Diversity and evolution of polyketide biosynthesis gene clusters in the *Ceratocystidaceae*

Mohammad Sayari, Emma T. Steenkamp, Magriet A. van der Nest and Brenda D. Wingfield*

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

Corresponding author

*Brenda D Wingfield

Email: Brenda.wingfield@fabi.up.ac.za

Tel: +27 124206288

Fax: +27 124203960

Abstract

Polyketides are low-molecular weight secondary metabolites with diverse biological activities. Polyketide synthase (PKS) and other proteins responsible for producing these compounds are often encoded from genes that are clustered in the same genomic region. Functional analyses and recent genomic studies are increasingly showing that most fungi are capable of producing a repertoire of PKS compounds. In this study, we considered the potential of *Ceratocystidaceae* for producing polyketides by making use of a comparative genomics approach. The fungi in this family have diverse lifestyles ranging from being saprotrophs, insect associates through to plant pathogens. Our specific aims were to identify the putative polyketide biosynthesis gene clusters, to characterize them at the structural and phylogenetic levels and to predict the types of polyketide compounds they might produce. For this purpose, we used the whole genome sequences from nineteen species in the genera, *Ceratocystis, Endoconidiophora, Davidsoniella, Huntiella, Thielaviopsis* and *Bretziella*, to identify and characterise PKS gene clusters, by employing a range of bioinformatics and phylogenetic tools. Our results showed that all of the examined genomes contained putative clusters containing a non-reducing type I PKS and a type III PKS. Phylogenetic analyses suggested that these genes were already present in the ancestor of the *Ceratocystidaceae* after which they diverged together with the species harbouring them. By contrast, the various reducing type I PKS-containing clusters identified in these genomes, appeared to have distinct origins during the evolution of this family. Although one of the identified clusters potentially allows for the production of melanin, their functional characterization will undoubtedly reveal many novel and important compounds implicated in the biology of the *Ceratocystidaceae*.

Keywords: Polyketide synthase, gene, cluster

1.1. Introduction

Polyketides belong to a group of low-molecular weight compounds typically referred to as secondary metabolites (Keller et al 2005). Polyketides are structurally diverse and show a range of biological and pharmacological activities (Keller et al 2005). They are produced in a process reminiscent of fatty acid biosynthesis from short chain carboxylic molecules (i.e., acyl coenzyme A [CoA]) (Rawlings 1997; Keller et al 2005). The process is catalysed by polyketide synthases (PKSs), which may represent one of three types (PKS-I, PKS-II or PKS-III) depending on their primary structures, sequences and catalytic mechanisms (Chan et al 2009; Cox and Simpson 2010).

As is true for many other classes of secondary metabolites, the genes required for producing polyketides typically occur in the same genomic region where they are usually expressed together (Keller et al 2005; Shwab and Keller 2008; Osbourn 2010). In fact, the need for coordinated gene expression has been suggested as one of the evolutionary drivers for clustering the genetic determinants of secondary metabolite biosynthesis (Khaldi and Wolfe 2011). Another hypothesis is that clustering confers a selective advantage to the gene cluster itself (Walton 2000; Keller et al 2005). Nevertheless, a typical polyketide biosynthesis cluster would thus include the core PKS synthase gene for producing the parent compound and various additional genes whose products may be involved in processes such as expression regulation of the cluster and additional post-translational modifications of the parent polyketide (also referred to as tailoring enzymes), as well as its cellular transport (Brown et al 2012; Osbourn 2010).

In fungi, polyketides are the most abundant secondary metabolites (Keller et al 2005). For instance, a recent comparative genomic study of *Fusarium graminearum*, *Fusarium oxysporum*, and *Fusarium verticillioides* identified 46 likely secondary metabolite biosynthesis gene clusters, of which 87 % comprised PKS genes (Ma et al 2013). Of the three known types of PKS, only the PKS-I and PKS-III types have so far been reported from fungi, with PKS-II only known so far from bacteria.

Fungal PKS-I closely resembles the fatty acyl synthases of *Saccharomyces cerevisiae* and animals and is a large multi-domain enzyme (Cane et al 1998). Most are iterative and thus perform repeated cycles of polyketide chain elongation and associated

functional group modification to produce the final polyketide molecule (e.g., a non-iterative enzyme would carry out only one cycle of biosynthetic and modification reactions to make a diketide). A PKS-I always comprises beta-ketoacyl synthase (KS), acyl carrier protein (ACP) and acyl transferase (AT) domains. Some may also contain domains for starter acyltransferase (SAT), thioesterase (TE) and C-methyltransferase (MET), as well as the product template domain (PT). Those that reduce the ketone groups in their ketide subunits may also contain beta-ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains (Kroken et al 2003). Therefore, apart from the KS, AT and ACP domains, a non-reducing (NR) PKS-I (i.e., NR-PKS-1) may include the SAT, TE, PT and MET domains. Its reducing (R) counterpart (i.e., R-PKS-1) additionally contains one or more of the KR, DH and ER domains, depending on the degree to which keto groups need reduction to hydroxyl groups (Cox 2007; Kroken et al 2003).

Compared to PKS-I, PKS-III is much less abundant in fungi (Hashimoto et al 2014). However, interest in this protein type is growing due to the numerous compounds it has been reported to produce (Austin and Noel 2003) including chalcones, pyrones, phloroglucinols, acridones, stilbenes, and resorcinolic lipids (Yu et al 2012; Hashimoto et al 2014). PKS-III is also known as the chalcone synthase-like PKS and is architecturally relatively simple, consisting of homodimers of identical monomeric KS domains (Austin and Noel 2003). In an ACP-independent manner, the single active site in each monomer catalyses full cycles of polyketide chain elongation, together with the associated functional group modifications, to iteratively form polyketide products (Austin and Noel 2003).

While a wide array of polyketides have been identified and characterized in filamentous fungi (Beck et al 1990; Chang et al 1995; Fulton et al 1999; Graziani et al 2004), little information is available for Ascomycota classified in the *Ceratocystidaceae* (Order: Microascales) (De Beer et al 2014). Based on the number of anticipated secondary metabolite clusters per fungal genome, earlier studies propose that most fungi are capable of synthesizing up to 10-fold more secondary metabolites than formerly characterized and identified by chemical analysis (Ma et al 2013). Due to their abundance in fungi, the same pattern would potentially also hold true for polyketide biosynthesis in the *Ceratocystidaceae*. Therefore, our overall objectives were to identify in these fungi putative polyketide biosynthesis gene clusters (hereafter referred to as PKS clusters), to characterize them at the structural and phylogenetic levels and to predict the types of polyketide compounds they might produce. For this purpose, we

utilized as wide a taxon sampling as possible by including all of the published genome sequences for members of this family. In total, we used 20 genomes representing 19 species and six genera (i.e., *Ceratocystis*, *Huntiella*, *Thielaviopsis*, *Endoconidiophora*, *Bretziella* and *Davidsoniella*) in this family (De Beer et al 2014; De Beer et al 2017). We also included the genome for *C. adiposa* that is known to represent a discrete genus in the family, but whose taxonomy has not yet been revised (De Beer et al 2014).

1.2. Materials and methods

1.2.1. Identification and annotation of PKS clusters

For each of the 20 genomes included in this study (Table 1), we evaluated the completeness of the assemblies using version 2.0 of the programme Benchmarking Universal Single-Copy Orthologs (BUSCO) with the Fungal dataset (Simão et al 2015). To identify contigs containing secondary metabolite biosynthesis clusters we used version 4.0 of the Antibiotics and Secondary Metabolite Analysis Shell (antiSMASH) software (Blin et al 2017) at the Secondary Metabolite Bioinformatics Portal (http://www.secondarymetabolites.org/; Weber and Kim 2016). antiSMASH utilizes the Cluster Assignment by Islands of Sites (CASSIS) algorithm to predict and confirm the borders of secondary metabolite biosynthesis clusters, which assumes the existence of common regulatory patterns in cluster promoters for cluster delimitation (Wolf et al 2016). These annotations were confirmed using programme Secondary Metabolite Unknown Regions Finder (SMURF; www.jcvi.org/smurf/) (Khaldi et al 2010).

The open reading frames (ORFs) encoded on these contigs (containing putative PKS clusters) were then predicted with Web AUGUSTUS (http://bioinf.uni-greifswald.de/augustus/; Stanke and Morgenstern 2005). Assigned ORFs were examined for signature domains of the target secondary metabolite biosynthesis genes by InterProScan (Zdobnov and Apweiler 2001). PKS genes and domains were also annotated with ClustScan (Starcevic et al 2008), which additionally predicted the chemical structure of the polyketide product's linear backbone. These domain annotations were verified with the Secondary Metabolites by InterProScan (SMIPS) tool (Wolf et al 2016), as well as comparison to those produced manually through using the PKS/NRPS analysis website (http://nrps.igs.umaryland.edu/nrps; Bachmann and Ravel 2009) and the MOTIF search tool (http://www.genome.jp/tools/motif/).

Table 1. Isolates numbers and genome sequence and completeness information for the species used in this study.

Species [a]	Isolate number [b]	GenBank Accession number	BUSCO completeness scores [c]				References
			Complete and Single-copy	Complete and Duplicated	Fragmented	Missing	
B. fagacearum	CMW2656	MKGJ00000000	95,2	0	3,1	1,7	Wingfield et al 2016b
C. adiposa	CMW2573	LXGU00000000	99,3	0	0,3	0,4	Wingfield et al 2016a
C. albifundus	CMW13980	JSSU000000000	97,6	0	1,4	1	Van der Nest et al 2014a
C. eucalypticola	CMW11536	LJOA00000000	97,9	0	1,4	0,7	Wingfield et al 2015b
C. fimbriata	CMW15049	APWK00000000	91	0,3	1,4	7,6	Wilken et al 2013
C. harringtonii	CMW14789	MKGM0000000	97,9	0	0,7	1,4	Wingfield et al 2016b
C. manginecans	CMW17570	JJRZ000000000	97,2	0,3	1,7	1,1	Van der Nest et al 2014b
C. platani	CF0	LBBL00000000	99,3	0	0,4	0	Belbahri 2015
C. smalleyii	CMW14800	NETT01000000	99	0	0,7	0,3	
D. virescens	CMW17339	LJZU000000000	97,6	0	1,4	1	Wingfield et al 2015b
E. laricicola	CMW20928	LXGT00000000	98,9	0,3	0,7	0,4	Wingfield et al 2016a
E. polonica	CMW20930	LXKZ00000000	97,6	0	1,4	1	Wingfield et al 2016a
H. bhutanensis	CMW8217	MJMS00000000	96,2	0,3	2,1	1,7	Wingfield et al 2016b
H. decipiens	CMW30855	NETU00000000	96,9	0,3	2,1	1	Wingfield et al 2017
H. moniliformis	CMW10134	JMSH00000000	94,8	0,3	3,4	1,8	Van der Nest et al 2014b
H. omanensis	CMW11056	JSUI000000000	88,6	0,3	7,9	3,5	Van der Nest et al 2014a
H. savannae	CMW17300	LCZG00000000	96,2	1,7	2,1	1,7	Van der Nest et al 2015

T. musarum	CMW1546	LKBB00000000	98,6	0	1,4	0	Wingfield et al 2015b
T. punctulata 1	BPI 893173	LAEV00000000	99,3	0,7	0,7	0	Wingfield et al 2015a
T. punctulata 2	CMW1032	MJMR01000000	99,7	0,7	0,3	0	Wilken et al 2018

[a] The two isolates of *T. punctulata* included here are indicated with the digits "1" and '2'.

[b] Isolates with CMW numbers may be obtained from the culture collection of the Tree Protection Cooperative Programme (TPCP) Forestry and Agricultural Biotechnology Institute (FABI) University of Pretoria Pretoria South Africa. Those with CFO and BPI numbers may be obtained from respectively Istituto per la Protezione delle Piante (Consiglio Nazionale delel Ricerche Florence Italy) and the US National Fungus Collections Systematic Botany and Mycology Laboratory Maryland U.S.A

[c] The completeness of each genome assembly and annotation was evaluated with the Benchmarking Universal Single-Copy Orthologs (BUSCO) v. 2 tool, by employing the set of single copy genes conserved among all fungi (Simão et al 2015).

All predicted PKS clusters and PKS domain organizations were evaluated manually (Ichikawa et al 2012). For this purpose, similarity searches were done using tblastn and blastp against the non-redundant (nr) protein sequence database of the National Centre for Biotechnology Information (NCBI; http://blast.ncbi.nlm.nih.gov/). We considered our genes homologous to these known ones when they had more than 60% similarity in a blast search, and also when their domain architectures had high similarity.

We employed a PCR-based approach for confirming the organization of the genes in each of the identified PKS clusters, by making use of a representative from each the five genera included in this study. For this purpose, CLC Genomics Workbench v. 8.0.1 (CLCBio, Aarhus Denmark) and Primer3 v. 0.4.0 (http://bioinfo.ut.ee/primer3-0.4.0/) were used to design primers that would allow amplification of individual genes, as well as the regions between them. Correlation between predicted and observed amplicon sizes were used as evidence that the specific cluster was correctly assembled (see supplementary file S1 for details regarding DNA extraction methods, primer sequences, PCR conditions and expected product sizes).

1.2.2. Phylogenetic analysis

The inferred amino acid sequences for the PKS gene or only its KS domain (i.e., encoding beta-ketoacyl synthase) were subjected to phylogenetic analysis. The datasets used in these analyses consisted of the sequences of the various *Ceratocystidaceae* species used in this study, as well as those for representative taxa obtained from NCBI and Joint Genome Institute (JGI; https://jgi.doe.gov/our-science/science-programs/fungal-genomics/). For this purpose, blastp with the *Ceratocystidaceae* sequences were used to query NCBI's protein database, as well as that of the JGI via MycoCosm (Grigoriev et al 2014). In both cases, the top sequence hits were downloaded and included in our gene datasets.

Individual gene sequence sets were subjected to multiple sequence alignment based on fast Fourier transform (MAFFT; http://mafft.cbrc.jp/alignment/server/) (Katoh et al 2017). Phylogenetic trees were inferred with Molecular Evolutionary Genetics Analysis (MEGA) v. 7 (Tamura et al 2013) by making use of amino acid distances (based on the number of differences) and neighbor-joining (Saitou and Nei 1987). We also used MEGA to infer maximum likelihood phylogenies for the different datasets by making us of the model

parameters indicated by ProtTest (Abascal et al 2005). In both cases, branch support was estimated using 1000 bootstrap replicates.

1.3. Results

1.3.1. Identification and annotation of PKS clusters

Before identifying PKS genes and clusters, we first evaluated the relative completeness of the genome assemblies used in this study. The analyses with BUSCO showed that most assemblies were characterized by a high level of completeness as they lacked only 0-1.8% of the conserved single-copy genes common to all fungi (Table 1). These data thus showed that the genomes examined would allow for meaningful comparisons. To compensate for the two assemblies with slightly poorer completeness scores (i.e., *C. fimbriata* and *H. omanensis* that respectively lacked 7.6% and 3.5%), we included multiple other representatives from the two genera with high-quality assemblies in our dataset.

Our analyses with antiSMASH, SMURF, InterProScan, SMIPs, MOTIF and blastp against NCBI's non-redundant protein database indicated that all 20 of the *Ceratocystidaceae* genomes encode PKS genes with high similarity to homologs in various *Sordariomycetes* (Tables 2-4; Supplementary file S2). Examination of the regions surrounding the respective PKS genes revealed that they occur in clusters that appeared to be conserved in some members of the *Ceratocystidaceae*. A total of eight such clusters were identified among the genomes examined (Figures 1 and 2; see supplementary file S3 for details regarding the genomic annotation of these cluster). By using a representative from each genus, the organization of these clusters were confirmed using PCR analysis of the inter- and intragenic regions. In these analyses, correlation between predicted and observed PCR fragment sizes provided evidence that the specific cluster was correctly assembled (Supplementary file S1).

Table 2. Comparison of the size (in amino acid residues, aa) and domain structure of the *Ceratocystidaceae* PKS-III proteins with that of the top BlastP hit in the NCBI database.

***Ceratocystidaceae* PKS-III**	**Size (aa)**	**Domain Organization***	**Information regarding the top BlastP hit**			
			Accession number	**Species**	**Size (aa)**	**Domain organization***
Davidsoniella	426	CHS	KFA65321	*Stachybotrys chlorohalonata*	1159	CHS
Thielaviopsis	435	CHS	KXH46773	*Colletotrichum simondsii*	442	CHS
Ceratocystis	434	CHS	KFA54710	*Stachybotrys chlorohalonata*	449	CHS
Bretziella	839	CHS-MviM	OAA68032	*Sporothrix insectorum*	464	CHS
Huntiella	739	CHS-MviM	KDN65911	*Colletotrichum sublineola*	442	CHS
Endoconidiophora	426	CHS	XP_016585873	*Sporothrix schenckii*	440	CHS

*CHS and MviM correspond to chalcone and stilbene synthase and dehydrogenase domains respectively.

Table 3. Comparison of the size (in amino acid residues, aa) and domain structure of the *Ceratocystidaceae* NR-PKS-I proteins with that of the top BlastP hit in the NCBI database.

***Ceratocystidaceae* NR-PKS-I**	**Size (aa)**	**Domain Organization***	**Information regarding the top BlastP hit**			
			Accession number	**Species**	**Size (aa)**	**Domain Organization***
Davidsoniella	2185	KS-AT-DH-ACP-ACP-TE	XP016640516	*Scedosporidium apiospermum*	2166	KS-AT-DH-ACP-ACP-TE
Thielaviopsis	2184	KS-AT-DH-ACP-ACP-TE	AHA15770	*Scedosporidium boydii*	2171	KS-AT-DH-ACP-ACP-TE
Ceratocystis	2195	KS-AT-DH-ACP-ACP-TE	AHA15770	*Scedosporidium boydii*	2171	KS-AT-DH-ACP-ACP-TE
Bretziella	2214	KS-AT-DH-ACP-ACP-TE	OLN85997	*Colletotrichum chlorophyti*	2184	KS-AT-DH-ACP-ACP-TE
Huntiella	2175	KS-AT-DH-ACP-ACP-TE	XP_018155915	*Colletotrichum higginsanum*	2198	KS-AT-DH-ACP-ACP-TE
Endoconidiophora	2192	KS-AT-DH-ACP-ACP-TE	XP016640516	*Scedosporidium apiospermum*	2166	KS-AT-DH-ACP-ACP-TE

*KS AT DH ACP and TE correspond to ketosynthase acyltransferase dehydratase acyl carrier protein and thioesterase domains respectively.

Table 4. Comparison of the size (in amino acid residues) and domain structure of the *Ceratocystidaceae* R-PKS-I proteins with that of the top BlastP hit in the NCBI database.

***Ceratocystidaceae* R-PKS-I**	**Size (aa)**	**Domain Organization***	**Information regarding the top BlastP hit**			
			Accession	**Name**	**Size (aa)**	**Domain organization***
R-PKS-I homolog A						
Ceratocystis	2272	KS-AT-DH-ER-KR-ACP	ALQ32773	*Fusarium babinda*	2354	KS-AT-DH-ER-KR-ACP
C. adiposa	2299	KS-AT-DH-ER-KR-ACP	ETS00110	*Trichoderma reesei*	2348	KS-AT-DH-ER-KR
B. fagacearum	2274	KS-AT-DH-ER-KR-ACP	XP_018157330	*Colletotrichum higginsianum*	2358	KS-AT-DH-ER-KR
Davidsoniella	2312	KS-AT-DH-ER-KR-ACP	ALQ32896	*Fusarium poae*	2346	KS-AT-DH-ER-KR-ACP
Endoconidiophora	2312	KS-AT-DH-ER-KR-ACP	ALQ32896	*Fusarium poae*	2346	KS-AT-DH-ER-KR-ACP
Thielaviopsis	2322	KS-AT-DH-ER-KR-ACP	ALQ32804	*Fusarium commune*	2351	KS-AT-DH-ER-KR-ACP
Huntiella	2300	KS-AT-DH-MET-ER-KR-ACP	ALQ32840	*Fusarium euwallaceae*	2318	KS-AT-DH-ER-KR-ACP
R-PKS-I homolog B						
Davidsoniella	2539	KS-AT-DH-MET-ER-KR-ACP	KJZ77319	*Hirsutella minnesotensis*	2477	KS-AT-DH-ER-KR-ACP
Endoconidiophora	2641	KS-AT-DH-MET-ER-KR-ACP	KPM39648	*Neonectria ditissima*	2595	KS-AT-DH-MET-ER-KR-ACP
Huntiella	2488	KS-AT-DH-MET-ER-KR-ACP	KXX79732	*Madurella mycetomatis*	2625	KS-AT-DH-MET-ER-KR-ACP
R-PKS-I homolog C						
T. punctulata	2551	KS-AT-DH-MET-ER-KR-ACP	KJZ77319	*Hirsutella minnesotensis*	2477	KS-AT-DH-MET-ER-KR-ACP
Davidsoniella	2565	KS-AT-DH-MET-ER-KR-ACP	KJZ77319	*Hirsutella minnesotensis*	2477	KS-AT-DH-MET-ER-KR-ACP
Endoconidiophora	2547	KS-AT-DH-MET-ER-KR-ACP	KJZ77319	*Hirsutella minnesotensis*	2477	KS-AT-DH-MET-ER-KR-ACP
R-PKS-I homolog D						
Davidsoniella	2513	KS-AT-DH-MET-ER-KR-ACP	KPM39648	*Neonectria ditissima*	2595	KS-AT-DH-MET-ER-KR-ACP
Thielaviopsis	2546	KS-AT-DH-MET-ER-KR-ACP	KJZ77319	*Hirsutella minnesotensis*	2477	KS-AT-DH-MET-ER-KR-ACP

*KS AT ACP DH KR ER and MET correspond to ketosynthase acyltransferase acyl carrier protein dehydratase ketoreductase Enoylreductase and Methyltransferase domains respectively.

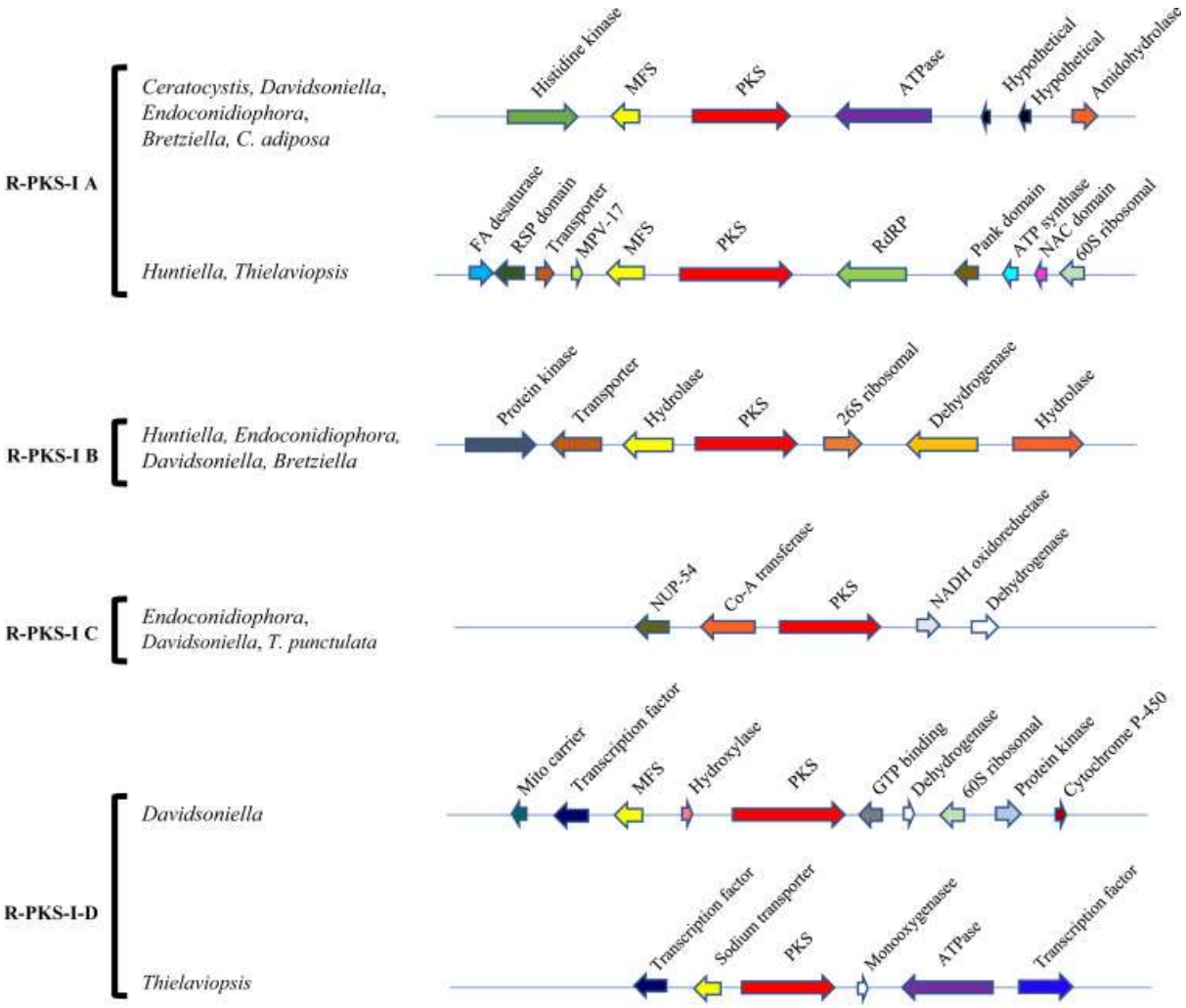

Fig 1. Gene clusters identified in the *Ceratocystidaceae* as containing type I reducing polyketide synthase (R-PKS-I) genes. The clusters were named (A-D) based on the phylogenetic relatedness among their PKS genes (see Figure 5). Coding region and orientation are indicated with block arrows. Annotation information in the genomes of the respective isolates examined are provided in Supplementary File S3. Abbreviations are as follows: MFS = Major facilitator protein, PKS = Polyketide synthase, FA desaturase = Fatty acid desaturase, RDRP = RNA dependent RNA polymerase, NAC = NAC transcription factor domain containing protein, GTP binding = guanine trimeric nucleotide-binding protein, Co-A transferase = Co acetyl transferase and NUP-54 = Nucleoporin 54.

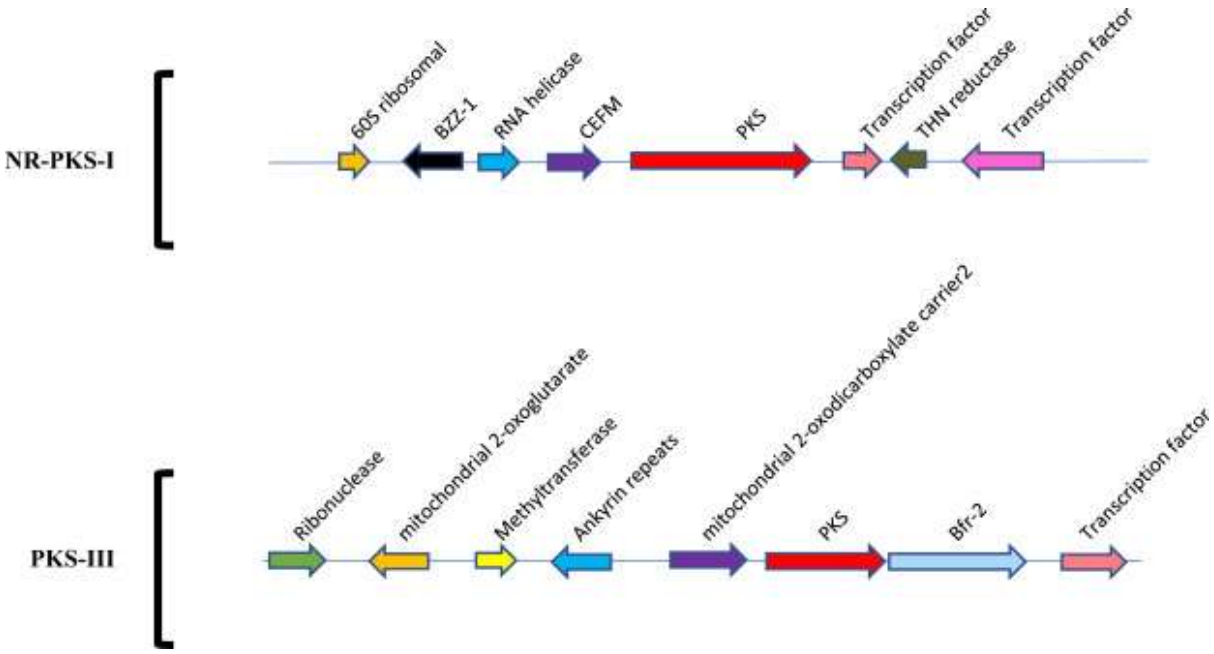

Fig 2. Gene clusters identified in the *Ceratocystidaceae* as containing respectively type I non-reducing polyketide synthase (NR-PKS-I) and type III polyketide synthase (PKS-III) genes. Coding region and orientation are indicated with block arrows. Annotation information in the genomes of the respective isolates examined are provided in Supplementary File S3. Abbreviations are as follows: BZZ-1 = polymerization protein, CEFM = major facilitator superfamily, THN reductase = Tetrahydroxynaphthalene reductase.

The identified clusters were grouped based on the domain structure of the predicted PKS proteins (Tables 2-4). One of the clusters included a putative PKS-III (Figure 2; Table 2) and a single copy of this cluster was also present in all of the *Ceratocystidaceae* genomes examined. The remaining seven clusters all encoded a putative PKS-I that, based on the presence/absence of certain domains, were further grouped into those containing R-PKS-I (Figure 1) and NR-PKS-I (Figure 2) genes. A single copy of the cluster containing NR-PKS-I (designated as such due to absence of KR, ER or DH domains) also occurred in all of the genomes examined (Figure 2, Table 3). The remaining six clusters all contained a putative R-PKS-I because of the presence of domains implicated in ketide reduction (i.e., KR, ER and DH) (Figure 5, Table 4).

1.3.2. Phylogenetic analysis

Phylogenetic analysis separated the *Ceratocystidaceae* sequences into a distinct and well supported clade in both the PKS-III and NR-PKS-I trees (Figures 3 and 4). In these trees, similar groupings among the *Ceratocystidaceae* sequences were also observed. For example, the *Ceratocystis* sequences grouped together and formed the sister clade of a broader clade

containing all the other *Ceratocystidaceae* sequences. The only exceptions were the *B. fagacearum* and *C. adiposa* sequences, which were nested within this broader *Ceratocystidaceae* clade. These groupings generally mirrored the taxonomy and known relationships of the isolates from which the sequences came; i.e., the respective *Huntiella* and *Thielaviopsis* formed distinct clades, and those from *Endoconidiophora* and *Davidsoniella* formed a clade (De Beer et al 2014).

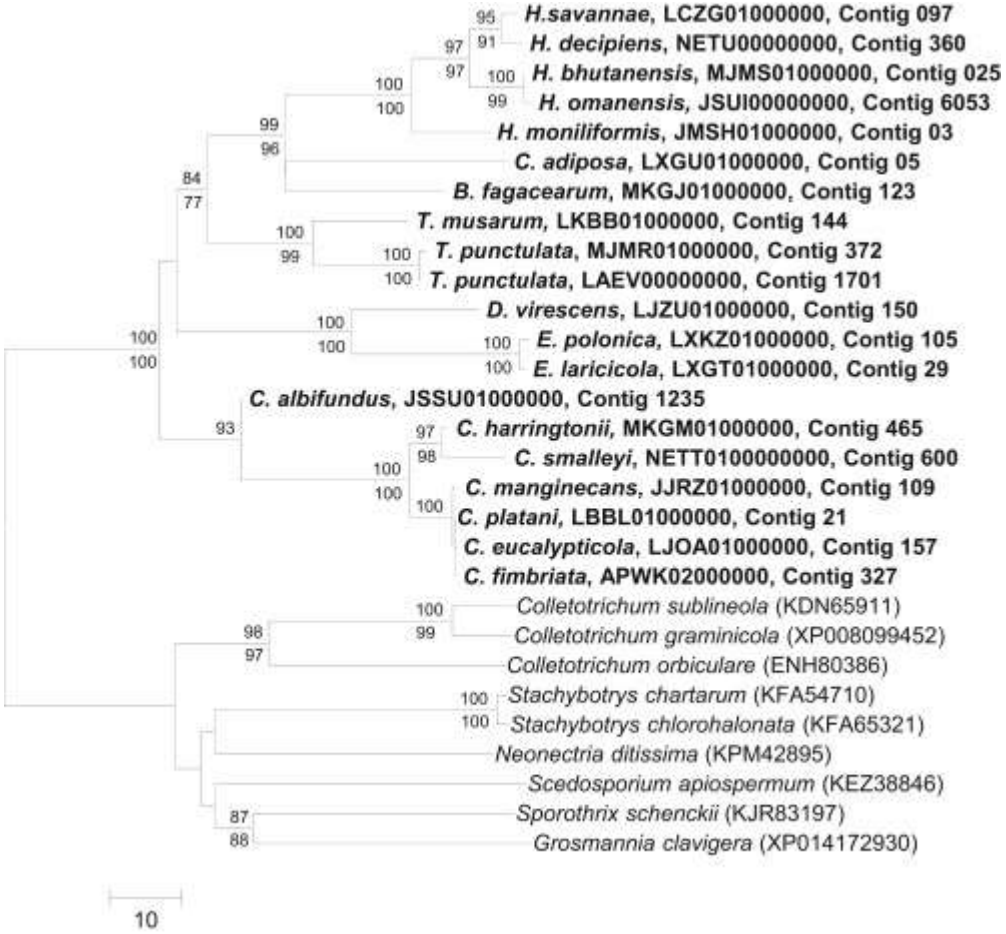

Fig 3. Neighbor-joining tree inferred from the PKS-III gene sequences examined in this study**.** GenBank accession numbers or sequence identifiers from genome projects for each sequence are provided in parentheses. Similar groups were obtained with the maximum likelihood analysis. Percentage bootstrap support (based on a 1000 repeats) is indicated at the internodes, with those for the neighbour-joining distance-based analysis above the branches and those for the maximum likelihood below the branches.

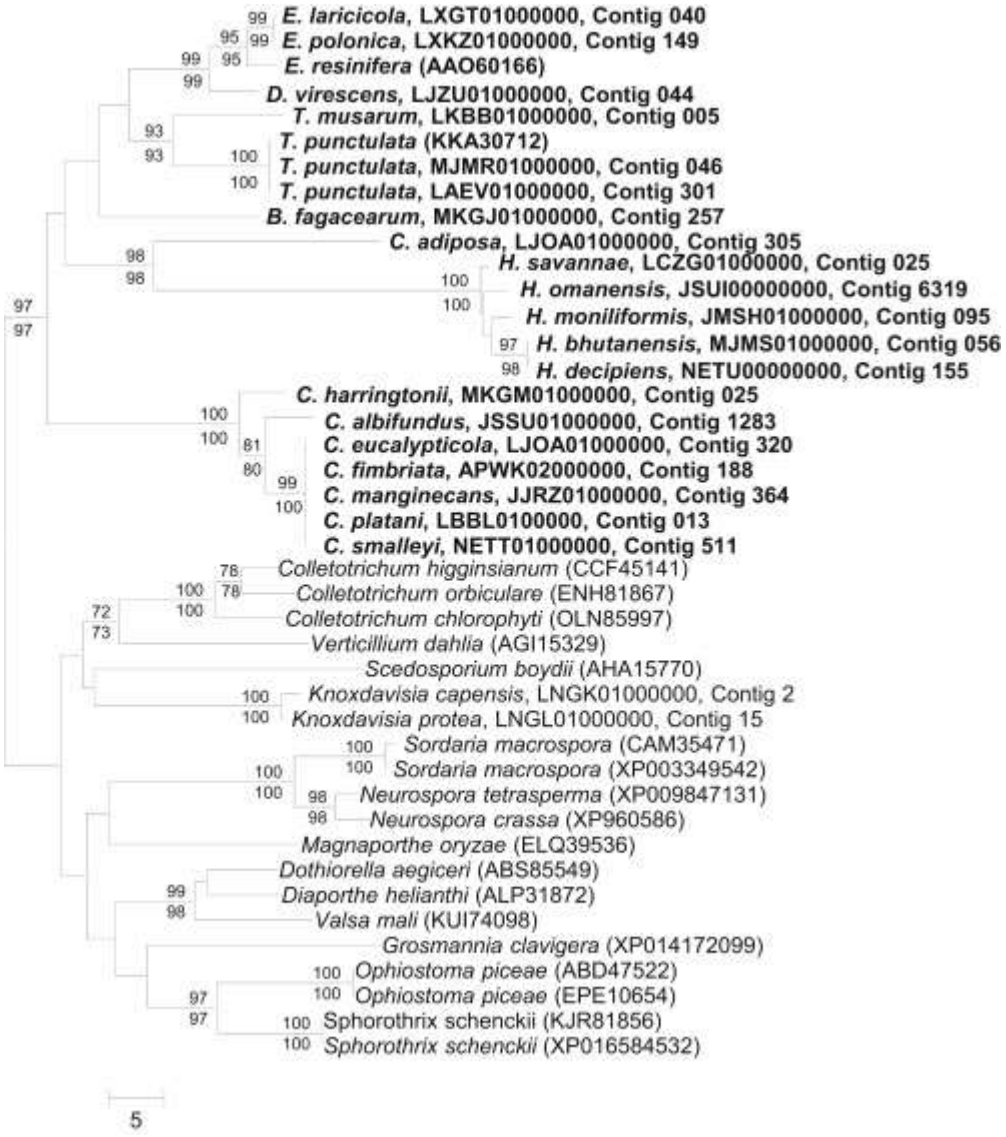

Fig 4. Neighbor-joining tree inferred from the beta-ketoacyl synthase (KS) domain of the non-reducing PKS-I gene sequences examined in this study**.** GenBank accession numbers or sequence identifiers from genome projects for each sequence are provided in parentheses. Similar groups were obtained with the maximum likelihood analysis. Percentage bootstrap support (based on a 1000 repeats) is indicated at the internodes, with those for the neighbour-joining distance-based analysis above the branches and those for the maximum likelihood below the branches.

Phylogenetic analysis of the KS domain of the predicted R-PKS-I protein separated the *Ceratocystidaceae* into four well-supported clades, which we designated A-D (Figure 5). The respective R-PKS-I genes included in each of these clades were designated accordingly (i.e., R-PKS-I A through to R-PKS-I D) (see Table 3). We also applied these designations to the gene clusters in which the respective PKS genes occurred (see Figure 1).

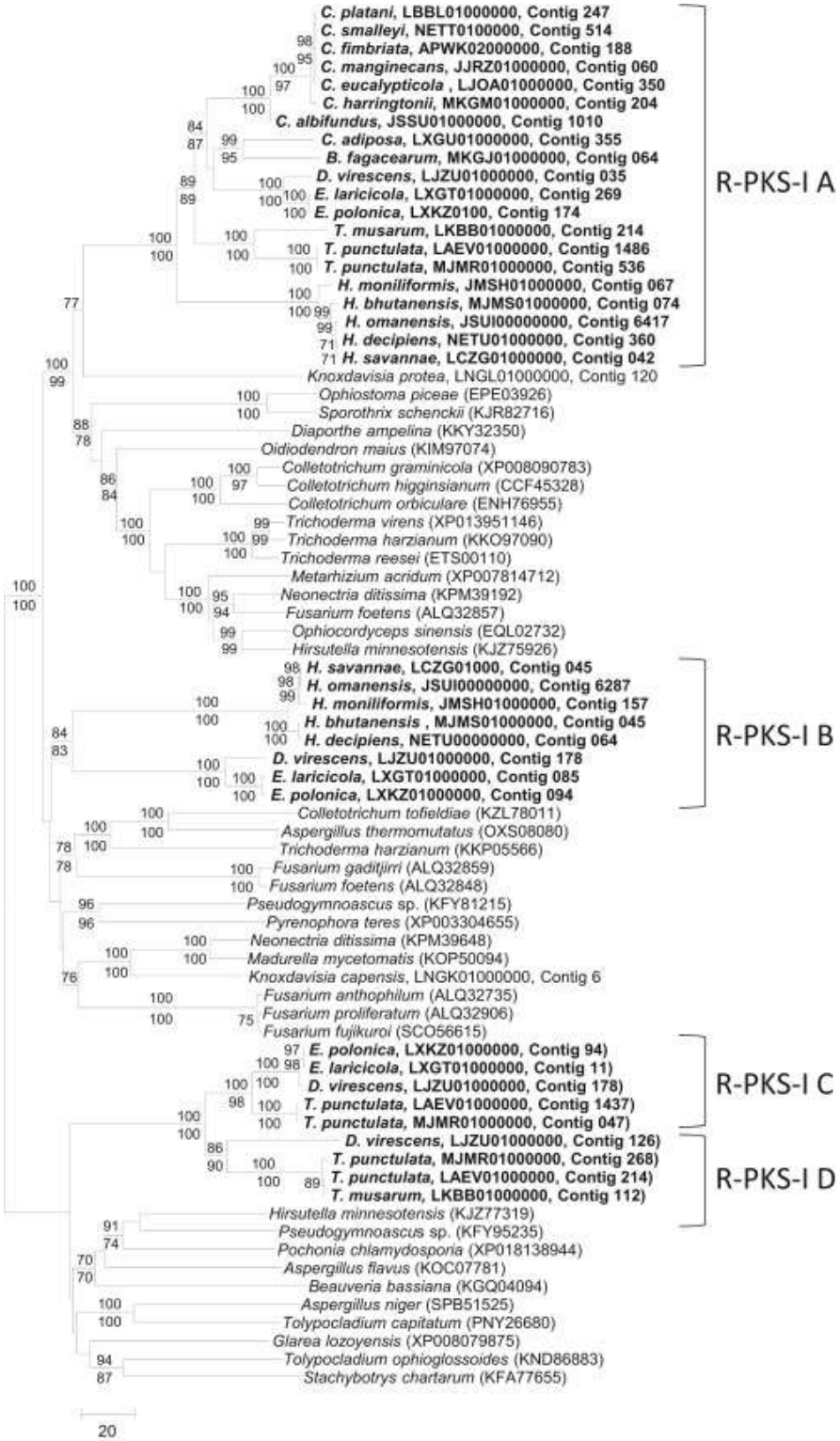

Fig 5. Neighbor-joining tree inferred from the beta-ketoacyl synthase (KS) domain of the reducing PKS-I gene sequences examined in this study. The four main clades recovered for the *Ceratocystidaceae* are indicated in bold with the letters A-D. Similar groups were obtained with the maximum likelihood analysis. GenBank accession numbers or sequence identifiers from genome projects for each of these sequences are provided in parentheses. Percentage bootstrap support (based on a 1000 repeats) is indicated at the internodes, with those for the neighbour-joining distance-based analysis above the branches and those for the maximum likelihood below the branches.

R-PKS-I clade A included sequences from all 20 of the genomes examined in this study, although the corresponding cluster of the *Huntiella* species differed dramatically from that of the other *Ceratocystidaceae* species (Figures 1 and 5). The groupings among the sequences in clade A broadly corresponded with taxonomic expectations (De Beer et al 2014). Clade D included sequences from all the *Thielaviopsis* and *Davidsoniella* genomes examined, although the gene clusters containing these R-PKS-I genes were markedly different between the two genera (Figure 1). Clade B included KS sequences from all the *Endocondiophora* and *Huntiella* species examined, as well as *Davidsoniella*. Clade C included sequences from all the analysed *Davisoniella* and *Endoconidiophora* genomes, as well as *Thielaviopsis*. In both clades B and C, the gene clusters in which their corresponding R-PKS-I genes occurred were similar in content and organization (Figure 1). Overall, the R-PKS-I phylogeny suggested multiple and independent evolutionary origins for the gene in the *Ceratocystidaceae* (Figure 5). Clades A and B were respectively nested within larger clades containing homologous sequences from other *Sordariomycetes*, including *Knoxdaviesia sp.* (Microascales). Clades C and D had a sister-group relationship and was also nested within a larger clade containing *Sordariomycetes* homologs.

1.3.3. PKS biosynthetic gene cluster structure and predicted products

The NR-PKS-I cluster was conserved and present in all of the *Ceratocystidaceae* genomes examined (Figure 2). It consisted of genes encoding 60S ribosomal protein, actin polymerization protein BZZ-1, ATP-dependent RNA-helicase, a transcription factor, 1,3,8-tetrahydroxynaphthalene reductase, a dehydrogenase and a transmembrane protein. These protein sequences showed high similarity to those from a range of *Sordariomycetes* including *Sedosporium boydii* and *Colletotrichum higginsianum* (supplementary file S3). All of the *Ceratocystidaceae* NR-PKS-I genes also encoded a protein consisting of 2175-2195 amino acid residues with the same domain structure (i.e., KS-AT-DH-ACP-ACP-TE).

ClustScan could not predict a product for the *Ceratocystidaceae* NR-PKS-I cluster, but it likely allows for the formation of melanin in these fungi. This is because the NR-PKS-I gene and cluster showed high similarity to those from other fungi that produce melanin (Figure 2 and supplementary files S2 and S3). Also, fungal melanin biosynthesis is known to require a 1,3,8-trihydroxynaphthalene reductase (Chumley and Valent 1990; Perpetua et al 1996) and a

transcription factor (Cho et al 2012), and the genes for both are contained within the *Ceratocystidaceae* NR-PKS-I cluster (see Figure 2).

Although, ClustScan could not predict products for any of the gene clusters containing R-PKS-I homologs, similarity with known PKS clusters indicated a possible role for the R-PKS-I B cluster of *Davidsoniella*, *Huntiella* and *Endoconidiophora*, and the R-PKS-I D cluster in *Davidsoniella*. The latter comprised genes coding for the PKS (designated here as R-PKS-I), a GTP-binding protein, dehydrogenase, 60S ribosomal protein, protein kinase, hydroxylase, cytochrome P-450, Mitochondrial carrier protein, transcription factor and Major facilitator protein (Figure 1; supplementary file S3). A cluster with similar architecture and gene content also has been identified in *Acremonium strictum*, where it is involved in methylorcinaldehyde synthesis (Bailey et al 2007).

The R-PKS-I B cluster identified in *Davidsoniella*, *Huntiella*, *Endoconidiophora* and *Bretziella* included genes coding for a dehydrogenase, transporter, a hydrolase, the PKS (designated here as R-PKS-I), cytochrome P-450, a 26S ribosomal protein, dehydrogenase and a hydrolase (Figure 1; supplementary file S3). A cluster with this organization and gene content has been previously implicated in the production of mycophenoloc acid in *Penicillium brevicompactum* (Regueira et al 2011; Hansen et al 2011). The core gene of the cluster also showed a very high similarity to *Madurella mycetomattis* lovastatine diketide (supplementary file S3).

Although the remaining R-PKS-I clusters and their genes showed high similarity to those in other fungi (Figure 2 and supplementary files S2 and S3), ClustScan could also not predict a product for any of them. Based on their gene content, however, the R-PKS-I A cluster probably does not represent a functional PKS biosynthesis cluster. This is because the region surrounding the PKS gene did not contain genes encoding regulatory proteins, proteins that would facilitate post-translational modification of the parent polyketide, and cellular transport (Brown et al 2012; Osbourn 2010). This is contrast to the situation for the R-PKS-I D cluster in *Thielaviopsis* and the R-PKS-I C cluster in *Davidsoniella*, *Bretziella* and *Endoconidiophora* and *Huntiella*. The latter cluster includes genes encoding multiple dehydrogenases, an oxidorecutase and co-A transferase, while the *Thielaviopsis* R-PKS-I D cluster contains genes encoding transcription factors, a transporter and monooxygenase (Figure 2 and S3). These genes are typically encoded in functional PKS biosynthetic clusters

(Khaldi et al 2010; Keller and Hohn 1997), but their functionality in the *Ceratocystidacea* requires experimental verification.

The PKS-III containing gene cluster, is highly conserved and present in all 20 of the *Ceratocystidaceae* genomes examined (Figure 2). It is composed of a ribonuclease enzyme, a mitochondrial 2-oxoglutarate/malate carrier protein, a methylteransferase domain containing protein, an ankyrine repeats containing protein, mitochondrial 2-oxodicarboxylate carrier 2, a chalcone synthase A (here designated as PKS-III), Bfr-2 and a transcription factor. This cluster thus has the requisite gene repertoires of a functional PKS cluster (Seshime et al 2005), and these genes showed high similarity to those in a range of *Sordariomycetes*, including *Stachybotrys*, *Colletotrichum* and *Neonectria* (supplementary file S3).

1.4. Discussion

In this study we identified and characterized the repertoire of PKS-containing gene clusters potentially involved the production of polyketides in the *Ceratocystidaceae*. Our results showed that the *Ceratocystidaceae* species harbour between three and six putative PKS clusters (two of which are likely not functional). This is comparable to what is known for other Ophiostomatoid fungi (Seifert et al 2013), e.g. *Ophiostoma ulmi* and *O. novo-ulmi* have ten PKS clusters (Sbaraini et al 2017) and *Grosmannia clavigera* has six (Lah et al 201). In general, however, the *Ceratocystidaceae* have much fewer PKS clusters than fungi such as *Trichoderma* (Mukherjee et al 2012) and *Fusarium* (Hansen et al 2015). All of the genomes examined in our study contained clusters containing a PKS-III, a NR-PKS-I and at least one cluster with a R-PKS-I gene. The only other member of this family for which PKS genes and/or clusters have been characterized is *E. resinifera* in which NR-PKS-I was needed for melanin production (Loppnau et al 2004). Our study thus presents a comprehensive foundation for exploring the roles of polyketide biosynthesis in this economically important group of fungi.

The *Ceratocystidaceae* NR-PKS-I sequence and domain structure is highly similar to its homologs in *Neurospora crassa* and *Colletothrichum sublineola* that are involved in melanin production (Baroncelli et al 2014). The *Ceratocystidaceae* cluster containing this gene did not exactly match those of other fungi, but the organization of this cluster is not conserved in all melanin producing fungi (Keller and Hohn 1997). The fungal melanin biosynthesis PKS is

a large multifunctional protein containing KS and AT domains, two tandem ACP domains, and a TE domain (sometimes Claisen cyclase) (Vagstad et al 2012). It catalyses the synthesis of the pentaketide 1,3,6,8-tetrahydroxynaphthalene, which then undergoes successive reduction-dehydration steps to yield 1,8-dihydroxynaphthalene (DHN) from which melanin (in fungi referred to as DHN-melanin) is ultimately produced (Butler and Day 1998). Given its high similarity to the PKS involved in DHN-melanin biosynthesis and its presence in all of the genomes examined, our future research will thus seek to determine whether the *Ceratocystidaceae* NR-PKS-I is indeed involved in the production of this important and widely distributed compound (Gómez and Nosanchuk 2003). For example, in *E. resinifera*, disruption of this PKS led to the generation of albino strains (Loppnau et al 2004). Melanin production is also known to be an important factor for pathogenicity in high number of Ascomycetes (Butler and Day 1998).

Despite the fact that PKS-III is not generally known from the fungi, all of the *Ceratocystidaceae* genomes that were examined here showed a highly conserved cluster that contained a homolog of the gene encoding chalcone synthase A (CHS). This cluster was also very similar to that reported in *Stachybotrys chlorohalonata* (Semeiks et al 2014). Few studies have examined the role of this enzyme in fungi, but expression analyses have revealed that all of the PKS-III genes examined so far, are responsible for the biosynthesis of polyketide pyrones and resorcinols (Hashimoto et al 2014). It is thus attractive to speculate that the PKS-III identified in this study might have a similar function in the *Ceratocystidaceae*.

PKS-III does not normally have multiple catalytic domains (Chan et al 2009), however our study presents an example of where this might not be the case. The PKS-III in all of the *Huntiella* species examined included a dehydrogenase domain together with the normal CHS domain. However, structural diversity in the polyketide compounds produced by PKS-III usually stem from the diversity of starter and extender units used and other enzymatic modifications of the parent compound (Chan et al 2009; Brown et al 2012). Therefore, the inclusion of additional catalytic domains may contribute to the diversity of products that could potentially be produced by such PKS-III. It is anticipated that with the increasing availability of whole genome sequences and development of bioinformatics tools, other fungal PKS-IIIs with new and unusual biosynthetic capabilities will be discovered (Vijayan et al 2011), as have been shown in plants and bacteria (Morita et al 2010). In fact, according to

the knowledge gained from these previous studies on diverse organisms, it should be likely to fast track our understanding of the biosynthetic machinery associated with fungal PKS-III clusters.

As we mentioned earlier, except for NR-PKS-I, none of the other PKS clusters examined were predicted to produce a known compound. Although some of the clusters and core genes displayed to other known PKSs, functional analysis is still required. It is thus possible that some of our clusters may produce novel compounds. The thioesterase domain is absent in some of the identified R-PKS-I genes (TE or CYC), which suggests that the fungi might be unable to release their products. As a result, these PKS clusters could probably synthesize starter units for other enzymatic systems that have the domain needed for release of the final product, as described for zearalenone (Gaffoor and Trail 2006; Kim et al 2005; Lysøe et al 2006). Alternatively, these clusters could rely on strange accessory enzymes for catalysing the metabolites to be released.

Distinctive evolutionary histories were inferred for the three *Ceratocystidaceae* PKS types. The NR-PKS-I and PKS-III clusters were conserved across the family, and the phylogeny inferred from their PKS sequences broadly corresponded with the expected relationships among the fungi included (De Beer et al 2014). This suggests that the respective clusters were already present in the ancestor of the *Ceratocystidaceae* and that their genetic divergence generally matched that of the species hosting them. However, our data for the R-PKS-I genes and clusters indicated a scenario involving multiple and independent origins. In the case of R-PKS-I A, a homolog might have been present in the *Ceratocystidaceae* ancestor, but its domain structure and overall gene cluster architecture have changed to a point where the *Huntiella* cluster differed dramatically from those of the other species. In the case of the R-PKS-I B and C clusters, homologous regions were likely acquired only in specific lineages of the *Ceratocystidaceae*. For example, the R-PKS-I B cluster was present in all of the *Endoconidiophora, Davidsoniella* and *Huntiella* genomes examined, although *Huntiella* do not represent a close relative of these fungi (De Beer et al 2014). The same was also true for the R-PKS-I D in the more distantly related *Thielaviopsis* and *Davidsoniella* species, although acquisition of the gene was also associated with the development of distinctly different gene clusters in the two groups of fungi.

Various possible evolutionary mechanisms explain the PKS clusters' presence and distribution within the *Ceratocystidaceae* genomes examined. Vertical inheritance from parent to offspring seems to be the most suitable explanation for the evolution of the family's highly conserved NR-PKS-I and PKS-III clusters. By contrast, some of the R-PKS-I clusters were likely acquired in a lineage-specific fashion via horizontal gene transfer (HGT), as have been shown for many secondary metabolite biosynthesis clusters in other fungi (Keller et al 2005; Osbourn 2010). For example, HGT was recently used to explain the distribution of a fujikurin-like gene cluster among a wide range and often unrelated plant-associated fungi (Sbaraini et al 2013). This is particularly true for the R-PKS-I B and C clusters, although an alternative hypothesis for their distribution would involve ancestral duplication, followed by subsequent divergence and gene/cluster loss (Kroken et al 2003). We note, however, that neither of these hypotheses (i.e., HGT and duplication followed by loss) alone can explain the non-homologous structures observed among the R-PKS-I A clusters and among those for R-PKS-I D. In these instances, initial acquisition (via HGT or gene/cluster duplication and losses) combined with *de novo* cluster formation through rearrangement of existing genetic elements most probably led to the clusters observed in extant *Ceratocystidaceae* (Osbourn 2010).

Acknowledgments

We thank the University of Pretoria and members of the Tree Protection Cooperative Programme for funding, as well as the South African Department of Science and Technology and National Research Foundation (NRF) for funding through the Centre of excellence (Centre of Excellence in Tree Health Biotechnology) and the DST-NRF-SARChI chair (Fungal genomics) programmes.

References

Abascal F, Zardova R, Posada D, 2005. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **21**: 2104-2105.

Austin MB, Noel JP, 2003. The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports* **20**: 79–110.

Bachmann BO, Ravel J, 2009. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods in Enzymology* **458**: 181-217.

Bailey AM, Cox RJ, Harley K, Lazarus CM, Simpson TJ, Skellam E, 2007. Characterisation of 3-methylorcinaldehyde synthase (MOS) in *Acremonium strictum*: First observation of a reductive release mechanism during polyketide biosynthesis. *Chemical Communications* **39**: 4053-4055.

Baroncelli R, Sanz-Martín JM, Rech GE, Sukno SA, Thon MR, 2014. Draft genome sequence of *Colletotrichum sublineola*, a destructive pathogen of cultivated sorghum. *Genome Announcments* **2**: e00540-14.

Beck J, Ripka S, Siegner A, Schiltz E, Schweizer E, 1990. The multifunctional-methylsalicylic acid synthase gene of *Penicillium patulum*. Its gene structure relative to that of other polyketide synthases. *European Journal of Biochemistry* **192**: 487–498.

Belbahri L, 2015. Genome sequence of *Ceratocystis platani*, a major pathogen of plane trees. URL http://www.ncbi.nlm.nih.gov/ nuccore/814603118.

Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, De Los Santos EL, Kim HU, Nave M, Dickschat JS, 2017. antiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research* **45**: W36-41.

Brown DW, Butchko RA, Baker SE, Proctor RH, 2012. Phylogenomic and functional domain analysis of polyketide synthases in *Fusarium*. *Fungal Biology* **116:** 318–31.

Butler MJ, Day AW, 1998. Fungal melanins: a review. *Canadian Journal of Microbiology* **44**: 1115-36.

Cane DE, Walsh CT, Khosla C, 1998. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**: 63–68.

Chan YA, Podevels AM, Kevany BM, Thomas MG, 2009. Biosynthesis of polyketide synthase extender units. *Natural Product Reports* **26**: 90–114.

Chang PK, Cary JW, Ju J, Bhatnagar D, Cleveland TE, 1995. The *Aspergillus parasiticus* polyketide synthase gene pksA, a homolog of *Aspergillus nidulans* wA, is required for aflatoxin B1 biosynthesis. *Molecular Genomics and Genetics* **248:** 270–277.

Cho Y, Srivastava A, Ohm RA, Lawrence CB, Wang K-H, Grigoriev IV, 2012. Transcription factor Amr1 induces melanin biosynthesis and suppresses virulence in *Alternaria brassicicola*. *Plos Pathogens* **8**: e1002974.

Chumley FG, Valent B, 1990. Genetic analysis of melanin-deficient, nonpathogenic mutants of *Magnaporthe grisea*. *Molecular Plant-Microbe Interactions* **3**: 135–143.

Cox RJ, 2007. Polyketides proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Organic and Biomolecular Chemistry* **5:** 2010-2026.

Cox RJ, Simpson TJ, 2010. Comprehensive Natural Products II: chemistry and biology. ed. L Mander and H-W Liu, Elsevier, Oxford **1:** 347.

De Beer ZW, Marincowitz S, Duong TA, Wingfield MJ, 2017. *Bretziella*, a new genus to accommodate the oak wilt fungus, *Ceratocystis fagacearum* (Microascales, Ascomycota). *MycoKeys* **27**: 1-19

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

Fulton TR, Ibrahim N, Losada MC, Grzegorski D, Tkacz JS, 1999. A melanin polyketide synthase (PKS) gene from *Nodulisporium* sp. that shows homology to the pks1 gene of *Colletotrichum lagenarium*. *Molecular Genomics and Genetics* **262:** 714–720.

Gaffoor I, Trail F, 2006. Characterization of two polyketide synthase genes involved in zearalenone biosynthesis in *Gibberella zeae*. *Applied and Environmental Microbiology* **72**: 1793–1799.

Gómez BL, Nosanchuk JD, 2003. Current opinion in infectious diseases, melanin and fungi. *Current Opinion in Infectious Diseases* **16**: 91–96.

Graziani S, Vasnier C, Daboussi MJ, 2004. Novel polyketide synthase from *Nectria haematococa*. *Applied and Environmental Microbiology* **70:** 2984–2988.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I, 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: D699-704.

Hansen BG, Salomonsen B, Nielsen, M, Nielsen J, Hansen N, Nielsen K, Regueira T, Nielsen J, Patil K, Mortensen UH, 2011. Versatile enzyme expression and characterization system for *Aspergillus nidulans*, with the *Penicillium brevicompactum* polyketide synthase gene from the mycophenolic acid gene cluster as a test case. *Applied and Environmental Microbiology* **77**: 3044 –3051.

Hansen FT, Gardiner DM, Lysøe E, Fuertes PR, Tudzynski B, Wiemann P, Sondergaard TE, Giese H, Brodersen DE, Sørensen JL, 2015. An update to polyketide synthase and non-

ribosomal synthetase genes and nomenclature in *Fusarium*. *Fungal Genetics and Biology* **75**: 20–29.

Hashimoto M, Nonaka T, Fujii I, 2014. Fungal type III polyketide synthases. *Natural Product Reports* **10**: 1306-1317.

Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, Yamazaki S, Fujita N, 2012. DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. *Nucleic Acids Research* **41:** D408–D414.

Katoh K, Rozewicki J, Yamada KD, 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. Bbx108. https://doi.org/10.1093/bib/bbx108.

Keller NP, Hohn TM, 1997. Metabolic pathway gene clusters in filamentous fungi. *Fungal Genetics and Biology* **21:** 17-29.

Keller NP, Turner G, Bennett JW, 2005. Fungal secondary metabolism – from biochemistry to genomics. *Natural Reviews Microbiology* **3:** 937–947.

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, et al, 2010. SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology* **47**: 736–741.

Khaldi N, Wolfe KH, 2011. Evolutionary origins of the fumonisin secondary metabolite gene cluster in *Fusarium verticillioides* and *Aspergillus niger*. *International Journal of Evolutionary Biology* **15**: 1–7.

Kim YT, Lee YR, Jin JM, Han KH, Kim H, Kim JC, Lee T, Yun SH, Lee YW, 2005. Two different polyketide synthase genes are required for synthesis of zearalenone in *Gibberella zeae*. *Molecular Microbiology* **58**: 1102–1113.

Kroken S, Glass NL, Taylor JW, Yoder OC, 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proceedings of the National Academy of Sciences* **100:** 15670-15675.

Lah L, Haridas S, Bohlmann J, Breuil C, 2013. The cytochromes P450 of *Grosmannia clavigera*: genome organization, phylogeny, and expression in response to pine host chemicals. *Fungal Genetics and Biology* **50**: 72-81.

Loppnau P, Tanguay P, Breuil C, 2004. Isolation and disruption of the melanin pathway polyketide synthase gene of the softwood deep stain fungus *Ceratocystis resinifera*. *Fungal Genetics and Biology* **41:** 33–41.

Lysøe E, Klemsdal SS, Bone KR, Frandsen RJN, Johansen T, Thrane U, Giese H, 2006. The PKS4 gene of *Fusarium graminearum* is essential for zearalenone production. *Applied and Environmental Microbiology* **72:** 3924–3932.

Ma LJ, Geiser DM, Proctor RH, Rooney AP, O' Donnell K, Trail F, Gardiner DM, Manners JM, Kazan K, 2013. *Fusarium* pathogenomics. *Annual Review of Microbiology*. **67**: 399-416.

Morita H, Abe I, Noguchi H, 2010. Comprehensive Natural Products II: chemistry and biology. ed. L Mander and H-W Liu, Elsevier, Oxford **1**: 171.

Mukherjee PK, Horwitz BA, Kenerley CM, 2012. Secondary metabolism in *Trichoderma* - a genomic perspective. *Microbiology* **158**: 35-45.

Osbourn A, 2010. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends in Genetics* **26**: 449-457.

Perpetua NS, Kubo Y, Takano Y, Furusawa I, 1996. Cloning and characterization of a melanin biosynthetic THR1 reductase gene essential for appressorial penetration of *Colletotrichum lagenarium*. *Molecular Plant-Microbe Interactions* **9**: 323–329.

Rawlings BJ, 1997. Biosynthesis of polyketides. *Natural Product Reports* **14:** 523-556.

Regueira TB, Kildegaard KR, Hansen BG, Mortensen UH, Hertweck C, Nielsen J, 2011. Molecular basis for mycophenolic acid biosynthesis in *Penicillium brevicompactum*. *Applied Environmental Microbiology* **77**: 3035–3043.

Saitou N, Nei M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406-425.

Sbaraini N, Andreis FC, Thompson CE, Guedes RLM, Junges Â, Campos T, Staats CC, Vainstein MH, Ribeiro de Vasconcelos AT, Schrank A, 2017. Genome-wide analysis of secondary metabolite gene clusters in *Ophiostoma ulmi* and *Ophiostoma novo-ulmi* reveals a Fujikurin-like gene cluster with putative role in infection. *Frontiers in Microbiology* **8**: 1063.

Seifert KA, de Beer ZW, Wingfield MJ, 2013. The ophiostomatoid fungi: expanding frontiers. CBS-KNAW Fungal Biodiversity Centre.

Semeiks J, Borek D, Otwinowski Z, Grishin NV, 2014. Comparative genome sequencing reveals chemotype-specific gene clusters in the toxigenic black mold *Stachybotrys*. *BMC Genomics* **15**: 590.

Seshime Y, Juvvadi PR, Fujii I, Kitamoto K, 2005. Discovery of a novel superfamily of type III polyketide synthases in *Aspergillus oryzae*. *Biochemical and Biophysical Research Communications* **331**: 253-60.

Shwab, EK, Keller NP, 2008. Regulation of secondary metabolite production in filamentous ascomycetes. *Mycological Research* **112:** 225–230.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva, EV, Zdobnov EM, 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

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.

Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D, 2008. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and *in silico* prediction of novel chemical structures. *Nucleic Acids Research* **36**: 6882– 6892.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725-2729.

Vagstad AL, Hill EA, Labonte JW, Townsend CA, 2012. Characterization of a fungal thioesterase having Claisen cyclase and deacetylase activities in melanin biosynthesis. *Chemical Biology* **19:** 1525–1534.

Van der Nest MA, Beirn LA, Crouch JA, Demers JE, De Beer ZW, De Vos L, Gordon TR, Moncalvo JM, Naidoo K, Sanchez-Ramirez S, Roodt 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**: 472-85.

Van der Nest MA, Bihon W, De Vos L, Naidoo K, Roodt D, Rubagotti E, Slippers B, Steenkamp ET, Wilken PM, Wilson A, Wingfield MJ, 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 MP, Wilken PM, Wingfield MJ, 2015. Saprophytic and pathogenic fungi in the Ceratocystidaceae differ in their ability to metabolize plant-derived sucrose. *BMC Evolutionary Biology* **15**: 273.

Vijayan M, Chandrika SK, Vasudevan, SE, 2011. PKSIIIexplorer: TSVM approach for predicting type III polyketide synthase proteins. *Bioinformation* **6**: 125–127.

Walton JD, 2000. Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: a hypothesis. *Fungal Genetics and Biology* **30**: 167-171.

Weber T, Kim HU, 2016. The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. *Synthetic and Systems Biotechnology* **1**: 69-79.

Wilken PM, Steenkamp ET, van der Nest MA, Wingfield MJ, De Beer ZW, Wingfield BD, 2018. Unexpected placement of the MAT1-1-2 gene in the MAT1-2 idiomorph of *Thielaviopsis*. *Fungal Genetics and Biology* DOI10.1016/j.fgb.2018.01.007.

Wilken PM, Steenkamp ET, Wingfield MJ, De Beer ZW, Wingfield BD, 2013. IMA Genome-F 1: Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. *IMA fungus* **4**: 357-358.

Wingfield BD, Ades PK, Al-Naemi FA, Beirn LA, Bihon W, Crouch JA, De Beer ZW, De Vos L, Duong TA, Fields CJ, Fourie G, 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**: 233-48.

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

Wingfield BD, Barnes I, De Beer ZW, De Vos L, Duong TA, Kanzi AM, Naidoo K, Nguyen HD, Santana QC, Sayari M, Seifert KA, 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 BD, Berger DK, Steenkamp ET, Lim HJ, Duong TA, Bluhm BH, De Beer ZW, De Vos L, Fourie G, Naidoo K, Olivier N, 2017. Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntiella decipiens* and *Ophiostoma ips*. *IMA fungus* **8**: 385-96.

Wingfield BD, Duong TA, Hammerbacher A, van der Nest MA, Wilson A, Chang R, De Beer W, Steenkamp ET, Wilken MP, Naidoo K, Wingfield MJ, 2016b. Draft genome sequences for *Ceratocystis fagacearum*, *C. harringtonii*, *Grosmannia penicillata*, and *Huntiella bhutanensis*. *IMA fungus* **7**: 317-323.

Wolf T, Shelest V, Nath N, Shelest E, 2016. CASSIS and SMIPS: promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes. *Bioinformatics* **32**: 1138-1134.

Yu D, Xu F, Zeng J, Zhan J, 2012. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64**: 285–295.

Zdobnov, EM, Apweiler R, 2001. InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**: 847-848.

Supplementary Material

Supplementary File S1: This file contains information on the PCR-based confirmations of gene content and organization of each PKS cluster.

Supplementary File S2: This file contains information of the top 10 BLAST hits for all PKS genes identified in this study.

Supplementary File S3: This file contains information regarding the annotation of each of the PKS clusters identified in the 19 *Ceratocystidaceae* genomes examined.