MAT1-1-1*, *MAT1-2-1,* and *MAT1-1-2*) at the *MAT1* locus (Wilken et al., 2014). Self-sterile isolates of this species arise from a putative recombination event at the *MAT1* locus that results in the deletion of the *MAT1-2-1* gene (Wilken et al., 2014). This process has been termed uni-directional mating-type switching (Webster and Butler, 1967; Witthuhn et al., 2000) and appears to be a feature of all species of *Ceratocystis* (Wilken et al., 2014; Witthuhn et al., 2000).

In *Huntiella* (Ceratocystidaceae) robust information regarding mating is available only for two species, the heterothallic *H. omanensis* and the homothallic *H. moniliformis* (Wilson et al., 2015a). Self-fertility in *H. moniliformis* arises from a form of secondary homothallism (Lin and Heitman, 2007) known as uni-sexual mating, where the *MAT1* locus of the studied isolates contains only a *MAT1-2-1* gene (Wilson et al., 2015a). For other genera in the Ceratocystidaceae, including *Thielaviopsis*, only phenotypic information is available regarding the mating-type systems. Mating studies in culture have identified three heterothallic (*T. punctulata*, *T. paradoxa* and *T. ethacetica*) and one homothallic (*T. cerberus*) species, with no sexual state known for the remaining two species (*T. musarum* and *T. euricoi*) (Dade, 1928; El-Ani et al., 1957; Mbenoun et al., 2014).

Genome sequences have recently become publicly available for several species in the Ceratocystidaceae (Van der Nest et al., 2014a, 2014b; Wilken et al., 2013; Wingfield et al., 2015a, 2015b, 2016a, 2016b). These include those for *T. punctulata* (Wingfield et al., 2015a),

T. musarum (Wingfield et al., 2015b), *T. euricoi* (JCM-Riken BioResource Center, 2016a) and *T. ethacetica* (JCM-Riken BioResource Center, 2016b). The availability of these sequences provides an opportunity to study the molecular basis of mating in *Thielaviopsis*. Our aim was, thus, to identify and characterize the *MAT1* locus from the available genomes of these fungi. We also complemented these existing genomic resources with genome data for three additional *Thielaviopsis* isolates including a second isolate of *T. punctulata*, as well as an isolate each of *T. paradoxa* and *T. euricoi*. The distribution of the mating-type genes inferred from the genomes was also confirmed, using conventional PCR and Sanger sequencing, in multiple isolates of *T. musarum*, *T. punctulata*, *T. paradoxa, T. euricoi* and *T. ethacetica*.

Materials and methods

Genomes and genome sequencing

The genome sequences for seven *Thielaviopsis* isolates were used. These included the available genome sequences of four isolates that were retrieved from NCBI. These were the genomes of *T. punctulata* isolate CRDP1 (accession number LAEV01000000; Wingfield et al., 2015a), *T. musarum* isolate CMW1546 (accession number LKBB01000000; Wingfield et al., 2015b), *T. euricoi* isolate JCM6020 (accession number BCHJ01000000; JCM-Riken BioResource Center, 2016a) and *T. ethacetica* isolate JCM6961 (accession number BCFY01000000; JCM-Riken BioResource Center, 2016b). Although the latter two isolates were initially deposited into the GenBank database as *T. paradoxa*, comparison of their tubulin (β-tubulin) and translation elongation factor 1-alpha (TEF-1α) sequences (data not shown) confirmed conspecificity with the respective species proposed by Mbenoun et al. (2014). To complement these public domain sequences, we also determined the whole genome sequences for a putative MAT1-1 isolate (CMW1032) of *T. punctulata*, a putative MAT1-2 isolate (CMW36654) of *T. paradoxa* and a *T. euricoi* isolate (CMW8799) of unknown mating type (Table 1; Mbenoun et al., 2014), preserved in the culture collection CMW of the Forestry and Agricultural Biotechnology Institute (FABI) at the University of Pretoria, South Africa.

To sequence the genome for isolate CMW1032 of *T. punctulata* (Table 1; Mbenoun et al., 2014), the fungus was grown for two weeks on 2 % MEA-TS media (20 g L^{-1} Malt Extract, 20 g L^{-1} Agar [Biolab, Merck], 100 mg L^{-1} Thiamine, 150 ml L^{-1} Streptomycin [SIGMA, Steinheim, Germany]) at 25 °C before being subjected to DNA extraction as described previously (Roux et al., 2004). Genome sequencing was carried out on the Illumina Genomics Analyzer llx platform at the University of California at Davis Genome Centre, California, USA. For sequencing, paired-end libraries with insert fragments of 350 and 600 bases were prepared and used to produce reads of a 100-base target length. These sequences were filtered by

removing poor quality reads (threshold limit of 0.05) and/or terminal residues using the CLC Genomics Workbench v. 8.5.1 (Qiagen Aarhus, Denmark). The same software package was used to *de novo* assemble a draft genome sequence with the following parameters: a word size of 64 and a bubble size of 100 base pairs.

To improve the quality of the assembly, multiple rounds of contig extension and scaffolding were performed using both SSPACE (Boetzer et al., 2011) and GapFiller (Boetzer and Pirovano, 2012), as well as the "join contigs" option of the genome finishing module implemented in the CLC Genomics Workbench (Supplementary table 1). A quantitative assessment of the genome assembly completeness was performed with the BUSCO (Benchmarking Universal Single-Copy Orthologs) program (Simão et al., 2015) using contigs greater than 500 bases in length. This subset of contigs was also submitted to the GenBank database of the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) as the genome sequence of this isolate.

For low-coverage sequencing of the *T. paradoxa* isolate CMW36654 and *T. euricoi* isolate CMW8799 (Mbenoun et al., 2014), the fungi were grown as described above. DNA was then extracted using the QIAamp DNA Mini Kit (Qiagen, USA) following the manufacturer's instructions. The extracted DNA was used for whole genome shotgun sequencing at the Central Analytical Facilities (CAF) of Stellenbosch University. Single reads were produced using the 200 bp chemistry and the 318-chip Ion Torrent PGM (Thermo Fisher Scientific, USA) platform. The raw sequences were imported into the CLC Genomics Workbench 9.5.4, filtered by removing poor quality reads (threshold limit of 0.05) and/or terminal residues, and then used in a *de novo* genome assembly allowing the program to estimate the optimal word and bubble sizes. The raw data produced were deposited in the NCBI Sequence Read Archive (SRA) database.

The mating-type genes in *T. punctulata*, *T. musarum, T. paradoxa, T. euricoi* and *T. ethacetica*

The genomes for the seven *Thielaviopsis* isolates were imported into the CLC Main Workbench version 7.9.1 to generate seven genome databases. Each of the databases was screened for the presence of mating-type genes using a tBLASTn search with the *MAT* proteins previously identified in *C. fimbriata* (NCBI accession numbers AHV84685, AHV84686 and AHV84687; Wilken et al., 2014) and *H. omanensis* (NCBI accession numbers AOY41710, AOY41711, AOY41712 and AOY41713; Wilson et al., 2015a). These included the putative mating-type proteins MAT1-1-1, MAT1-1-2 and MAT1-2-1 present in both *C. fimbriata* and *H. omanensis*, as well as MAT1-2-7 present only in the *Huntiella* species. These tBLASTn searches were carried out using CLC Main workbench, and contigs that produced matches

(expect [e]-values ≤ 0.01) were subjected to *de novo* gene prediction using the online version of the AUGUSTUS Gene Prediction software (Stanke et al., 2006). All predictions utilized the *Fusarium graminearum* gene models (the closest available relative of *Thielaviopsis*), with the program parameters set to predict any number of possibly partial genes on both strands. The same contigs were also submitted to the Fgenesh *ab initio* predictor (Solovyev et al., 2006) using the generic *Fusarium* gene models to identify any genes that might not have been predicted by AUGUSTUS. All resulting gene models were then used to annotate the respective contigs. The predicted coding sequences were translated to amino acid sequences and submitted to the BLASTp server at NCBI for identification. To identify conserved domains that might be present, all predicted MAT1-1 and MAT1-2 protein sequences were subjected to a Pfam domain analysis using CLC Main Workbench and the full Pfam database version 29 (Finn et al., 2014).

Regions that showed sequence similarity between the mating-type and flanking regions of the two *T. punctulata* isolates, as well as between *T. punctulata* and *C. fimbriata* (accession number KF033902; Wilken et al., 2014) were identified using BLASTn searches. To do this, these regions were subjected to BLASTn searches implemented in the BLAST+ software package with default settings (Camacho et al., 2009). The hit tables produced were used to plot the identity onto the respective genomic regions using the GenoPlotR package (Guy et al., 2010).

Phylogenetic analysis of the *Thielaviopsis* mating-type genes

The sequences of the full MAT1-1-1, MAT1-2-1 and MAT1-1-2 proteins predicted from the genomic gene sequences were compared with homologues of these mating-type genes from other Sordariomycetes (Supplementary table 2) by constructing a maximum likelihood tree for each individual dataset. Each dataset was first aligned using the online MAFFT server (Katoh and Standley, 2013), before being imported into Prottest v3.4 (Darriba et al., 2011; Guindon and Gascuel, 2003) to determine the most appropriate evolutionary model and parameters to use for the analysis. A maximum likelihood tree was then constructed using the PhyML v3.1 package (Guindon et al., 2010) and the appropriate model parameters. Branch support was evaluated using the same model parameters and 1 000 bootstrap replicates. The resulting tree was visualised and edited using TreeGraph v2.14.0-771 (Stöver and Müller, 2010).

PCR amplification of *MAT* genes from isolates of five *Thielaviopsis* species

The mating-type genes identified from the genomes of *T. punctulata*, *T. musarum*, *T. paradoxa, T. euricoi* and *T. ethacetica* were used to design nine primer sets targeting fragments of the individual genes in each species. To do this, each predicted gene was submitted to the online primer design tool Primer3web (Koressaar and Remm, 2007;

Untergasser et al., 2012) for designing (based on default settings) forward and reverse primers. This approach was used to design a primer set each for: the *MAT1-1-1* gene in *T. punctulata*, *T. euricoi* and *T. ethacetica* and the *MAT1-2-1*, *MAT1-2-7* and *MAT1-1-2* genes in *T. punctulata* and *T. paradoxa* (Table 2).

The designed mating-type primers were used to investigate the presence and distribution of the *MAT* genes in a collection of *Thielaviopsis* isolates that was available to us (Table 1). This included six isolates representing *T. punctulata* (including CMW1032 used for whole genome sequencing), four isolates of *T. euricoi* (including CMW8799 subjected to low-level sequencing in this study), four isolates of *T. ethacetica* and the single known isolate of *T. musarum*. Four isolates that had previously been identified as *T. paradoxa sensu stricto* (including CMW36654 that was used for low-level genome sequencing in the present study) were also used. All but one of these isolates were used in a taxonomic study of species in the *T. paradoxa* complex (Mbenoun et al., 2014), while isolate CMW37952 was used in a study by Suleman et al (2001). *Thielaviopsis punctulata* isolate CRDP1, *T. euricoi* isolate JCM6020 and *T. ethacetica* isolate JCM6961 were represented in this study by only genomic sequences, and no cultures were available for use in the PCR analyses.

The fungi were grown as described above and their DNA was extracted using the method described by Damm et al. (2008). The extracted DNA was used as the template in PCRs targeting the various mating-type genes. All reactions were conducted in 25 µl volumes that included 1U KapaTaq DNA polymerase (Kapa Biosystems, USA), 1X KapaTaq Buffer A supplemented with 1 mM $MgCl_2$, 0.25 mM of each dNTP, 0.4 mM of each primer (see below) and 90-150 ng of DNA. A touch-down PCR amplification protocol (Don et al., 1991; Korbie and Mattick, 2008) was carried out as follows: initial denaturation at 95 °C for 3 min, followed by 10 cycles of 95 °C for 30 seconds, 60 °C for 30 seconds (with a 1 °C decrease in temperature per cycle) and 72 °C for 1 min, followed by another 30 cycles of 95 °C for 30 seconds, 50 °C for 30 seconds and 72 °C for 1 min, and a final extension at 72 °C for 7 min. All PCR products were visualised under UV light after electrophoresis on a 1 % ($w.v^{-1}$) agarose (LE Agarose, SeaKem, USA) gel (Sambrook and Russell, 2001). Representative PCR products were purified using Sephadex G50 columns (Sigma-Aldrich, USA), and sequenced using the BigDye Terminator Cycle Sequencing Kit v3.1 (Life Technologies, USA), the original PCR primers and an ABI PRISM 3300 Genetic Analyser (Applied Biosystems, USA) at the Bioinformatics and Computational Biology Unit of the University of Pretoria. The electropherograms were assembled into contigs in CLC Main Workbench. This software was also used to compare the sequences of the individual PCR products with those of the mating-type genes identified from the *Thielaviopsis* genomes.

Results

Genomes and genome sequencing

Sequencing the DNA of the MAT1-1 *T. punctulata* isolate CMW1032 produced reads having 101 bp average length, split into two libraries. The 350 bp pair-end library contained 14 367 376 raw sequence pairs, while the 600 bp pair-end library included 19 611 914 pairs. After trimming for quality and ambiguous nucleotides, 14 345 406 and 19 559 958 paired reads were retained in the respective libraries, with an average read length of 95.78 bp. The initial draft genome sequence assembled using CLC Main Workbench contained 2 370 contigs (Supplementary table 1) that were produced from 99.24 % of the available reads. These contigs were ordered into 1 016 scaffolds (767 larger than 500 bp) using multiple rounds of scaffolding through SSPACE and the Genome Finishing Module in CLC Genomics Workbench (Supplementary table 1).

The assembly produced a high-quality draft genome sequence when considering general genome statistics. It had an average coverage of 111x, N50 value of 70 143 nucleotides, with the largest contig being 348 141 nucleotides in size. The genome was 28 028 969 nucleotides long (as assessed through the summation of all scaffold sizes), with a GC content of 48.13 %. The BUSCO analysis indicated a 98 % completeness score for the assembly, with 1 336 full single-copy BUSCOs, 73 complete duplicated BUSCOs, 25 fragmented BUSCOs and 4 missing BUSCOs. This Whole Genome Shotgun project was deposited at DDBJ/ENA/GenBank under the accession MJMR00000000. The version described in this paper is MJMR01000000.

The low-level genome sequencing of the MAT1-2 *T. paradoxa* isolate CMW36654 and MAT1-1 *T. euricoi* isolate CMW8799 produced 3 930 339 and 4 355 082 single raw reads, respectively. All the reads were retained for assembly but trimmed to an average length of 167.2 bp and 165.2 bp respectively to improve quality. The reads for *T. paradoxa* isolate CMW36654 were assembled into a 25.9 Mb genome consisting of 9 512 contigs with an N50 size of 3 998 bp and a coverage of 23x. The 26.1 Mb genome produced from the *T. euricoi* reads were contained in 7 891 contigs, which had a N50 value of 5 075 and a coverage of 25x. The raw sequence data for these isolates were deposited in the NCBI Sequence Read Archive (SRA) database under the accession number SRP115856.

The mating-type genes in *T. punctulata*, *T. musarum, T. paradoxa, T. euricoi* and *T. ethacetica*

The tBLASTn analyses consistently identified a single scaffold for each *Thielaviopsis* isolate with apparent similarity to one or more of the *C. fimbriata* and/or *H. omanensis* mating-type

proteins. Contig LAEV01001793 from the previously published *T. punctulata* CRDP1 genome and contig 6200 from *T. paradoxa* isolate CMW36654 (sequenced in this study) showed significant similarity to the MAT1-1-2 and MAT1-2-1 gene products of both *C. fimbriata* and *H. omanensis* (e-values ranging from 2.68E-21 to 4.25E-68), as well as to the *H. omanensis* MAT1-2-7 gene product (e-values ranging from 3.12E-34 to 2.37E-37). No contigs of either genome showed any similarity to the MAT1-1-1 gene product of either *C. fimbriata* or *H. omanensis*. Based on the presence of a *MAT1-2-1* gene and the absence of a *MAT1-1-1* gene (Wilken et al., 2017), these scaffolds were designated as containing the *MAT1-2* idiomorphs.

Contig 0000099 from the *T. punctulata* CMW1032 genome, contig LKBB01000260 from the *T. musarum* CMW1546 genome, contig BCFY01000016 from the *T. ethacetica* JCM6961 genome, contig BCHJ01000017 from the *T. euricoi* JCM6020 genome and contig6445 from the *T. euricoi* CMW8799 genome showed strong similarity (all with e-values ≤ 2.25 E-19) to both the *C. fimbriata* and *H. omanensis* MAT1-1-1 gene products. These scaffolds were accordingly designated as containing the *MAT1-1* idiomorph based on the presence of the archetypal *MAT1-1-1* gene (Wilken et al., 2017). No contig from any of these genomes showed any similarity to either the *C. fimbriata* or *H. omanensis* MAT1-1-2, MAT1-2-1 or MAT1-2-7 gene products.

De novo predictions using AUGUSTUS and Fgenesh produced annotations for multiple genes on all the contigs identified as containing the *MAT1-1* and *MAT1-2* idiomorphs. Among these were several genes known to be mating-type genes and typically present within the *MAT1* locus (Figure 1) of Sordariomycetes (Wilken et al., 2017). The contig from *Thielaviopsis punctulata* isolate CMW1032 contained only a *MAT1-1-1* homologue (1 292 bp), while that of *T. punctulata* (CRDP1) harboured the *MAT1-1-2* (1 269 bp), *MAT1-2-1* (1 018 bp) and *MAT1-2-7* (662 bp) genes. Similarly, the contigs from *T. euricoi* isolate JCM6020, *T. ethacetica* isolate JCM6961, *T. musarum* isolate CMW1546 and *T. euricoi* isolate CMW8799 contained only a *MAT1-1-1* homologue of, respectively, 1 266 bp, 1 270 bp, 1 276 bp and 1 266 bp in size. The contig of *T. paradoxa* isolate CMW36654 had homologues of the *MAT1-2-1* (979 bp), *MAT1-2-7* (673 bp), and *MAT1-1-2* (1 587 bp) genes. The annotated *Thielaviopsis* contigs containing the mating-type gene information were submitted to either the GenBank database under accession numbers KX989056, MF476807 and MF476808 or to the Third Party Annotation Section of the DDBJ/ENA/GenBank databases under the accession numbers TPA: BK010318-BK010321.

The presence of two full-length idiomorphs in the *T. punctulata* isolates made it possible to characterise the boundaries of the *MAT1* locus. The *MAT1-2* idiomorph in *T. punctulata* isolate CRDP1 is 4 700 bp in size and included the *MAT1-2-1*, *MAT1-1-2,* and *MAT1-2-7* genes

(Figure 2). The *MAT1-1* idiomorph in the *T. punctulata* isolate CMW1032 sequenced in this study was only 2 813 bp in size with a single *MAT1-1-1* gene (Figure 2). As expected for the idiomorphs of heterothallic species (McGuire et al., 2001; Turgeon, 1998; Wilken et al., 2017), sequence identity between the mating-type regions of the two *T. punctulata* isolates was restricted to the flanking regions of the locus, with no significant similarity within the mating-type idiomorphs (Figure 2). Similarly, a low level of sequence identity was present when the two *T. punctulata* idiomorphs were compared to the *MAT1* locus and flanking region of the homothallic *C. fimbriata* (Wilken et al., 2014), with similarity restricted to only to the genes predicted to flank the mating-type region (Figure 2).

The putative *MAT1-1-1* genes from the MAT-1 isolates of *T. punctulata* (CMW1032), *T. musarum* (CMW1546), *T. euricoi* (CMW8799 and JCM6020) and *T. ethacetica* (JCM6961) contained a conserved MATalpha_HMGbox domain (pfam domain PF04769) spanning the single intron present in this gene. The *MAT1-2-1* gene in *T. punctulata* isolate CMW1032 and *T. paradoxa* isolate CMW36654 harboured an HMG-box domain (pfam domain PF00505). For both the *MAT1-2-1* genes, an intron spanned a codon coding for a serine residue that forms part of the HMG-box domain, a characteristic trait of the Pezizomycotina *MAT1-2-1* gene (Arie et al., 1997; Debuchy and Turgeon, 2006; Wilken et al., 2017). The *MAT1-1-2* genes of *T. punctulata* isolate CRDP1 and *T. paradoxa* isolate CMW36654 contained a mating-type protein 1-1-2 domain (pfam domain PF17043), which did not span any of the predicted introns. The presence of the respective domains in the MAT1-1-1, MAT1-2-1 and MAT1-1-2 proteins is considered characteristic of these mating-type genes (Dyer et al., 2016; Wilken et al., 2017), adding further confidence to the identity of the predictions. Although no conserved domain was predicted for MAT1-2-7 of *T. punctulata* CRDP1 or *T. paradoxa* CMW36654, the BLASTp analysis confirmed their similarity (e-value ≤ 2.12E-45) to the MAT1-2-7 protein of *H. omanensis* (Wilson et al., 2015a).

Phylogenetic analysis of the *Thielaviopsis* mating-type genes

Phylogenetic comparison of the *Thielaviopsis* MAT1-1-1, MAT1-2-1 and MAT1-1-2 sequences inferred from the genome sequences with those present in other Sordariomycetes provided further support for the identities of these proteins (Figure 3). Each of these proteins grouped together with homologues of the respective proteins from *C. fimbriata, H. omanensis* and *H. moniliformis*. In all the phylogenetic analyses, species in the Ceratocystidaceae consistently grouped sister to the two *Knoxdaviesia* species in the Gondwanamycetaceae (Réblová et al., 2011), producing a single clade that contained all known mating-type gene sequences for the Microascales.

PCR amplification of *MAT* genes from isolates of five *Thielaviopsis* species

Using the primer pairs designed from the various *Thielaviopsis MAT1* gene sequences (Table 2), fragments of the representative mating-type genes were amplified from all isolates tested. In agreement with the genome sequence, only a *MAT1-1-1* fragment was amplified from the single known isolate (CMW1546) of *T. musarum* (Mbenoun et al., 2014). Similarly, we amplified only *MAT1-1-1* gene fragments from the three *T. ethacetica* isolates (CMW36725, CMW36771 and CMW36775) that were tested. No fragments of *MAT1-1-2*, *MAT1-2-1* or *MAT1-2-7* were amplified from either the *T. musarum* or *T. ethacetica* isolates. Six *T. punctulata* isolates were screened for mating-type genes, of which five (CMW1032, CMW26389, CMW42889, CMW42890 and CMW42891) produced the *MAT1-1-1* fragment, while parts of the *MAT1-2-1*, *MAT1-2-7* and *MAT1-1-2* genes were amplified in the sixth isolate (CMW37952). Of the four *T. paradoxa* isolates investigated, representative fragments of the *MAT1-2-1, MAT1-2-7* and *MAT1-1-2* genes were amplified from three isolates (CMW36654, CMW36689 and CMW36642). The remaining *T. paradoxa* isolate (CMW36655) contained only the *MAT1-1-1* gene. Similarly, one isolate of *T. euricoi* (CMW28538) contained the *MAT1-2-1*, *MAT1-2-7* and *MAT1-1-2* genes, while only the *MAT1-1-1* gene was present in the remaining three isolates (CMW8790, CMW8799 and CMW28537).

Discussion

Results of this study provide the first insights into the molecular basis of mating strategies in species of *Thielaviopsis*. This genus includes numerous important pathogens of mainly monocotyledonous plants (Abdullah et al., 2009; Bliss, 1941; Mbenoun et al., 2014), yet very little is known regarding their biology. Although studies on model species have framed our understanding of many complex biological processes (Yarden, 2016), it is increasingly apparent that knowledge based on these models substantially underestimates the diversity of the *MAT1* locus structure and its gene content (Dyer et al., 2016; Wilken et al., 2017). Our findings emphasise this view and illustrate the power of genome-based studies on non-model species to reveal the range of molecular idiosyncrasies potentially underlying fungal mating systems.

The *MAT1* locus of five *Thielaviopsis* species was characterized using a combination of genome sequencing and conventional PCR. Genome analyses allowed identification of a *MAT1-1* idiomorph in *T. punctulata*, *T. euricoi*, *T. ethacetica* and *T. musarum*. In all cases, it contained the *MAT1-1-1* gene, typical of other heterothallic Sordariomycetes (Dyer et al., 2016; Wilken et al., 2017). A *MAT1-2* idiomorph identified in an isolate each of *T. punctulata* and *T. paradoxa* contained three genes (i.e., *MAT1-2-1*, *MAT1-2-7* and *MAT1-1-2*). Of these,

MAT1-2-1 and *MAT1-2-7* are typically encountered in the *MAT1-2* idiomorphs of other Sordariomycetes including members of the Microascales (Aylward et al., 2016; Dyer et al., 2016; Wilken et al., 2017, 2014; Wilson et al., 2015a). However, the *MAT1-1-2* gene has been considered as exclusively present in the *MAT1-1* idiomorph (Dyer et al., 2016; Wilken et al., 2017), although a homologue of this gene is present in both idiomorphs of heterothallic *Diaporthe* species (Kanematsu et al., 2007). PCR analysis confirmed that the unique *Thielaviopsis* idiomorph organisation was reflected in isolates of the five species. This was expected as the *MAT1* locus structure is commonly conserved among closely related species (Duong et al., 2013; Martin et al., 2011; Yun et al., 2000). As such, it is plausible to expect that MAT1-2 isolates of *T. ethacetica* and *T. musarum* would show the same gene organization as the *MAT1-2* idiomorphs of *T. punctulata*, *T. paradoxa* and *T. euricoi*.

To the best of our knowledge, this study presents the first report of a *MAT1-1-2* gene exclusively present at the *MAT1-2* idiomorph of a heterothallic species. The predicted *Thielaviopsis* MAT1-1-2 protein showed the characteristic domain associated with this protein in other Ascomycota. This domain has been defined as the HPG (Debuchy and Turgeon, 2006) or PPF (Kanematsu et al., 2007) domain based on the presence of conserved amino acids in a limited number of species. Wilken et al (2014) reported that the common Proline in the H**P**G and P**P**F motifs was replaced by a tyrosine yielding a unique HYF/PYF motif in *C. fimbriata*, and suggested that this motif might be conserved among all Microascales. That view was in part supported by the presence of the domain in *H. omanensis* (Wilson et al., 2015a). However, in the predicted MAT1-1-2 proteins of *T. punctulata* and *T. paradoxa,* the tyrosine in question was replaced by an isoleucine residue. Nonetheless, the identity of the MAT1-1-2 domain was confirmed based on the newly-developed Pfam model for this protein (Dyer et al., 2016; Finn et al., 2016; Wilken et al., 2017), as well as by a gene phylogeny in which these MAT1-1-2 proteins grouped with those of other Sordariomycetes. This suggests that the putative *MAT1-1-2* gene in *Thielaviopsis* likely encodes a functional MAT1-1-2 protein.

Despite being encoded on the *MAT1-2* idiomorph of *Thielaviopsis*, our designation of "*MAT1-1-2*" was based on the similarity of this gene to the *MAT1-1-2* genes used by Turgeon and Yoder (2000) to name this gene. These authors designated the *MAT1-1-2* name based on the *N. crassa MAT A-2* homologue, using its consistent association with the *MAT1-1* idiomorph, which was the second mating-type gene (*MAT1-1-**2***) present at the *MAT**1-1*** idiomorph. Although the standardised nomenclatural system for naming mating-type genes (Turgeon and Yoder, 2000; Wilken et al., 2017) dictates the use of a *MAT1-2*-specific name for the *Thielaviopsis MAT1-1-2* gene, strict application of the system would likely cause unnecessary confusion. This is because homologues of the same gene would be called both *MAT1-1-2* and *MAT1-2-12* (the next sequential *MAT1-2* gene name; Wilken et al., 2017). Therefore, following

the recently revised nomenclatural guidelines for naming mating-type genes (Wilken et al., 2017), this *Thielaviopsis MAT* gene is designated as a *MAT1-2*-specific version of *MAT1-1-2*.

No definitive molecular function has been assigned to the MAT1-1-2 protein, although its biological role in the mating process has been studied. Deletion analyses have shown that MAT1-1-2 has a crucial role in controlling the development of the fruiting body for both homothallic and heterothallic Sordariomycetes (Arnaise et al., 2001; Dyer et al., 2016; Klix et al., 2010; Zheng et al., 2013). Such a general role in the sexual process would therefore not be idiomorph-specific, and could account for the presence of *MAT1-1-2* in the *Thielaviopsis MAT1-2* idiomorph. This view is congruent with the results of studies of the *SMR1* gene in the heterothallic *Podospora anserina*, which showed that this *MAT1-1-2* homologue was not a true mating-type gene (Arnaise et al., 1997). Through inter-nuclear complementation tests, the authors showed that *SMR1* would fulfil its role in the sexual process regardless of the mating-type idiomorph in which it was located. Such a function is in stark contrast to other mating-type genes (i.e. *MAT1-1-1* and *MAT1-2-1*) that are involved in nuclear identity, necessitating idiomorph-specific expression (Arnaise et al., 1997).

Formally, a *MAT* gene name is assigned to any gene present within the confines of the *MAT1* locus (Turgeon and Yoder, 2000; Wilken et al., 2017), although this might not necessarily reflect a role in the sexual process. This nomenclature system has been widely applied to members of the subdivision Pezizomycotina for more than 15 years (reviewed by Wilken et al., 2017), and has proven useful for the delineation of *MAT* genes from related taxa using known *MAT* gene sequences (Duong et al., 2013; Wilken et al., 2012; Wilken et al., 2014; Wilson et al., 2015a). However, a function in the sexual process (which is implied when a mating-type-specific name is assigned) has not yet been identified for many of the known *MAT* genes (Dyer et al., 2016; Wilken et al., 2017). What is known, however, is that certain *MAT* genes (notably *MAT1-1-1* and *MAT1-2-1*) are master regulators of sexual reproduction (Debuchy et al., 2010), and as such their position within opposite mating-type idiomorphs are crucial. Many of the other *MAT* genes have a secondary function (Debuchy et al., 2010; Dyer et al., 2016; Turgeon and Debuchy, 2007), and their role likely will not be idiomorph specific. Rather, their constant association with a particular idiomorph might be a reflection of the ancestral *MAT* gene arrangement.

The results of this study suggest that *T. musarum* and *T. euricoi* have heterothallic mating systems. Previously, the mating system of four *Thielaviopsis* species (*T. punctulata, T. paradoxa, T. ethacetica* and *T. cerberus*) was established using mating experiments in culture (El-Ani et al., 1957; Mbenoun et al., 2014), although this approach was not useful for assigning mating types to isolates of *T. musarum* and *T. euricoi* (Mbenoun et al., 2014). A *MAT1-1*

idiomorph was present in the genome sequences of the *T. musarum* and *T. euricoi* isolates studied here, while the *MAT1-2*-related genes were amplified from an additional isolate of *T. euricoi*. Therefore, identification of opposite idiomorphs of the *MAT1* locus in both species suggests that these fungi are heterothallic. Based on this view, characterisation of the *T. musarum* and *T. ethacetica MAT1* locus in a compatible mating partner to the *MAT1-1* isolate would likely yield a *MAT1-2* idiomorph.

The results of previous studies combined with the current findings underscore the complex evolutionary history of sexual reproduction in the Microascales. The *MAT1-2-7* gene is currently known only in the Microascales (Aylward et al., 2016; Wilson et al., 2015a), and regions of identity between this protein and MAT1-1-1 have been identified in *Knoxdaviesia* species (Aylward et al., 2016). Limited identity is also present between these proteins in *Thielaviopsis* (Supplementary figure 1), suggesting that *MAT1-2-7* arose from a *MAT1-1-1* ancestral gene in the progenitor of all Microascales (Aylward et al., 2016). Following this view, it was presumably retained in the *MAT1* locus of *Knoxdaviesia* (Aylward et al., 2016), *Huntiella* (Wilson et al., 2015a) and *Thielaviopsis*, but was lost in *Ceratocystis* (Wilken et al., 2014). An aberrant cross-cover event between the two idiomorphs of a heterothallic ancestor to *Thielaviopsis* could have produced the unusual *MAT1* locus structure currently seen (Figure 4). Such a cross-over would have relocated the *MAT1-1-2* gene from the *MAT1-1* to *MAT1-2* idiomorph. Characterising the locus structure and gene complement of homothallic species such as *T. cerberus* (Mbenoun et al., 2014) could provide insight into the processes that shaped the *MAT1* locus and associated genes in *Thielaviopsis* and other genera in the Microascales.

Acknowledgements

We acknowledge the assistance of Dr. Kershney Naidoo in sequencing of the *Thielaviopsis punctulata* CMW1032 genome. This work was supported by the Claude Leon Foundation in the form of Postdoctoral Fellowship to PMW. This project was financed by the University of Pretoria and the Department of Science and Technology (DST)/National Research Foundation (NRF) Centre of Excellence in Tree Health Biotechnology and the SARCHI chair in Fungal Genomics. The Grant holders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are those of the author(s), and that the NRF accepts no liability whatsoever in this regard.

References

Abdullah, S.K., Asensio, L., Monfort, E., Gomez-Vidal, S., Salinas, J., Lorca, L.V.L., Jansson, H.B., 2009. Incidence of the two date palm pathogens, *Thielaviopsis paradoxa* and *T. punctulata* in soil from date palm plantations in Elx, South-East Spain. J. Plant Prot. Res. 49**,** 276-279.

Arie, T., Christiansen, S.K., Yoder, O.C., Turgeon, B.G., 1997. Efficient cloning of Ascomycete mating type genes by PCR amplification of the conserved *MAT* HMG box. Fungal Genet. Biol. 21**,** 118-130.

Arnaise, S., Debuchy, R., Picard, M., 1997. What is a *bona fide* mating-type gene? Internuclear complementation of *mat* mutants in *Podospora anserina*. Mol. Gen. Genet. 256**,** 169-178.

Arnaise, S., Zickler, D., Le Bilcot, S., Poisier, C., Debuchy, R., 2001. Mutations in mating-type genes of the heterothallic fungus *Podospora anserina* lead to self-fertility. Genetics 159**,** 545-556.

Aylward, J., Steenkamp, E.T., Dreyer, L.L., Roets, F., Wingfield, M.J., Wingfield, B.D., 2016. Genetic basis for high population diversity in *Protea*-associated *Knoxdaviesia*. Fungal Genet. Biol. 96**,** 47-57.

Bihon, W., Wingfield, M.J., Slippers, B., Duong, T.A., Wingfield, B.D., 2014. *MAT* gene idiomorphs suggest a heterothallic sexual cycle in a predominantly asexual and important pine pathogen. Fungal Genet. Biol. 62**,** 55-61.

Billiard, S., Lopez-Villavicencio, M., Devier, B., Hood, M.E., Fairhead, C., Giraud, T., 2011. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biol. Rev. Camb. Philos. Soc. 86**,** 421-442.

Bliss, D.E., 1941. A new species of *Ceratostomella* on the date palm. Mycologia 33**,** 468-482.

Boetzer, M., Henkel, C.V., Jansen, H.J., 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 Biol. 13**,** R56.

Butler, G., 2007. The evolution of *MAT*: the Ascomycetes, in: J. Heitman, J.W. Kronstad, J.W. Taylor, L.A. Casselton (Eds.), Sex in Fungi: Molecular Determination and Evolutionary Implications. American Society of Microbiology, Washington, DC, pp. 3-18.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10**,** 421.

Dade, H.A., 1928. *Ceratostomella paradoxa*, the perfect stage of *Thielaviopsis paradoxa* (de Seynes) von Höhnel. Trans. Br. Mycol. Soc. 13**,** 184-193.

Damm, U., Mostert, L., Crous, P.W., Fourie, P.H., 2008. Novel *Phaeoacremonium* species associated with necrotic wood of *Prunus* trees. Persoonia 20**,** 87-102.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27**,** 1164-1165.

De Beer, Z.W., Duong, T.A., Barnes, I., Wingfield, B.D., Wingfield, M.J., 2014. Redefining *Ceratocystis* and allied genera. Stud. Mycol. 79**,** 187-219.

Debuchy, R., Berteaux-Lecellier, V., Silar, P., 2010. Mating systems and sexual morphogenesis in Ascomycetes, in: K.A. Borkovich, D.J. Ebbole (Eds.), Cellular and Molecular Biology of Filamentous Fungi. ASM Press, Washington, DC, pp. 501-535.

Debuchy, R., Turgeon, B.G., 2006. Mating-type structure, evolution, and function in Euascomycetes, in: U. Kües, R. Fischer (Eds.), The Mycota: Growth, Differentiation and Sexuality. Springer, Berlin Heidelberg, pp. 293-323.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., Mattick, J.S., 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 19**,** 4008.

Duong, T.A., De Beer, Z.W., Wingfield, B.D., Wingfield, M.J., 2013. Characterization of the mating-type genes in *Leptographium procerum* and *Leptographium profanum*. Fungal Biol. 117**,** 411-421.

Dyer, P.S., Inderbitzin, P., Debuchy, R., 2016. Mating-type structure, function, regulation and evolution in the Pezizomycotina, in: J. Wendland (Ed.), The Mycota I: Growth, Differentiation and Sexuality. Springer International Publishing, Switzerland, pp. 351-385.

El-Ani, A.S., Klotz, L.J., Wilbur, W.D., 1957. Heterothallism, heterokaryosis, and inheritance of brown perithecia in *Ceratostomella radicicola*. Mycologia 49**,** 181-187.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44**,** D279-D285.

Finn, R.D., Miller, B.L., Clements, J., Bateman, A., 2014. iPfam: a database of protein family and domain interactions found in the Protein Data Bank. Nucleic Acids Res. 42**,** D364-D373.

Gioti, A., Mushegian, A.A., Strandberg, R., Stajich, J.E., Johannesson, H., 2012. Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. Mol. Biol. Evol. 29**,** 3215-3226.

Glass, N.L., Grotelueschen, J., Metzenberg, R.L., 1990. *Neurospora crassa* A mating-type region. Proc. Natl. Acad. Sci. U.S.A. 87**,** 4912-4916.

Glass, N.L., Vollmer, S.J., Staben, C., Grotelueschen, J., Metzenberg, R.L., Yanofsky, C., 1988. DNAs of the two mating-type alleles of *Neurospora crassa* are highly dissimilar. Science 241**,** 570-573.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59**,** 307-321.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52**,** 696-704.

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

Heitman, J., Kronstad, J.W., Taylor, J.W., Casselton, L.A., 2007. Sex in fungi: molecular determination and evolutionary implications. ASM Press, Washington DC.

JCM-Riken BioResource Center, 2016a. NBRP: Genome sequencing of *Thielaviopsis paradoxa* JCM6020. http://www.ncbi.nlm.nih.gov/nuccore/BCHJ00000000.1.

JCM-Riken BioResource Center, 2016b. NBRP: Genome sequencing of *Thielaviopsis paradoxa* JCM6961. http://www.ncbi.nlm.nih.gov/nuccore/BCFY00000000.1.

Kanematsu, S., Adachi, Y., Ito, T., 2007. Mating-type loci of heterothallic *Diaporthe* spp.: homologous genes are present in opposite mating-types. Curr. Genet. 52**,** 11-22.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30**,** 772-780.

Klix, V., Nowrousian, M., Ringelberg, C., Loros, J.J., Dunlap, J.C., Pöggeler, S., 2010. Functional characterization of *MAT1-1*-specific mating-type genes in the homothallic ascomycete *Sordaria macrospora* provides new insights into essential and nonessential sexual regulators. Eukaryot. Cell 9**,** 894-905.

Korbie, D.J., Mattick, J.S., 2008. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat. Protoc. 3**,** 1452-1456.

Koressaar, T., Remm, M., 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23**,** 1289-1291.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33**,** 1870-1874.

Lee, S.C., Ni, M., Li, W.J., Shertz, C., Heitman, J., 2010. The evolution of sex: a perspective from the fungal kingdom. Microbiol. Mol. Biol. Rev. 74**,** 298-340.

Lin, X., Heitman, J., 2007. Mechanisms of homothallism in fungi and transitions between heterothallism and homothallism, in: J. Heitman, J.W. Kronstad, J.W. Taylor, L.A. Casselton (Eds.), Sex in Fungi: Molecular Determination and Evolutionary Implications. ASM Press, Washington, DC, pp. 35-57.

Martin, S.H., Wingfield, B.D., Wingfield, M.J., Steenkamp, E.T., 2011. Structure and evolution of the *Fusarium* mating type locus: new insights from the *Gibberella fujikuroi* complex. Fungal Genet. Biol. 48**,** 731-740.

Mbenoun, M., De Beer, Z.W., Wingfield, B.D., 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.

McGuire, I.C., Marra, R.E., Turgeon, B.G., Milgroom, M.G., 2001. Analysis of mating-type genes in the chestnut blight fungus, *Cryphonectria parasitica*. Fungal Genet. Biol. 34**,** 131-144.

Metzenberg, R.L., Glass, N.L., 1990. Mating type and mating strategies in *Neurospora*. Bioessays 12**,** 53-59.

Ni, M., Feretzaki, M., Sun, S., Wang, X., Heitman, J., 2011. Sex in fungi. Annu. Rev. Genet. 45**,** 405-430.

Réblová, M., Gams, W., Seifert, K.A., 2011. *Monilochaetes* and allied genera of the *Glomerellales*, and a reconsideration of families in the *Microascales*. Stud. Mycol. 68**,** 163-191.

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

Sambrook, J., Russell, D.W., 2001. Molecular Cloning - A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31**,** 3210-3212.

Solovyev, V., Kosarev, P., Seledsov, I., Vorobyev, D., 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 7 Suppl. 1**,** S10.1-S10.12.

Staben, C., Yanofsky, C., 1990. *Neurospora crassa* **a** mating-type region. Proc. Natl. Acad. Sci. U.S.A. 87**,** 4917-4921.

Stanke, M., Tzvetkova, A., Morgenstern, B., 2006. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. Genome Biol. 7 Suppl. 1**,** S11.1–S11.8.

Stöver, B.C., Müller, K.F., 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11**,** 7.

Suleman, P., Al-Musallam, A., Menezes, C.A., 2001. The effect of solute potential and water stress on black scorch caused by *Chalara paradoxa* and *Chalara radicicola* on date palms. Plant Dis. 85**,** 80-83.

Turgeon, B.G., 1998. Application of mating type gene technology to problems in fungal biology. Annu. Rev. Phytopathol. 36**,** 115-137.

Turgeon, B.G., Debuchy, R., 2007. *Cochliobolus* and *Podospora*: mechanisms of sex determination and the evolution of reproductive lifestyle, in: J. Heitman, J.W. Kronstad, J.W. Taylor, L.A. Casselton (Eds.), Sex in Fungi: Molecular Determination and Evolutionary Implications. ASM Press, Washington, DC, pp. 93-121.

Turgeon, B.G., Yoder, O.C., 2000. Proposed nomenclature for mating type genes of filamentous ascomycetes. Fungal Genet. Biol. 31**,** 1-5.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3-new capabilities and interfaces. Nucleic Acids Res. 40**,** e115.

Van der Nest, M.A., Beirn, L.A., Crouch, J.A., Demers, J.E., De Beer, Z.W., De Vos, L., Gordon, T.R., Moncalvo, J.-M., Naidoo, K., Sanchez-Ramirez, S., Roodt, D., Santana, Q.C., Slinski, S.L., Stata, M., Taerum, S.J., Wilken, P.M., Wilson, A.M., Wingfield, M.J., Wingfield, B.D., 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, M.A., Bihon, W., De Vos, L., Naidoo, K., Roodt, D., Rubagotti, E., Slippers, B., Steenkamp, E.T., Wilken, P.M., Wilson, A., Wingfield, M.J., Wingfield, B.D., 2014b. IMA Genome-F 2: Draft genome sequences of *Diplodia sapinea*, *Ceratocystis manginecans*, and *Ceratocystis moniliformis*. IMA Fungus 5**,** 135-140.

Webster, R.K., Butler, E.E., 1967. The origin of self-sterile, cross-fertile strains and culture sterility in *Ceratocystis fimbriata*. Mycologia 59**,** 212-221.

Whitehouse, H.L., 1949. Heterothallism and sex in the fungi. Biol. Rev. Camb. Philos. Soc. 24**,** 411-447.

Wilken, P.M., Steenkamp, E.T., De Beer, Z.W., Wingfield, M.J., Wingfield, B.D., 2013. IMA Genome-F1: Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. IMA Fungus 4**,** 357-358.

Wilken, P.M., Steenkamp, E.T., Hall, T.A., De Beer, Z.W., Wingfield, M.J., Wingfield, B.D., 2012. Both mating types in the heterothallic fungus *Ophiostoma quercus* contain *MAT1-1* and *MAT1-2* genes. Fungal Biol. 116, 427-437.

Wilken, P.M., Steenkamp, E.T., Wingfield, M.J., De Beer, Z.W., Wingfield, B.D., 2014. DNA loss at the *Ceratocystis fimbriata* mating locus results in self-sterility. PLoS ONE 9**,** e92180.

Wilken, P.M., Steenkamp, E.T., Wingfield, M.J., De Beer, Z.W., Wingfield, B.D., 2017. Which *MAT* gene? Pezizomycotina (Ascomycota) mating-type gene nomenclature reconsidered. Fungal Biol. Rev. 31**,** 199-211.

Wilson, A.M., Wilken, P.M., Van der Nest, M.A., Steenkamp, E.T., Wingfield, M.J., Wingfield, B.D., 2015b. Homothallism: an umbrella term for describing diverse sexual behaviours. IMA Fungus 6**,** 207-214.

Wingfield, B.D., Ades, P.K., Al-Naemi, F.A., Beirn, L.A., Bihon, W., Crouch, J.A., de Beer, Z.W., De Vos, L., Duong, T.A., Fields, C.J., Fourie, G., Kanzi, A.M., Malapi-Wight, M., Pethybridge, S.J., Radwan, O., Rendon, G., Slippers, B., Santana, Q.C., Steenkamp, E.T., Taylor, P.W., Vaghefi, N., van der Merwe, N.A., Veltri, D., Wingfield, M.J., 2015a. IMA Genome-F 4: Draft genome sequences of *Chrysoporthe austroafricana*, *Diplodia scrobiculata*, *Fusarium nygamai*, *Leptographium lundbergii*, *Limonomyces culmigenus*, *Stagonosporopsis tanaceti*, and *Thielaviopsis punctulata*. IMA fungus 6**,** 233-248.

Wingfield, B.D., Ambler, J.M., Coetzee, M., De Beer, Z.W., Duong, T.A., Joubert, F., Hammerbacher, A., McTaggart, A.R., Naidoo, K., Nguyen, H.D., 2016a. IMA Genome-F 6: 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, B.D., Barnes, I., De Beer, Z.W., De Vos, L., Duong, T.A., Kanzi, A.M., Naidoo, K., Nguyen, H.D., Santana, Q.C., Sayari, M., Seifert, K.A., Steenkamp, E.T., Trollip, C., Van der Merwe, N.A., Van der Nest, M.A., Wilken, P.M., Wingfield, M.J., 2015b. IMA Genome-F 5: Draft genome sequences of *Ceratocystis eucalypticola*, *Chrysoporthe cubensis*, *C. deuterocubensis*, *Davidsoniella virescens*, *Fusarium temperatum*, *Graphilbum fragrans*, *Penicillium nordicum*, and *Thielaviopsis musarum*. IMA Fungus 6**,** 493-506.

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

Witthuhn, R.C., Harrington, T.C., Wingfield, B.D., Steimel, J.P., Wingfield, M.J., 2000. Deletion of the *MAT-2* mating-type gene during uni-directional mating-type switching in *Ceratocystis*. Curr. Genet. 38**,** 48–52.

Yarden, O., 2016. Model fungi: Engines of scientific insight. Fungal Biol. Rev. 30**,** 33-35.

Yun, S.-H., Arie, T., Kaneko, I., Yoder, O.C., Turgeon, B.G., 2000. Molecular organization of mating type loci in heterothallic, homothallic, and asexual *Gibberella*/*Fusarium* species. Fungal Genet. Biol. 31**,** 7-20.

Yun, S.H., Kim, H.K., Lee, T., Turgeon, B.G., 2017. Self-fertility in *Chromocrea spinulosa* is a consequence of direct repeat-mediated loss of *MAT1-2*, subsequent imbalance of

nuclei differing in mating type, and recognition between unlike nuclei in a common cytoplasm. PLoS Genet. 13, e1006981.

Zheng, Q., Hou, R., Juanyu, Zhang, Ma, J., Wu, Z., Wang, G., Wang, C., Xu, J.R., 2013. The *MAT* locus genes play different roles in sexual reproduction and pathogenesis in *Fusarium graminearum*. PloS One 8**,** e66980.

Figure 1: A comparison of the mating-type regions from members of the Ceratocystidaceae. Shown here are the full *MAT1* locus of the homothallic *Ceratocystis fimbriata* (Wilken et al., 2014), as well as the *MAT1-1* idiomorph of *Thielaviopsis musarum* and *T. ethacetica*, and both the *MAT1-1 a*nd *MAT1-2* idiomorphs of *T. punctulata*, *T. paradoxa, T. euricoi* and *Huntiella omanensis* (Wilson et al., 2015a). The *MAT1-1* idiomorph of *T. paradoxa* and *MAT1-2* idiomorph of *T. euricoi* were inferred from PCR analysis. Gene names correspond to the colours shown in the key. The Neighbour Joining phylogeny was constructed from the combined 60S-LSU-MCM dataset of De Beer et al. (2014) using the MEGA7 software package (Kumar et al., 2016). The tree is rooted using the sequences of *Knoxdaviesia capensis* and *Graphium laricis*.

Figure 2: A plot of sequence identity between the *Thielaviopsis punctulata MAT1-1* (isolate 1032) and *MAT1-2* (isolate CRDP1) idiomorphs (Top), and between the *T. punctulata* idiomorphs and the *C. fimbriata MAT1* locus (Bottom). Red and blue lines indicate sequence identity based on BLASTn comparisons of the respective regions, with direct comparisons in red hues and reverse comparisons in blue hues. Details of the BLASTn results are given in Supplementary table 3.

Figure 3: Maximum likelihood phylogenies of the complete MAT1-1-1 (top), MAT1-1-2 (middle) and MAT1-2-1 (bottom) proteins from *Thielaviopsis* and representative Hypocreomycetidae. The clade including *Thielaviopsis* species and other members of the Microascales are indicated in the blue blocks. The trees were rooted using representative sequences from species of the Sordariomycetidae. The values at the nodes indicate bootstrap support values ≥ 60 %. The scale bars show the number of amino acid substitutions per site. The partial protein sequences of *T. paradoxa* MAT1-1-1 and *T. euricoi* MAT1-1-2 and MAT1-2-1 produced through PCR analysis were not included in the phylogenies.

Figure 4: A putative model for the evolution of the heterothallic *Thielaviopsis* mating-type locus from a heterothallic ancestor. The ancestral *MAT1* locus harboured the same gene complement as that of the extant *Thielaviopsis* species, although split differently across the two idiomorphs: *MAT1-1-1* and *MAT1-1-2* in the *MAT1-1* idiomorph and *MAT1-2-1* and *MAT1-2-7* in the *MAT1-2* idiomorph. An aberrant crossover event would have moved the *MAT1-1-2* gene from the *MAT1-1* to the *MAT1-2* idiomorph, producing an idiomorph organisation like that seen in modern *Thielaviopsis* species.

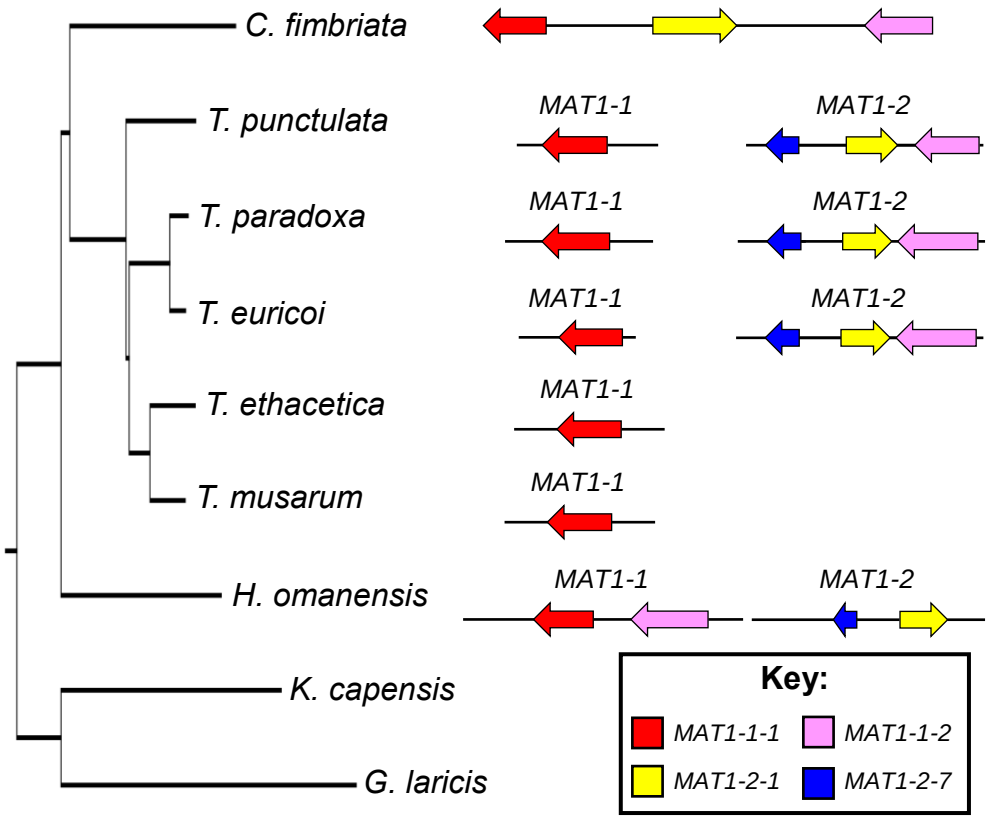
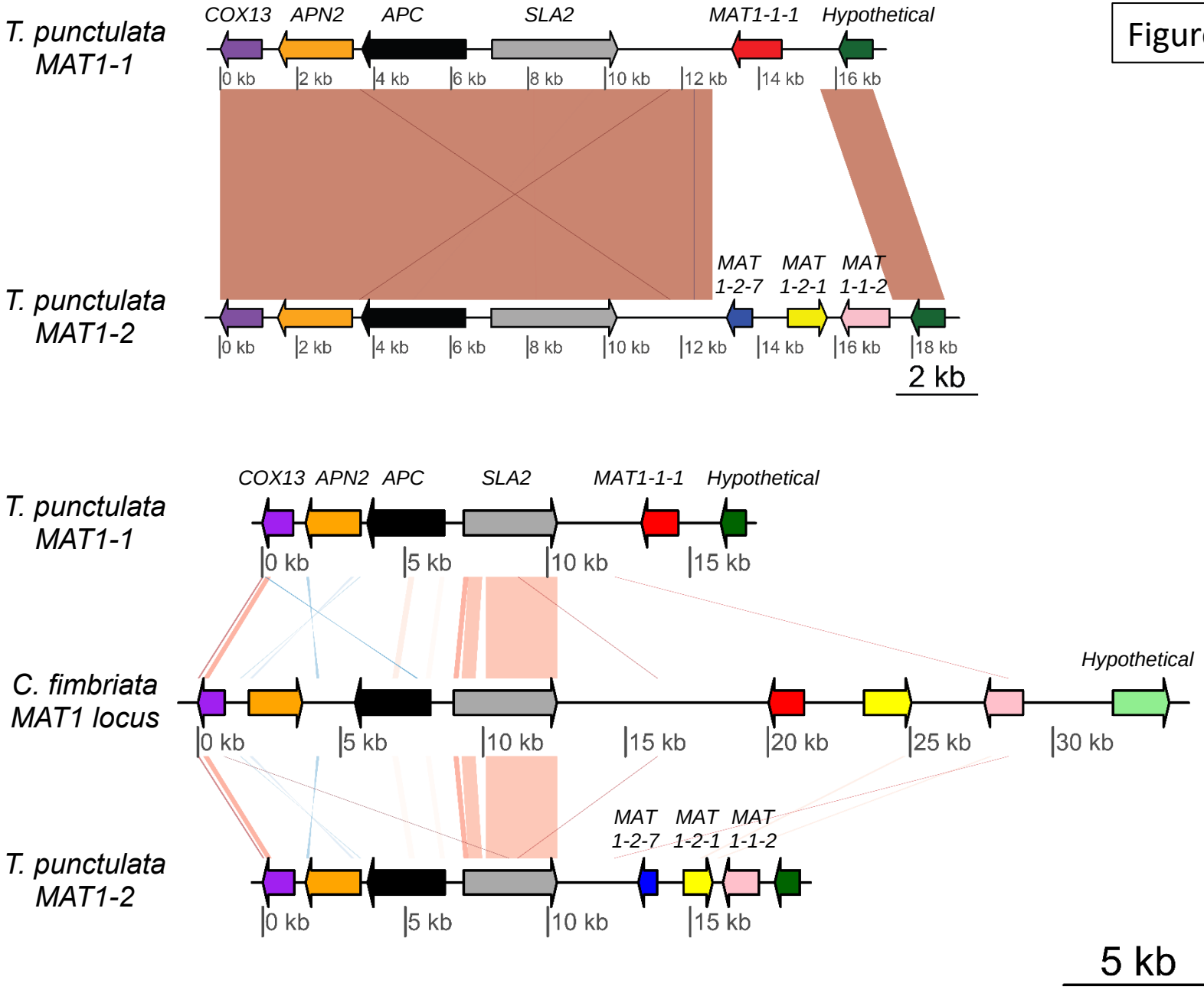
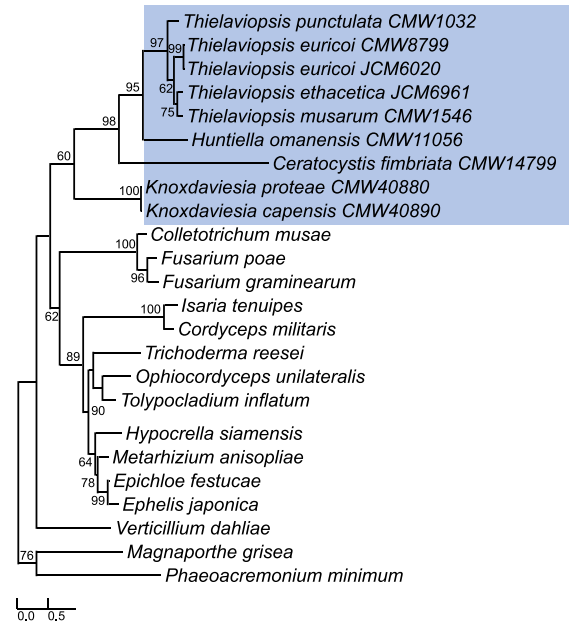
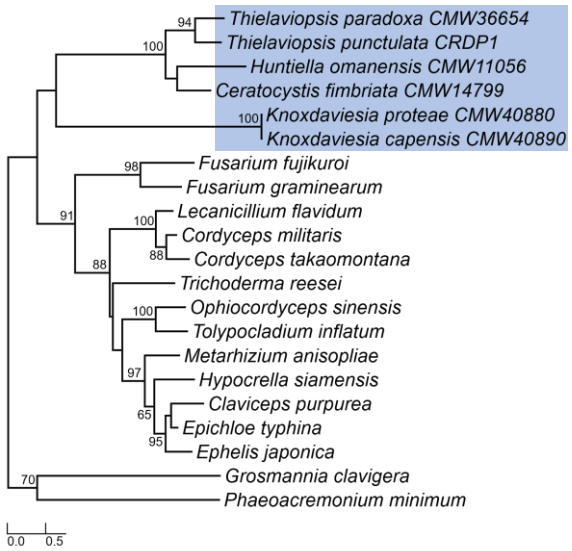

C. fimbriata
T. punctulata
T. paradoxa
T. euricoi
T. ethacetica
T. musarum
H. omanensis
K. capensis
G. laricis
MAT1-1
MAT1-2
MAT1-1
MAT1-2
MAT1-1
MAT1-2
MAT1-1
MAT1-1
MAT1-1
MAT1-2
Key:
MAT1-1-1
MAT1-1-2
MAT1-2-1
MAT1-2-7

Figure 1

Figure 2


T. punctulata MAT1-1
COX13
APN2
APC
SLA2
MAT1-1-1
Hypothetical
0 kb
2 kb
4 kb
6 kb
8 kb
10 kb
12 kb
14 kb
16 kb
T. punctulata MAT1-2
MAT 1-2-7
MAT 1-2-1
MAT 1-1-2
0 kb
2 kb
4 kb
6 kb
8 kb
10 kb
12 kb
14 kb
16 kb
18 kb
2 kb
T. punctulata MAT1-1
COX13
APN2
APC
SLA2
MAT1-1-1
Hypothetical
0 kb
5 kb
10 kb
15 kb
C. fimbriata MAT1 locus
Hypothetical
0 kb
5 kb
10 kb
15 kb
20 kb
25 kb
30 kb
T. punctulata MAT1-2
MAT 1-2-7
MAT 1-2-1
MAT 1-1-2
0 kb
5 kb
10 kb
15 kb
5 kb


Thielaviopsis punctulata CMW1032
Thielaviopsis euricoi CMW8799
Thielaviopsis euricoi JCM6020
Thielaviopsis ethacetica JCM6961
Thielaviopsis musarum CMW1546
Huntiella omanensis CMW11056
Ceratocystis fimbriata CMW14799
Knoxdaviesia proteae CMW40880
Knoxdaviesia capensis CMW40890
Colletotrichum musae
Fusarium poae
Fusarium graminearum
Isaria tenuipes
Cordyceps militaris
Trichoderma reesei
Ophiocordyceps unilateralis
Tolypocladium inflatum
Hypocrella siamensis
Metarhizium anisopliae
Epichloe festucae
Ephelis japonica
Verticillium dahliae
Magnaporthe grisea
Phaeoacremonium minimum
0.0 0.5


Thielaviopsis paradoxa CMW36654
Thielaviopsis punctulata CRDP1
Huntiella omanensis CMW11056
Ceratocystis fimbriata CMW14799
Knoxdaviesia proteae CMW40880
Knoxdaviesia capensis CMW40890
Fusarium fujikuroi
Fusarium graminearum
Lecanicillium flavidum
Cordyceps militaris
Cordyceps takaomontana
Trichoderma reesei
Ophiocordyceps sinensis
Tolypocladium inflatum
Metarhizium anisopliae
Hypocrella siamensis
Claviceps purpurea
Epichloe typhina
Ephelis japonica
Grosmannia clavigera
Phaeoacremonium minimum
0.0 0.5

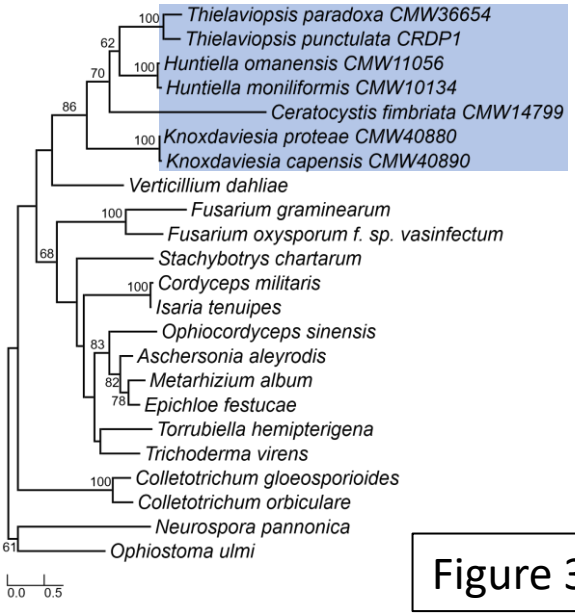

Thielaviopsis paradoxa CMW36654
Thielaviopsis punctulata CRDP1
Huntiella omanensis CMW11056
Huntiella moniliformis CMW10134
Ceratocystis fimbriata CMW14799
Knoxdaviesia proteae CMW40880
Knoxdaviesia capensis CMW40890
Verticillium dahliae
Fusarium graminearum
Fusarium oxysporum f. sp. vasinfectum
Stachybotrys chartarum
Cordyceps militaris
Isaria tenuipes
Ophiocordyceps sinensis
Aschersonia aleyrodis
Metarhizium album
Epichloe festucae
Torrubiella hemipterigena
Trichoderma virens
Colletotrichum gloeosporioides
Colletotrichum orbiculare
Neurospora pannonica
Ophiostoma ulmi
0.0 0.5

Figure 3

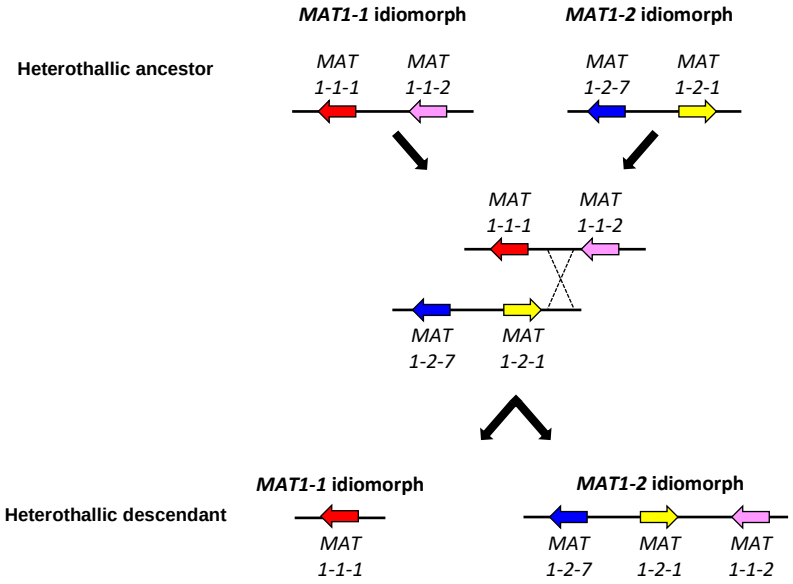

MAT1-1 idiomorph
MAT1-2 idiomorph
Heterothallic ancestor
MAT 1-1-1
MAT 1-1-2
MAT 1-2-7
MAT 1-2-1
MAT 1-1-1
MAT 1-1-2
MAT 1-2-7
MAT 1-2-1
MAT1-1 idiomorph
MAT1-2 idiomorph
Heterothallic descendant
MAT 1-1-1
MAT 1-2-7
MAT 1-2-1
MAT 1-1-2

Figure 4

Table 1: *Thielaviopsis* isolates used in the current study

Species	Culture number [a]	Mating-type	Mating genes present	*MAT* gene primers [b]
T. musarum	CMW1546 [c]	MAT1-1	*MAT1-1-1*	ThPara_111_F ThPara_111_R
T. punctulata	CMW1032 [d]	MAT1-1	*MAT1-1-1*	ThPun_111_F ThPunc_111_R
	CMW26389			
	CMW42889			
	CMW42890			
	CMW42891			
	CMW37952	MAT1-2	*MAT1-2-1*	ThPunc_121_F ThPunc_121_R
			MAT1-2-7	ThPun_127_R ThPun_127_F
	CRDP1 [c]			
			MAT1-1-2	ThPun112F ThPunct112R
T. paradoxa	CMW36655	MAT1-1	*MAT1-1-1*	ThPara_111_F ThPara_111_R
	CMW36654 [d]	MAT1-2	*MAT1-2-1*	ThPara_G_121F ThPara_G_121R
	CMW36689		*MAT1-2-7*	ThPara_G_127F ThPara_G_127R

	CMW36642		*MAT1-1-2*	ThPara_G_112F ThPara_G_112R
T. euricoi	CMW8790	MAT1-1	*MAT1-1-1*	ThEur_G_111R ThEur_G_111F
	CMW8799 [d]			
	CMW28537			
	JCM6020 [c]			
	CMW28538	MAT1-2	*MAT1-2-1*	ThPara_G_121F ThPara_G_121R
			MAT1-2-7	ThPara_G_127F ThPara_G_127R
			MAT1-1-2	ThPara_G_112F ThPara_G_112R
T. ethacetica	CMW36725	MAT1-1	*MAT1-1-1*	ThPara_111_F ThPara_111_R
	CMW36771			
	CMW36775			
	JCM6961 [c]			

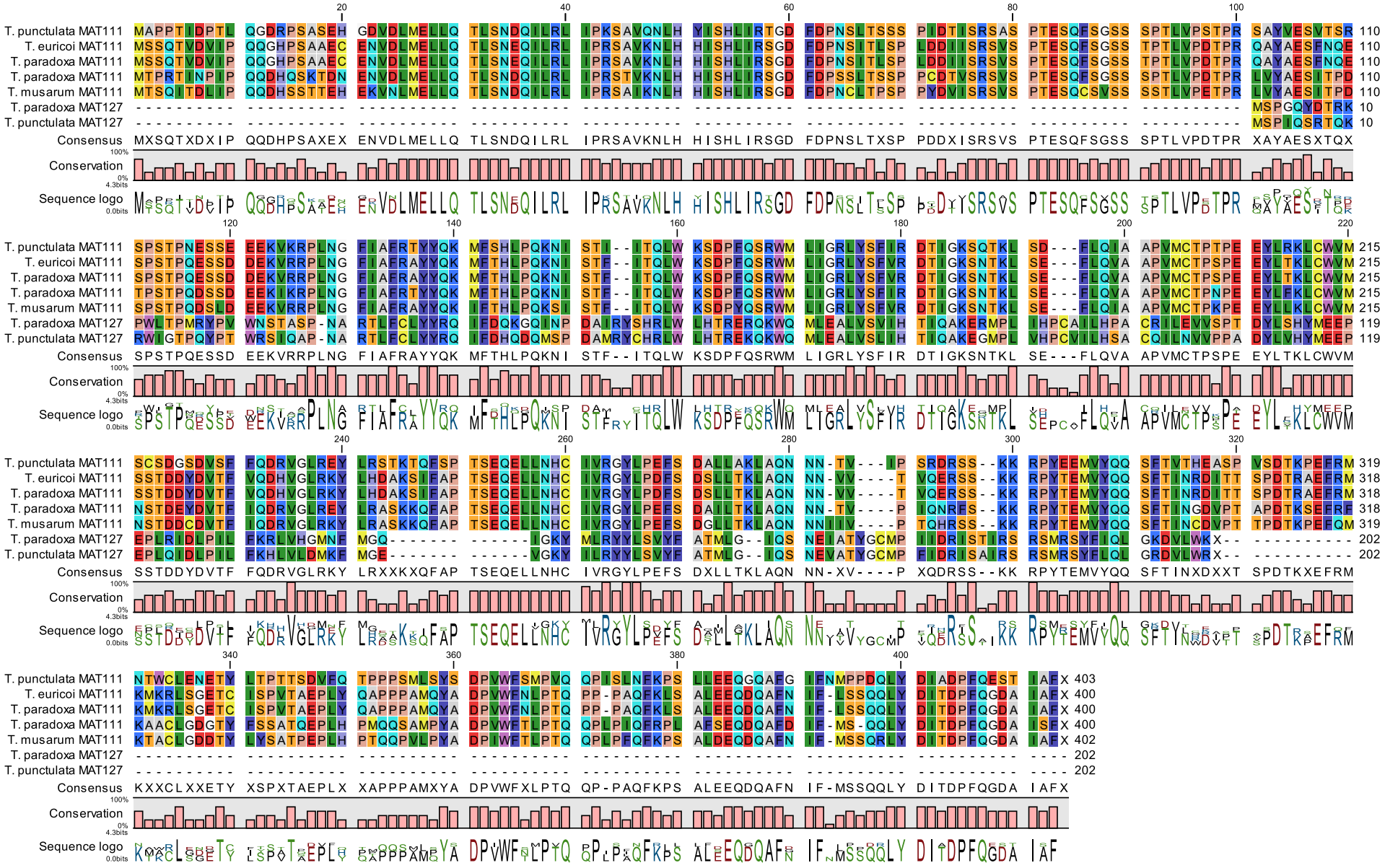
[a] All isolates from the study of Mbenoun et al. (2014), apart from CMW37952 which comes from the study of Suleman et al. (2001). Isolates CRDP1, JCM6960 and JCM6961 were represented solely by genome sequences, and as such were not used in the PCR analysis.
[b] See table 2 for primer sequences.
[c] Isolates for which the genome sequence was obtained from NCBI.
[d] Isolates for which the genome was sequenced in this study.

Table 2: Primers designed to amplify fragments of the mating-type genes.

Primer Name	Primer sequence (5' to 3')	Primer position in gene [1]	Gene sequence used to design primer	*MAT* gene targeted
ThPun112F	CAGTTCGTACATGACTGCCG	474…493	*T. punctulata MAT1-1-2*	*MAT1-1-2*
ThPun112R	GCCTTGACTTCTGTTGCGAA	885…866		
ThPara_G_112F	GGAACTTCCCCTGACAGTCA	45…64	*T. paradoxa MAT1-1-2*	
ThPara_G_112R	TTCAGATCCGTTGCAGGGTA	1427…1408		
ThPun_121_F	TTCAGGTGTCGCAGTTCTCT	137…156	*T. punctulata MAT1-2-1*	*MAT1-2-1*
ThPun_121_R	CACTTCAGACGATGGCATGG	937…918		
ThPara_G_121F	CACTGGGTGCAGAGCTTCT	29…47	*T. paradoxa MAT1-2-1*	
ThPara_G_121R	GCCCGATATAGAGACAGGCT	928…909		
ThPun_127_F	CACGAGATAGTCTGCGGGAG	10…29	*T. punctulata MAT1-2-7*	*MAT1-2-7*
ThPun_127_R	ATTCAGTCTCGCACGCAAAA	395…376		
ThPara_G_127F	CCTGGTCAATATGATACGCGG	7…27	*T. paradoxa MAT1-2-7*	
ThPara_G_127R	TTGGATTCCGAGCATTGTAGC	556…536		
ThPun_111_F	TTGATCCCGAAGTCTGCTGT	118…137	*T. punctulata MAT1-1-1*	*MAT1-1-1*
ThPun_111_R	TGAAACAGGAGAGGCTTCGT	1016…997		
ThPara_111_F	CCACATCAGCCATTTGATTC	150…169	*T. ethacetica MAT1-1-1*	
ThPara_111_R	TCTCCCTGAAAAGGGTCCGT	1250…1231		
ThEur_G_111F	CTGCACCACATCAGCCATTT	145…164	*T. euricoi MAT1-1-1* [2]	
ThEur_G_111R	CGCGGTTTATGGTGAACGAT	979…960		

[1] Nucleotide positions spanned by the primer. Positions assigned with the A of the start codon as gene position 1.

[2] Based on the genome of isolate JCM 6020.



Supplementary figure 1: A comparison of the *Thielaviopsis* MAT1-1-1 and MAT1-2-7 proteins. Proteins were aligned with the NCBI Constraint-based Multiple Alignment Tool COBALT (https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?LINK_LOC=BlastHomeLink) using default settings. The consensus sequence is based on majority amino acid presence, while the graph and sequence logo shows the level of conservation and amino acid distribution among the alignment, respectively.

Supplementary table 1: Assembly and scaffolding steps for the T. punctulata CMW1032 genome

Assembly/ Scaffold	Program	Input data	Parameters	Resulting number of contigs
1	De novo assembly on CLC Genomics	• Trimmed paired reads	Word size:64 Bubble size: 100 Min. contig lenght:200	2 370
2	SSPACE and Gapfiller	• Trimmed paired reads • Assembly 1	• Min. overlap (-m): 50 • Min read number for base calling (-o): 10 • Bases to trim from contig end (-t): 2 • Min. number of read pair links (-k): 10 • Max ratio between the best two contig pairs (-a): 0.70 • Extend contig option (-x): Enabled	1 723
3	Join contigs option of Genome Finishing Module – Reference sequence	• Assembly 2 • Assembly 1 as reference	Default settings	1 705
4	Join contigs option of Genome Finishing Module – Map contigs to themselves	• Assembly 3 • Use Assembly 3 as reference	Default settings	1 597
5	Join contigs option of Genome Finishing Module – Pair-end overlap	• Assembly 4 • Trimmed pair-end data mapped	Default settings, with min number of overlaps = 10	1 306
6, 7	SSPACE and Gapfiller	• Assembly 5 • Trimmed pair-end reads	As for run 2	1 290
8	Join contigs option of Genome Finishing Module – Reference sequence	• Assembly 8 • Assembly 1 as reference	Default settings	1 262
9	Join contigs option of Genome Finishing Module – Map contigs to themselves	• Assembly 8 • Use Assembly 3 as reference	Default settings	1 254
10	Join contigs option of Genome Finishing Module – Pair-end overlap	• Assembly 9 • Trimmed pair-end data mapped	Default settings, with min number of overlaps = 10	1 216
11	Join contigs option of Genome Finishing Module – Map long reads	• Assembly 10 • Trimmed Pair-end reads as long reads	Default settings	1 215
12, 13	SSPACE and Gapfiller	• Assembly 11 • Trimmed pair-end reads	As for run 2	1 169
14	Join contigs option of Genome Finishing Module – Reference sequence	• Assembly 13 • Assembly 1 as reference	Default settings	1 151

15	Join contigs option of Genome Finishing Module – Map contigs to themselves	• Assembly 14 • Use Assembly 3 as reference	Default settings	1 140
16	Join contigs option of Genome Finishing Module – Pair-end overlap	• Assembly 15 • Trimmed pair-end data mapped	Default settings, with min number of overlaps = 10	1 104
17, 18	SSPACE and Gapfiller	• Assembly 16 • Trimmed pair-end reads	As for run 2	1 093
19	Join contigs option of Genome Finishing Module – Map contigs to themselves	• Assembly 18 • Use Assembly 3 as reference	Default settings	1 081
20	Join contigs option of Genome Finishing Module – Pair-end overlap	• Assembly 19 • Trimmed pair-end data mapped	Default settings, with min number of overlaps = 10	1 079
21, 22	SSPACE and Gapfiller	• Assembly 20 • Trimmed pair-end reads	As for run 2	1 073
23	Join contigs option of Genome Finishing Module – Reference sequence	• Assembly 22 • Assembly 1 as reference	Default settings	1 065
24	Join contigs option of Genome Finishing Module – Map contigs to themselves	• Assembly 23 • Use Assembly 3 as reference	Default settings	1 064
25	Join contigs option of Genome Finishing Module – Pair-end overlap	• Assembly 15 • Trimmed pair-end data mapped	Default settings, with min number of overlaps = 10	1 059
26, 27	SSPACE and Gapfiller	• Assembly 25 • Trimmed pair-end reads	As for run 2	1 055
28	Join contigs option of Genome Finishing Module – Reference sequence	• Assembly 27 • Assembly 1 as reference	Default settings	1 044
29	Join contigs option of Genome Finishing Module – Map contigs to themselves	• Assembly 28 • Use Assembly 3 as reference	Default settings	1 039
30	Join contigs option of Genome Finishing Module – Pair-end overlap	• Assembly 29 • Trimmed pair-end data mapped	Default settings, with min number of overlaps = 10	1 033
31, 32	SSPACE and Gapfiller	• Assembly 30 • Trimmed pair-end reads	As for run 2	1 033
33	Join contigs option of Genome Finishing Module – Reference sequence	• Assembly 32 • Assembly 1 as reference	Default settings	1 029
34	Join contigs option of Genome Finishing Module – Pair-end overlap	• Assembly 33 • Trimmed pair-end data mapped	Default settings, with min number of overlaps = 10	1 028

35, 36	SSPACE and Gapfiller	• Assembly 34 • Trimmed pair-end reads	As for run 2	1 025
37	Join contigs option of Genome Finishing Module – Reference sequence	• Assembly 32 • Assembly 1 as reference	Default settings	1 022
38, 39	SSPACE and Gapfiller	• Assembly 37 • Trimmed pair-end reads	As for run 2	1 019
40	Join contigs option of Genome Finishing Module – Pair-end overlap	• Assembly 39 • Trimmed pair-end data mapped	Default settings, with min number of overlaps = 10	1 017

Supplementary table 2: Accession numbers of the proteins used for phylogenetic analysis.

Gene Name	Accession	Latin name
121	AAX83065	*Ophiostoma ulmi*
121	CCD57792	*Neurospora pannonica*
121	This study	*Thielaviopsis paradoxa*
121	This study	*Thielaviopsis punctulata*
121	KFA75378	*Stachybotrys chartarum IBT 40288*
121	KZZ99689	*Aschersonia aleyrodis RCEF 2490*
121	AAG42810	*Fusarium graminearum*
121	AEO15074	*Fusarium oxysporum f. sp. vasinfectum*
121	AOY41710	*Huntiella omanensis*
121	AOY41709	*Huntiella moniliformis*
121	APB91644	*Knoxdaviesia proteae*
121	APB91653	*Knoxdaviesia capensis*
121	AHV84686	*Ceratocystis fimbriata*
121	AEI72619	*Epichloe festucae*
121	KHO00962	*Metarhizium album ARSEF 1941*
121	CEJ91387	*Torrubiella hemipterigena*
121	XP_013957388	*Trichoderma virens Gv29-8*
121	BAC66500	*Cordyceps militaris*
121	AFX66472	*Ophiocordyceps sinensis*
121	EQB45543	*Colletotrichum gloeosporioides Cg-14*
121	BAC66503	*Isaria tenuipes*
121	ENH77678	*Colletotrichum orbiculare MAFF 240422*
121	BAG12301	*Verticillium dahliae*
111	XP_007913910	*Phaeoacremonium minimum UCRPA7*

111	AHV84685	*Ceratocystis fimbriata*
111	AOY41712	*Huntiella omanensis*
111	This study	*Thielaviopsis euricoi*
111	This study	*Thielaviopsis paradoxa*
111	This study	*Thielaviopsis paradoxa*
111	This study	*Thielaviopsis musarum*
111	APB91650	*Knoxdaviesia proteae*
111	APB91636	*Knoxdaviesia capensis*
111	CAD59610	*Fusarium poae*
111	BAC67541	*Isaria tenuipes*
111	ABE98363	*Fusarium graminearum*
111	KOM19990	*Ophiocordyceps unilateralis*
111	ACR78244	*Trichoderma reesei*
111	AGW27562	*Tolypocladium inflatum*
111	AKM95173	*Cordyceps militaris*
111	KFG83432	*Metarhizium anisopliae*
111	ALI93545	*Hypocrella siamensis*
111	BAD72608	*Ephelis japonica*
111	CAD59611	*Colletotrichum musae*
111	BAG83052	*Verticillium dahliae*
111	ACN59937	*Epichloe festucae*
111	BAC65083	*Magnaporthe grisea*
112	BAD72599	*Cordyceps militaris*
112	BAD95879	*Cordyceps takaomontana*
112	BAD72607	*Ephelis japonica*
112	BAD72611	*Epichloe typhina*

112	BAD72603	*Claviceps purpurea*
112	AAC71054	*Fusarium fujikuroi*
112	AMP43944	*Fusarium graminearum*
112	ALI93546	*Hypocrella siamensis*
112	BAE93604	*Lecanicillium flavidum*
112	BAE93597	*Metarhizium anisopliae*
112	AGW27559	*Ophiocordyceps sinensis*
112	BAE93601	*Tolypocladium inflatum*
112	ACR78245	*Trichoderma reesei*
112	AHV84687	*Ceratocystis fimbriata*
112	APB91637	*Knoxdaviesia capensis*
112	APB91651	*Knoxdaviesia proteae*
112	AOY41713	*Huntiella omanensis*
112	AGH03157	*Grosmannia clavigera*
112	XP_007913911	*Phaeoacremonium minimum UCRPA7*
112	This study	*Thielaviopsis punctulata*
112	This study	*Thielaviopsis paradoxa*

Supplementary table 3: BLASTn results used to construct the plots in figure 2.

Query Sequence	Subject Sequence	Alignment start to stop (for query)	Alignment start to stop (for subject)	Alignment length	% identical matches	Number of mismatches	Number of gap openings	E-value	Bit score
T. punctulata MAT1-1	*T. punctulata MAT1-2*	1 – 12788	1 – 12799	12870	97.82	128	22	0.0	21982
		15601 – 16967	17490 – 18846	1367	98.46	11	2	0.0	2370
		11686 – 11701	3622 – 3637	16	100	0	0	0.27	30.1
		3627 – 3642	11698 – 11713	16	100	0	0	0.27	30.1
		12311 – 12326	12338 – 12323	16	100	0	0	0.27	30.1
		9690 – 9714	5035 – 5011	25	84	4	0	0.94	28.3
		8131 - 8155	8204 – 8228	25	84	4	0	0.94	28.3
T. punctulata MAT1-1	*C. fimbriata*	7841 – 10336	10102 – 12597	2497	77.93	549	2	0.0	2012
		7231 – 7713	9256 – 9738	484	79.13	99	2	1,00E-115	410
		123 – 303	199 – 379	183	83.06	27	2	1,00E-47	185

		7062 – 7227	8975 – 9140	166	84.34	26	0	4,00E-47	183
		1547 – 1639	4255 – 4163	93	83.87	15	0	3,00E-22	100
		5138 – 5332	6828 – 7022	195	68.72	61	0	4,00E-15	77.0
		1 – 49	1 – 49	49	93.88	3	0	1,00E-14	75.2
		3118 – 3212	1921 – 1827	95	73.68	25	0	1,00E-09	59.0
		80 – 121	7706 – 7664	43	83.72	6	1	8,00E-05	42.8
		83 – 122	7705 – 7666	40	82.50	7	0	3,00E-04	41.0
		83 – 122	7708 – 7669	40	82.50	7	0	3,00E-04	41.0
		83 – 122	7711 – 7672	40	82.50	7	0	3,00E-04	41.0
		83 – 122	7714 – 7675	40	82.50	7	0	3,00E-04	41.0
		84 – 122	7716 – 7678	39	82.05	7	0	0.001	39.2
		6229 – 6390	8002 – 8163	163	66.26	53	2	0.001	39.2
		3418 – 3453	1519 – 1484	36	80.56	7	0	0.040	33.7
		8957 – 8975	16114 – 16132	19	94.74	1	0	0.49	30.1
		12335 – 12355	28415 – 28435	21	90.48	2	0	0.49	30.1
C. fimbriata	*T. punctulata MAT1-2*	10102 – 12597	7827 – 10322	2497	77.97	548	2	0.0	2017

		9256 – 9738	7217 – 7699	484	78.72	101	2	8,00E-113	401
		199 – 379	125 – 305	183	83.06	27	2	1,00E-47	185
		8975 – 9140	7048 – 7213	166	84.34	26	0	4,00E-47	183
		4163 – 4255	1624 – 1532	93	83.87	15	0	4,00E-22	100
		1 – 49	1 – 49	49	93.88	3	0	1,00E-14	75.2
		6828 – 7022	5133 – 5327	195	68.21	62	0	5,00E-14	73.4
		1827 – 1921	3197 – 3103	95	73.68	25	0	1,00E-09	59.0
		8002 – 8163	6224 – 6385	163	66.87	52	2	2,00E-05	44.6
		24717 – 24812	15348 – 15443	96	68.75	30	0	0.001	39.2
		1484 – 1519	3440 – 3405	36	80.56	7	0	0.044	33.7
		27762 – 27842	15959 – 16039	81	69.14	25	0	0.044	33.7
		915 – 930	8617 – 8632	16	100.00	0	0	0.54	30.1
		16114 – 16132	8943 – 8961	19	94.74	1	0	0.54	30.1
		28415 – 28435	12346 – 12366	21	90.48	2	0	0.54	30.1