G3, 2022, 12(3), jkac008
https://doi.org/10.1093/g3journal/jkac008
Advance Access Publication Date: 13 January 2022
Fungal Genetics and Genomics

Phenolic degradation by catechol dioxygenases is associated with pathogenic fungi with a necrotrophic lifestyle in the Ceratocystidaceae

Nicole C. Soal [1], Martin P. A. Coetzee [1], Magriet A. van der Nest [1,2], Almuth Hammerbacher [3,*] and Brenda D. Wingfield [1]

[1]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria 0028, South Africa
[2]Biotechnology Platform, Agricultural Research Council (ARC), Pretoria 0110, South Africa
[3]Department of Zoology and Entomology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria 0028, South Africa

*Corresponding author: Department of Zoology and Entomology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria 0028, South Africa. Email: almuth.hammerbacher@fabi.up.ac.za

Abstract

Fungal species of the Ceratocystidaceae grow on their host plants using a variety of different lifestyles, from saprophytic to highly pathogenic. Although many genomes of fungi in the Ceratocystidaceae are publicly available, it is not known how the genes that encode catechol dioxygenases (*CDOs*), enzymes involved in the degradation of phenolic plant defense compounds, differ among members of the Ceratocystidaceae. The aim of this study was therefore to identify and characterize the genes encoding CDOs in the genomes of Ceratocystidaceae representatives. We found that genes encoding CDOs are more abundant in pathogenic necrotrophic species of the Ceratocystidaceae and less abundant in saprophytic species. The loss of the *CDO* genes and the associated 3-oxoadipate catabolic pathway appears to have occurred in a lineage-specific manner. Taken together, this study revealed a positive association between *CDO* gene copy number and fungal lifestyle in Ceratocystidaceae representatives.

Keywords: catechol dioxygenase; Ceratocystidaceae; gene loss; catechol degradation

Introduction

The interaction between plants and pathogenic fungi is complex, with the continuous coevolution of plant defense mechanisms and virulence mechanisms of pathogens (Anderson *et al.* 2010). Upon invasion by a pathogen, plants defend themselves from infection through various mechanisms, such as the formation of physical barriers, synthesis of antimicrobial proteins, and chemical defense compounds (Jones and Dangl 2006; Ferreira *et al.* 2007). Plants synthesize a variety of chemical compounds (terpenes, alkaloids, and different types of phenolics) that are either preformed or produced upon infection (Lattanzio *et al.* 2006, 2008). Some of these phenolic compounds include known antifungal agents, such as catechin, stilbenes, isoflavonoids, and condensed tannins (Jeandet *et al.* 2002; Kocaçalişkan *et al.* 2006; Lattanzio *et al.* 2006; Liu *et al.* 2017; Ullah *et al.* 2017).

Phytopathogenic fungi have evolved to overcome host defense responses. For example, some fungi can degrade phenolics that are part of the host defense response via specialized enzymatic pathways (Westrick *et al.* 2021). In many instances, the phenolic degradation pathways in fungi are clustered and connected by key classes of enzymes (Gluck-Thaler *et al.* 2018; Gluck-Thaler and Slot 2018). This includes the enzyme class catechol dioxygenases (CDO), which can cleave aromatic rings at the catecholic bond through the addition of molecular oxygen (Hayaishi 1966; Broderick 1999). The CDO enzymes are divided into 2 families either catalyzing the first reaction of the *ortho*-cleavage or *meta*-cleavage pathways. These enzyme families have evolved independently from one another and are found in both eukaryotic and prokaryotic microorganisms (Harayama and Rekik 1989; Eltis and Bolin 1996; Vaillancourt *et al.* 2006).

The intradiol dioxygenases (EC 1.13.11.1), are part of the catechol branch of the 3-oxoadiptae pathway (also known as the *ortho*-cleavage pathway or the β-ketoadipate pathway) (Harwood and Parales 1996; Martins *et al.* 2015). Intradiol dioxygenases cleave the carbon–carbon bond of catecholic aromatic compounds between the 2 adjacent hydroxy substituents (Fig. 1) (Broderick 1999; Vaillancourt *et al.* 2006). In this pathway, the catecholic substrate is cleaved to form a muconolactone (Harwood and Parales 1996), which is further broken down via multiple enzymatic reactions to form the products succinyl-CoA and acetyl-CoA which enter the tricarboxylic acid cycle and provide substrates for the production of the high-energy molecules, NADH and ATP (Fig. 1) (Ornston and Stanier 1966; Harwood and Parales 1996; Martins *et al.* 2015). In the *meta*-cleavage pathway, on the other hand, an extradiol dioxygenase cleaves the catecholic substrate adjacent to a hydroxy substituent to form a muconate semialdehyde (Vaillancourt *et al.* 2006). This compound is further broken down

Received: November 03, 2021. **Accepted:** January 07, 2022

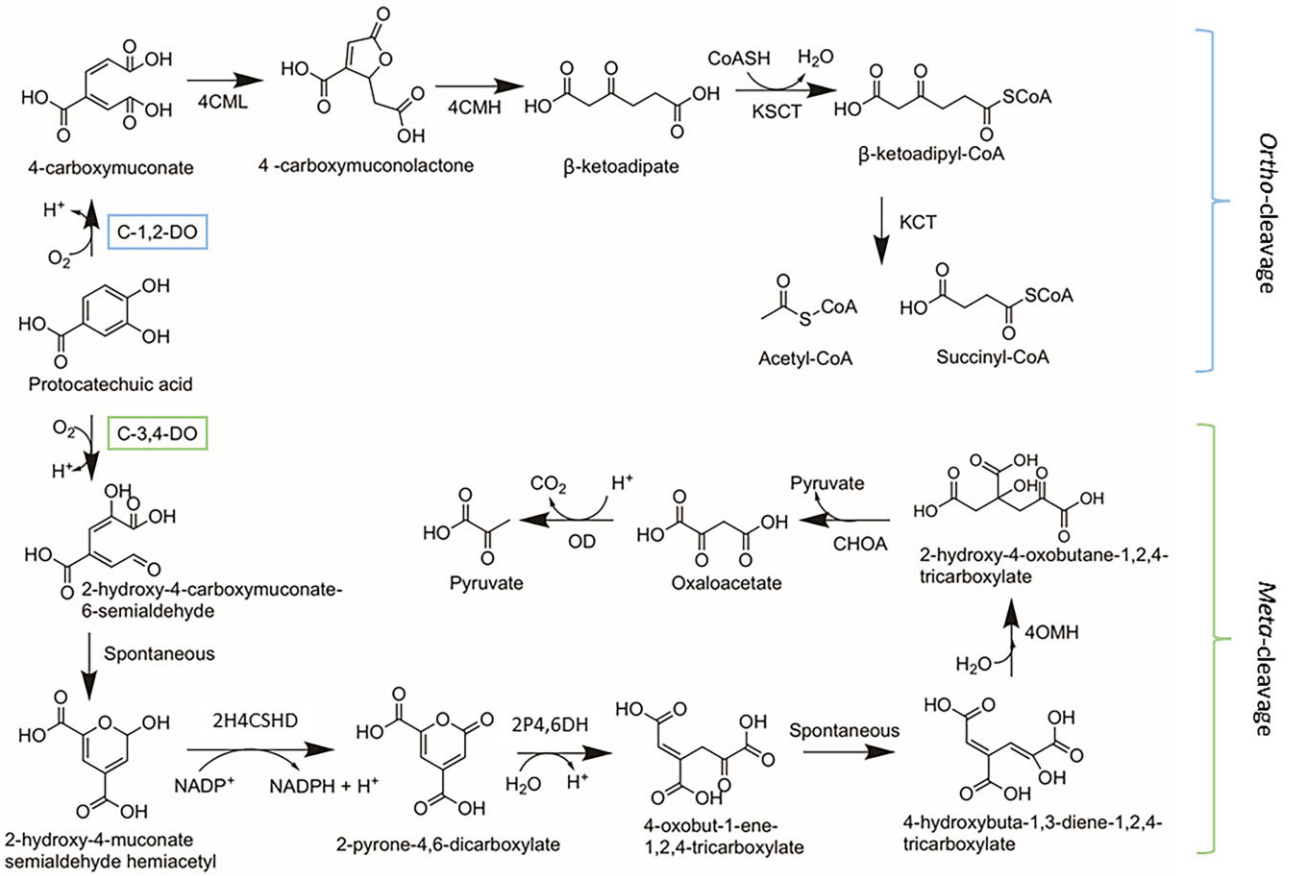

Fig. 1. The *ortho*- and *meta*-cleavage routes of phenolic degradation by microbial species. Blue bracket over C-1,2-DO represents the step in the *ortho*-cleavage pathway that CDO1, 2, and 4 of the Ceratocystidaceae are involved in. Green bracket over C-3,4-DO indicates the step in the *meta*-cleavage pathway -in which CDO3 of the Ceratocystidaceae is involved in. (Enzymes: C-1,2-DO—catechol 1,2-dioxygenase, 4CML—4-carboxymuconate lactonase, 4CMH—4-carboxymuconolactone hydrolase, KSCT—β-ketoadipate:succinyl-CoA transferase, KCT—β-ketoadipate-CoA thiolase; C-3,4-DO—catechol 3,4-dioxygenase, 2H4CSHD—2-hydroxy-4-carboxymuconate-6-semialdehyde dehydrogenase, 2P4,6DH—2-pyrone-4,6-dicarboxylate hydrolase, 4OMH—4-oxalomesaconate hydratase, CHOA—4-carboxy-4-hydroxy-2-oxoadipate aldolase, OD—oxaloacetate β-decarboxylase). Chemical structures and pathways, adapted from Wadke et al. (2016).

to form pyruvate and acetaldehyde which are metabolized by the glycolysis pathway (Fig. 1) (Maruyama *et al.* 2004).

Fungi are thus able to utilize phenolic plant defense metabolites as carbon sources through the use of CDOs and their respective catabolic pathways (Camarero *et al.* 1994; Bugg *et al.* 2011). The action of CDOs during plant infection has been shown to contribute to the pathogenicity and virulence of some phytopathogenic fungi (Shanmugam *et al.* 2010; Hammerbacher *et al.* 2013; Wadke *et al.* 2016). For example, the fungal maize pathogen, *Cochliobolus heterostrophus*, produces an intradiol CDO in response to phenolic compounds present in its environment (Shanmugam *et al.* 2010). Members of the family Ceratocystidaceae utilize CDOs to overcome host responses. For example, the pathogen *Endoconidiophora polonica* makes use of CDOs in the degradation of flavonoids and stilbenes produced by its conifer host (Wadke *et al.* 2016). Previous studies revealed that increased expression of the *CDO* genes in this fungus increased virulence of the pathogen (Hammerbacher *et al.* 2013; Wadke *et al.* 2016).

In spite of the importance of CDOs in the plant host-fungal pathogen interaction, very little is known about CDOs in the Ceratocystidaceae. This fungal family includes agricultural crop pathogens such as the sweet potato black rot pathogen, *Ceratocystis fimbriata*, the banana crown rot pathogen, *Thielaviopsis musarum*, and the causal agent of carrot root rot, *Berkeleyomyces basicola* (Halsted and Fairchild 1891; Melo *et al.* 2016; Nel *et al.* 2018). There are also many species which cause disease on important forest trees, including the conifer blue-stain fungi, *E. polonica* and *Endoconidiophora laricicola*, the wattle wilt pathogen, *Ce. albifundus*, and the causal agent of sapstreak in sugar maple trees, *Davidsoniella virescens* (Krokene and Solheim 1998; Roux *et al.* 2007; Richter 2012). Apart from these pathogenic species there are also a number of saprophytic species in the Ceratocystidaceae that do not cause disease or which are considered weakly pathogenic. These include *Huntiella* species, particularly *Huntiella moniliformis* and *Huntiella bhutanensis* (Wilson *et al.* 2015; Wingfield *et al.* 2016b), weakly pathogenic fungi such as *D. eucalypti*, found on *Eucalyptus* (Kile *et al.* 1996), and the ambrosia beetle symbionts, *Ambrosiella xylebori* and *Ambrosiella cleistominuta* (von Arx and Hennebert 1965; Mayers *et al.* 2020).

There are a large number of sequenced genomes in the Ceratocystidaceae (Wilken *et al.* 2013; Van der Nest *et al.* 2014; Wingfield *et al.* 2015, 2016a, 2017) which provide an important resource for determining the presence of these enzymes across different genera. However, to date, only 4 CDOs were identified in *E. polonica* (Wadke *et al.* 2016). It is still unknown whether other members of the Ceratocystidaceae with different ecological lifestyles have similar enzymes. The aim of this study was to identify

and characterize the *CDOs* in pathogenic and nonpathogenic species of the Ceratocystidaceae, as well as understand the evolutionary history of the catechol dioxygenases and their involvement in fungal nutrition.

Materials and methods

Identification and characterization of putative CDOs in the Ceratocystidaceae

To identify putative CDO homologs present in the genomes of the fungal species, we used the sequences identified in *E. polonica* to search the publicly available genomes. These included the nucleotide sequences of 3 intradiol dioxygenases (*EpCDO1*, GenBank Accession: KU221039; *EpCDO2*, GenBank Accession: KU221040; *EpCDO4*, GenBank Accession: KU221042) and 1 extradiol dioxygenase (*EpCDO3*, GenBank Accession: KU221041) (Wadke *et al.* 2016).

Putative CDO homologs were identified using local BLAST searches (tBLASTn, expect [E]-values > 10^5) using CLC Genomics Workbench v 8.1 (https://www.qiagenbioinformatics.com/). The corresponding scaffold or contig on which there was a hit was extracted in CLC Main Workbench and open reading frames were predicted using Web AUGUSTUS (http://bioinf.uni-greifswald.de/augustus/) (Hoff *et al.* 2012; Hoff and Stanke 2013). The genes identified from the annotations were confirmed with a reciprocal BLASTp on NCBI. The functional domains of the CDOs were annotated using InterProScan (Jones *et al.* 2014), the conserved domains were identified using the conserved domain (CD) function on NCBI (http://www.ncbi.nlm.-nih.gov/Structure/cdd/wrpsb.cgi). The predicted proteins were also assessed for transmembrane domains using TMHMM (Krogh *et al.* 2001) and for signal peptides using SignalP (Petersen *et al.* 2011). Orthologous proteins were identified using OrthoMCL (Li *et al.* 2003).

Phylogenetic analysis

A phylogenetic approach was taken to determine if the genes could be considered orthologous and to determine the evolutionary history and relationships of the respective genes. The genome sequence information for 30 species of the Ceratocystidaceae was used, in addition the genome sequences from *Fusarium circinatum* and *Fusarium fujikuroi* were used to provide outgroup gene sequences (Table 1). They were obtained from the GenBank database of the National Centre for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/).

The amino acid sequences for each of the individual genes (CDO1-4) were aligned using MAFFT (Multiple sequence alignment based on fast Fourier transform) v 7 (https://mafft.cbrc.jp/alignment/server/) using the L-INS-i option (Katoh and Toh 2008) to generate a single alignment per gene. The alignments were trimmed and in the case of CDO4, the N-terminal signal peptide was excluded. The alignments were then analyzed using a Maximum Likelihood (ML) method with RAxML v8 (Stamatakis 2014). The predicted best-fit substitution model for amino acid data as indicated by ProtTest 3 (Darriba *et al.* 2011) was incorporated in the ML analyses. The evolutionary models were the Whelan and Goldman (2001) with gamma distribution (WAG+G) for CDO1 and CDO2 data, and the Le-Gascuel model (Le and Gascuel 2008) with gamma distribution (LG+G) for CDO3 and CDO4. The branch support was determined using 1000 bootstrap replicates. The phylogenetic trees were viewed using MEGA v.7 (Kumar *et al.* 2018) and FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

A multi locus species tree was generated for all the species used in this study. The nucleotide sequences for the Minichromosome Maintenance Complex Component 7 (MCM7), DNA-directed RNA polymerase II subunit RPB1 and RPB2 (RPB1, RPB2), and elongation factor 2 (EF2) genes were identified in the genomes using the BLASTn function in CLC Main Workbench v 8.1. The nucleotide sequences were aligned using MAFFT v 7 and concatenated using FASconCAT-G v 1.04 (Kück and Longo 2014). The alignments were converted to Phylip files using PAUP v 4.0a (Swofford 2002). An ML analysis was conducted on the data using the predicted evolutionary model of a General Time Reversible with gamma distribution (GTR+G) as determined with jModelTest (Darriba *et al.* 2012). RAxML v8 was used to perform the ML analysis, with the data being partitioned into the respective loci, and the parameters for each partition allowed to vary independently. Bootstrap support was determined with 1000 rapid bootstrap replicates. The tree was viewed using MEGA v.7 and FigTree v1.3.1.

Gene synteny flanking *CDOs*

The gene synteny surrounding the *CDO* genes was investigated to determine whether the *CDOs* are found in specialized enzymatic clusters. For each of the *CDOs*, the 5 genes flanking either side of the genes, where available, were identified using BLASTx searches on NCBI. These flanking regions were assessed for synteny using EasyFIG 2.2.3 (Sullivan *et al.* 2011). This was done to identify any possible gene clusters involved in the degradation of catechol derivatives, and to determine whether there were regions in the surrounding genes that were indicative of gene expansions, gains or losses. The programme uses a BLASTn function to compare the sequences of the input fasta files using a similarity threshold with a maximum E-value of 0.001.

Analysis of *CDO* gene family evolution

The gene family contractions and expansions across the Ceratocystidaceae were inferred using the Count software package (http://www.iro.umontreal.ca/~csuros/gene_content/count.html) through inferring the character's history given the phylogeny (Csűrös and Miklós 2006; Csűrös 2010). The analysis was performed using a Wagner parsimony method that penalizes individual family members for their loss (loss penalty = 1) and gains, inferring the history with a minimum penalty on the previously described multilocus species tree generated. To further analyze the data, the ancestral character states were constructed for each individual gene using Mesquite v 3.0.4 (Maddison and Maddison 2019). The analyses were based on Parsimony reconstruction and used the tree generated from the combined data (see above) as the phylogenetic hypothesis.

Identification of the *ortho*- and *meta*-cleavage pathways in the Ceratocystidaceae

To identify the putative homologs of the *ortho*- and *meta*-cleavage pathways, the characterized genes of the pathways from closely related species were used in a tBLASTn search against the 30 genomes of the Ceratocystidaceae (Table 1) using CLC Genomics Workbench v 8.1. The reference genes used as queries include carboxy-cis, cis-muconate cyclase (*Ce. fimbriata* PHH5432.1); 3-oxoadipate enol-lactonase (Ce. platani KKF95260.1); 3-oxoadipate CoA-transferase (*Ce. platani* KKF92536.1); acetyl-CoA acyltransferase (*Ce. fimbriata* PHH52801.1). The corresponding scaffold or contig on which a gene was identified was extracted in CLC Main Workbench and open reading frames were predicted using web AUGUSTUS. The genes identified from the annotations were confirmed with a reciprocal BLASTp on NCBI.

Table 1. Presence of the *CDO* 1-4 gene transcripts across the Ceratocystidaceae with the genome information and ecological lifestyle of each species used in this study.

Species	Genome ID	*CDO1*	*CDO2*	*CDO3*	*CDO4*	Lifestyle
A. cleistominuta	JABFIG000000000.1	—	—	√	√	Beetle-vectored, weakly pathogenic
A. xylebori	PCDO00000000.1	—	—	√	√	Beetle-vectored, weakly pathogenic
Be. basicola	PJAC00000000.1	√	√	√	√	Necrotroph
Br. fagacearum	MKGJ00000000.1	—	√	√	√	Necrotroph
Ca. adiposa	LXGU00000000.1	—	—	√	√	Weakly pathogenic
Ce. albifundus	JSSU000000000.1	√	√	√	√	Necrotroph
Ce. eucalypticola	LJOA00000000.1	√	√	√	√	Necrotroph
Ce. cacaofunesta	PEJQ00000000.1	√	√	√	√	Necrotroph
Ce. fimbriata	APWK00000000	√	√	√	√	Necrotroph
Ce. harringtonii	MKGM00000000.1	√	√	√	√	Necrotroph
Ce. manginecans	JJRZ01000000	√	√	√	√	Necrotroph
Ce. platani	LBBL00000000.1	√	√	√	√	Necrotroph
Ce. smalleyi	NETT00000000.1	√	√	√	√	Necrotroph
Ch. thielavioides	BCGU00000000.1	—	√	√	√	Necrotroph
Davidsoniella australis	RHLR00000000.1	√	√	√	√	Necrotroph
D. eucalypti	RMBW00000000.1	√	√	√	√	Weakly pathogenic
D. neocaledoniae	RHDR00000000.1	√	√	√	√	Necrotroph
D. virescens	LJZU00000000.1	√	√	√	√	Necrotroph
E. laricicola	LXGT00000000.1	√	√	√	√√	Necrotroph
E. polonica	LXKZ00000000.1	√	√	√	√√	Necrotroph
H. bhutanensis	MJMS00000000.1	—	—	√	—	Saprotroph
H. decipiens	NETU00000000.1	—	—	√	—	Saprotroph
H. moniliformis	JMSH00000000.1	—	—	√	—	Saprotroph
H. omanensis	JSUI00000000.1	—	—	√	—	Saprotroph
H. savannae	LCZG00000000.1	—	—	√	—	Saprotroph
Thielaviopsis cerberus	JACYXV00000000.1	√	√	√	√	Necrotroph
T. ethacetica	BCFY00000000.1	√	√	√	√	Necrotroph
T. euricoi	BCHJ00000000.1	√	√	√	√	Necrotroph
T. musarum	LKBB00000000.1	√	√	√	√	Necrotroph
T. punctulata	LAEV00000000.1	√	√	√	√	Necrotroph
F. circinatum	AYJV02000000	√	√	√	√	Necrotroph
F. fujikuroi	GCA_900079805.1	√	√	√	√	Necrotroph

Statistical analysis

The data were separated into pathogens and nonpathogens (as seen in Table 1), and the number of species in each category were further assigned to a category of (1) having all 4 *CDOs* present or (2) not having all 4 *CDOs* present. This count data were subjected to a Fisher's Exact Test using a 95% confidence interval (2-tailed), with an alternate hypothesis that the true odds ratio is not equal to 1. All analyses were conducted using R (www.r-project.org).

Results

Identification and characterization of putative *CDOs* in the Ceratocystidaceae

From the 32 genomes used in this study (Table 1), putative CDOs (*CDO1*—*CDO4*) were identified based on their sequence identity to previously identified genes from *E. polonica* (Wadke *et al.* 2016). The genes identified were consistent between species within the respective genera (Table 1). Only one of the genes, *CDO3*, was identified in all the species included in this study (Table 1). *Ceratocystis* spp., *Davidsoniella* spp., 1 *Berkeleyomyces* sp., and *Thielaviopsis* spp. had single copies of all 4 of the previously described *CDO* genes. *Bretziella fagacearum* and *Chalaropsis thielavioides* genomes contained gene sequences for *CDO2*, 3, and 4. *Ambrosiella* spp. and *Catunica adiposa* contained gene sequences for *CDO3* and *CDO4*. Species of *Huntiella* had only 1 extradiol dioxygenase, *CDO3*. We also identified a gene duplication of *CDO4* in *E. polonica* and *E. laricicola*. The genomes of the outgroup species *F. circinatum* and *F. fujikuroi* contained 1 copy of each of the CDOs (Table 1).

The genome architecture of *CDO1* using InterProScan, CD and SignalP identified a conserved intradiol ring-cleavage dioxygenase core domain (IPR015889) adjacent to a conserved intradiol ring-cleavage dioxygenase C-terminal (IPR000627). The sequences contained the conserved residues of the active site and dimer interface, including the residues involved in binding the iron(III) cofactor.

The predicted conserved domains identified in the *CDO2* genes included an N-terminal domain (IPR007535) characteristic of an intradiol cleaving dioxygenase and an intradiol ring-cleavage dioxygenase core domain (IPR015889). A third intradiol dioxygenase was identified, *CDO4*, was identified. All the *CDO4* homologs contained an N-terminal signal peptide as predicted by InterProScan and SignalP, which was consistent with findings by Wadke *et al.* (2016). They were well conserved and spanned 20 amino acids into the gene. Thus, the CDO4 protein is predicted to be secreted to the extracellular space. The analyses identified the ASA-HP cleavage site between amino acids 20 and 21. This is a standard secretory peptide transported by the Sec translocon protein and cleaved by Signal Peptidase I (Auclair *et al.* 2012). This was consistent in each *CDO4* gene included in this study. The intradiol ring-cleaving dioxygenase core C-terminal domain (IPR000627) was present in each of the *CDO4* transcripts.

One extradiol cleaving dioxygenase, *CDO3*, was identified. This extradiol ring-cleavage dioxygenase was similar to a class III enzyme, with a conserved subunit B domain (IPR004183), with similarity to a 4,5-DOPA extradiol dioxygenase (EC 1.13.11.29). A gene duplication of *CDO4* was identified in the *E. polonica* and *E. laricicola* genomes. The genes were located on different scaffolds, and were surrounded by different flanking genes. The sequence

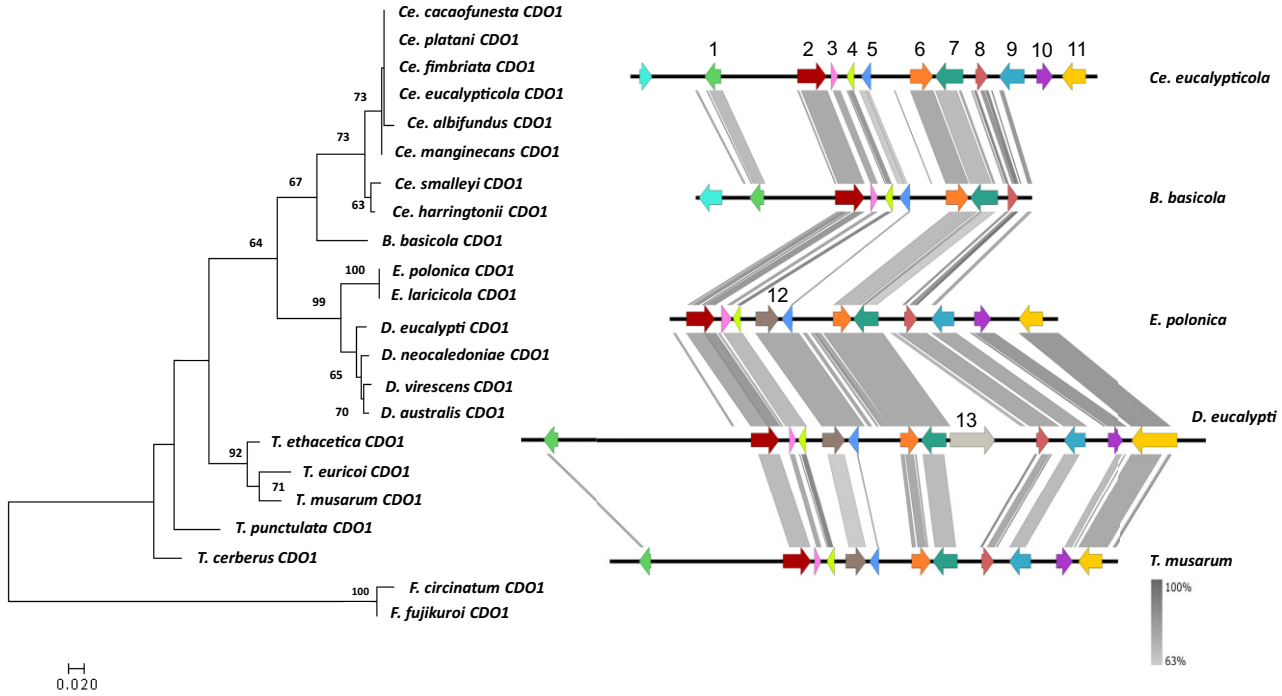

Fig. 2. Maximum likelihood gene tree (left) for CDO1 amino acid sequences (1000 bootstrap replicates) with outgroup species *F. circinatum* and *F. fujikuroi*. Synteny map (right) for representative species of each of the genera present in the gene tree representing the sequence similarity of the genomic regions flanking *CDO1*. **CDO1* (6, orange) including 5 genes upstream and downstream of the *CDO*. Genes which are numbered and colored are shown to be syntenic and conserved through the majority of the species, with some unique genes also included. The genes without numbers or colored gray were not found to be syntenic across genera. From the left: 1—hypothetical protein; 2—putative phosphatidylinositol N-acetylglucosaminyltransferase subunit C; 3—mediator of RNA polymerase II transcription subunit 31; 4—DNA-directed RNA polymerase III subunit RPC10; 5—hypothetical protein containing CFEM domain; 6—*CDO1*; 7—tRNA (uracil-O(2)-)-methyltransferase; 8—eukaryotic translation initiation factor eIF-1; 9—N-glycosylation protein; 10—hypothetical protein; 11—sodium/pantothenate symporter; 12—Ser/Thr protein phosphatase family protein; 13—reverse transcriptase.

similarity of the coding regions of the 2 genes was 92.73% in *E. polonica* and 93.9% in *E. laricicola* (Supplementary Fig. 1). When the amino acid sequences of the homologous *CDO4* genes at the same genomic location were compared to one another, they were 99.13% similar in percentage identity to each other (Supplementary Fig. 1). The duplicate gene is located on scaffolds that are 24,364 and 24,420 bp in length in *E. polonica* and *E. laricicola*, respectively (Wingfield *et al.* 2016a). Each scaffold had 3 genes with a predicted ORF present. The other genes present on the scaffold are predicted to be an isoamyl alcohol oxidase/dehydrogenase and a hypothetical protein. The cleavage site for the secretory signal was identified in the *Endoconidiophora CDO4* gene duplicates.

Phylogenetic analysis

The ML phylogenies of the individual genes showed close congruency to the species tree, and each gene formed a clade that included members of the same genus (Figs. 2–5). The duplicated *CDO4* genes identified in the *Endoconidiophora* species grouped with the *CDO4* genes identified previously (Fig. 5), but formed their own clade, suggesting that the gene expansion occurred in an ancestral *Endoconidiophora* lineage.

Gene synteny flanking *CDO*s

The genes flanking the *CDO*s differed between genera. Within-genus comparisons revealed high conservation of gene order and orientation. When compared across genera, there were certain genera that had conserved gene order and orientation, such as *Ceratocystis* and *Davidsoniella* (Figs. 2–5). However, there were some outliers, such as *Br. fagacearum* and the genus *Huntiella*, showing low synteny compared to the other genera included in the dataset (Figs. 2–5).

The gene orientation and order for *CDO1* showed gene synteny between the *Ceratocystis* species and the *Berkeleyomyces* species (Fig. 2). There are many similarities between the genera, with only a few genes that differed between them. A similar pattern is observed for *CDO2*, with genes downstream of the *CDO2* gene showing high synteny between the genera, whereas, upstream of the gene, the gene order is less conserved in *Thielaviopsis* (Fig. 3). One of the genes upstream of *CDO2* included a carboxy-*cis*, *cis*-muconate cyclase (see gene 5 in Fig. 3) which is one of the enzymes used in the catechol branch of the *ortho*-cleavage pathway. The genes upstream of *CDO3* were more conserved than those downstream in some of the species. There were striking differences between the flanking genes of *CDO3* of *H. moniliformis* compared to the rest of the species in the other genera (Fig. 4).

In many genera, the genes flanking *CDO4* are predicted to be involved in the shikimate pathway and the catabolism of quinate, a product of the shikimate pathway (Fig. 5). Quinate is degraded by various enzymes to a form that can enter the β-ketoadipate pathway. The genes identified as flanking *CDO4* include a catabolic 3-dehydroquinase (EC 1.1.5.8), 3-dehydroshikimate dehydratase (EC 4.2.1.10), quinate dehydrogenase (EC 1.1.1.282), and a quinate permease (Pf00083) which collectively are involved in the transformation of quinic acid to protocatechuate. This gene cluster was not present in *Br. fagacearum* and the *Ambrosiella* species (Fig. 5). There were also large differences in the gene orientation and order between the different species for

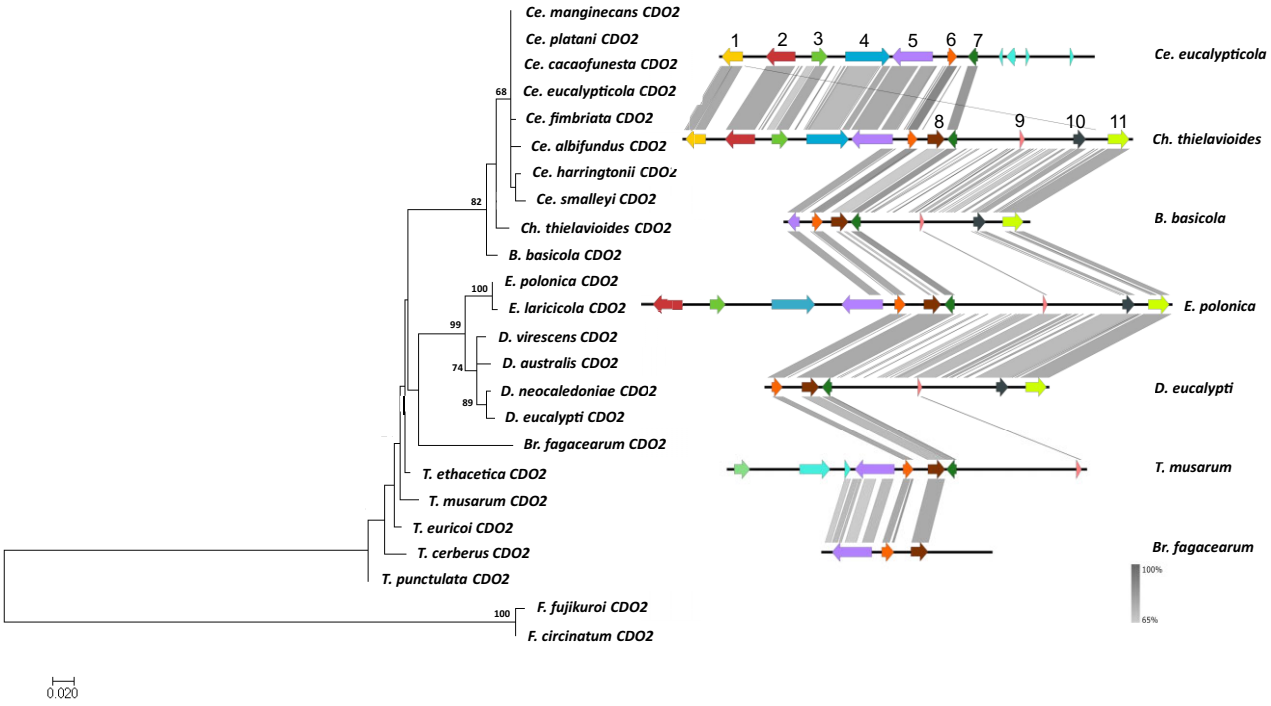

Fig. 3. Maximum likelihood gene tree (left) for CDO2 amino acid sequences (1000 bootstrap replicates) with outgroup species *F. circinatum* and *F. fujikuroi*. Synteny map (right) for representative species of each of the genera present in the gene tree representing the sequence similarity of the genomic regions flanking *CDO2*. **CDO2* (6, orange) including 5 genes upstream and downstream of the *CDO*. Genes which are numbered and colored are shown to be syntenic and conserved through the majority of the species, with some unique genes also included. The genes without numbers or colored gray were not found to be syntenic across genera. From the left: 1—neutral ceramidase; 2—ATP-dependent DNA helicase srs2; 3—hypothetical protein containing DUF 4591 domain; 4—sterol regulatory element-binding protein 1; 5—carboxy-*cis*, *cis*-muconate cyclase; 6—*CDO2*; 7—nitronate monooxygenase; 8—endonuclease/exonuclease/phosphatase, COG 2374; 9—double-stranded RNA binding motif; 10–37S ribosomal protein S7 mitochondrial; 11—hypothetical protein.

the genes flanking *CDO4* (Fig. 5), as the syntenic genes found in the *Ambrosiella* species were dispersed in a different order, with some of the genes appearing upstream of *CDO4* on the contig. In *Br. fagacearum* and the genus *Ambrosiella* only 1 syntenic gene was found on the scaffold (Histone-lysine N-methyltransferase SET9), where the rest of the genes lacked synteny to the other genera (Fig. 5).

Analysis of *CDO* gene family evolution

Parsimony analysis of the *CDO*s showed a simple evolutionary history of the genes (Fig. 6). Based on the results of the parsimony analysis, a loss of the *CDO1* gene occurred in the common ancestor of *Bretziella, Ambrosiella*, and *Huntiella*. This gene was also independently lost in the genus *Chalaropsis* when it split from the *Ceratocystis*. *Huntiella* incurred a simultaneous loss of *CDO2* and *CDO4* after it split from its closest relatives, the genera *Ambrosiella* and *Bretziella*. *CDO2* was also lost in *Ambrosiella* after diverging from *Bretziella*. A duplication of *CDO4* occurred in the common ancestor of *E. polonica* and *E. laricicola*, suggesting that they are paralogs which have continued to diverge over time (Fig. 6).

An alternative evolutionary history was determined using a Count analysis, which determines the distribution of homolog family sizes across the different genera using a Wagner parsimony approach (Supplementary Fig. 2). The predicted gains and losses on the tree (Supplementary Fig. 2) showed that many were lineage specific. Consistent with the previous results, the ancestral state of the genes predicted that the most recent common ancestor of the Ceratocystidaceae contained a single copy of each of the CDOs. A single gene loss for *CDO1* was predicted for the *Ch. thielavioides* (Supplementary Fig. 2). A gene loss event of both *CDO1* and *CDO2* was predicted to have occurred in the lineage leading to the *Huntiella*, *Bretziella*, and *Ambrosiella* lineages, which included *Ca. adiposa*. In contrast to the ancestral state reconstruction in Mesquite using parsimony, a gene gain of *CDO2* was predicted to have occurred for *Br. fagacearum*. The gene expansion predicted for the *Endoconidiophora* species was consistent with that of the method used in Mesquite.

When observed as a whole, the gene loss events were found to have occurred in the clades containing the saprophytic fungi and weakly pathogenic fungi, while the pathogenic fungi contained all 4 *CDO* copies (Table 1 and Fig. 6). The Fisher's Exact Test showed an association between *CDO* copy number and lifestyle of the Ceratocystidaceae ($P < 0.01$).

Identification of the *ortho*- and *meta*-cleavage pathways in the Ceratocystidaceae

The genes for the entire *ortho*-cleavage pathway in the Ceratocystidaceae were identified in all of the genera except for those of *Huntiella, Ambrosiella*, *Catunica*, and *Bretziella* (Table 2). There was one exception, the acetyl-CoA acyltransferase gene (EC 2.3.1.16) was present in all Ceratocystidaceae represented genomes (Table 2). The genes involved in the entire *meta*-cleavage pathway could not be identified in any of the fungal species, apart from 1 gene, a putative aldehyde dehydrogenase (EC 1.2.1.85).

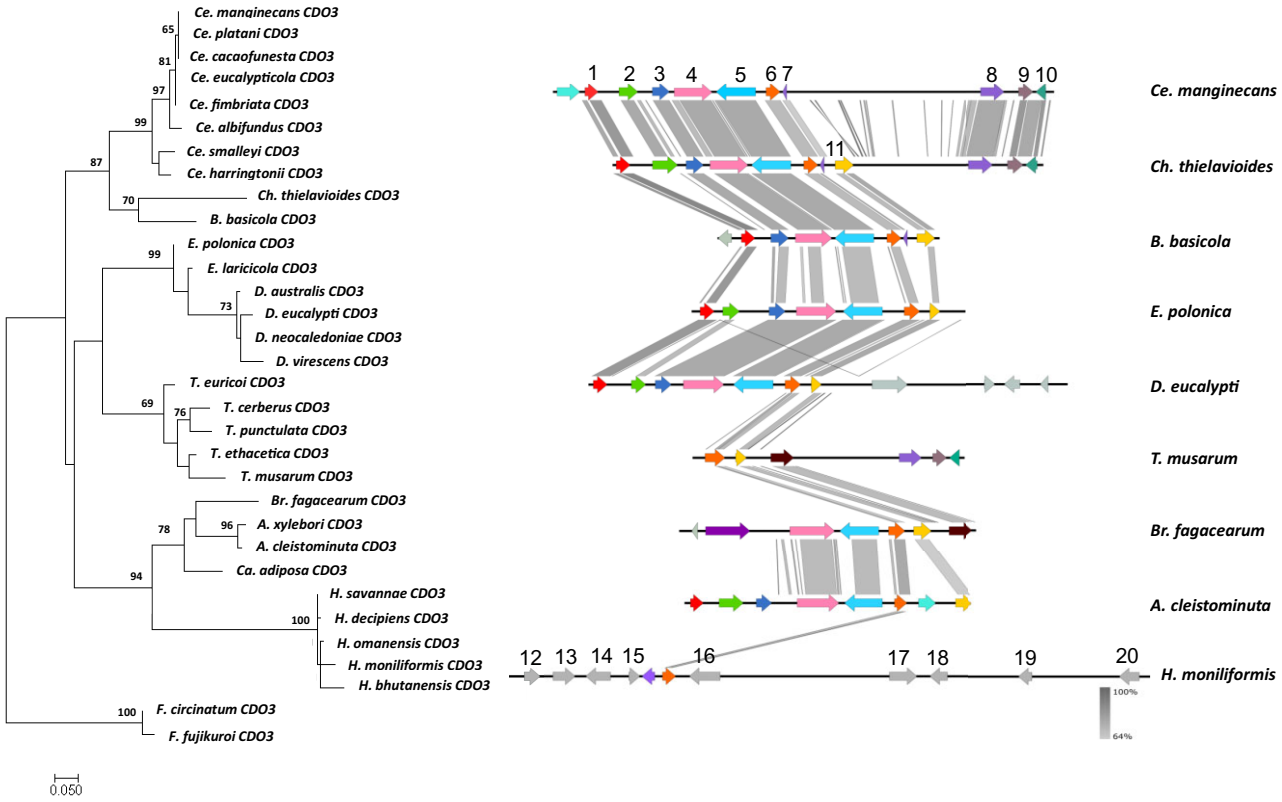

Fig. 4. Maximum likelihood gene tree (left) for CDO3 amino acid sequences (1000 bootstrap replicates) with outgroup species *F. circinatum* and *F. fujikuroi*. Synteny map (right) for representative species of each of the genera present in the gene tree representing the sequence similarity of the genomic regions flanking *CDO3*. **CDO3* (6, orange) including 5 genes upstream and downstream of the *CDO*. Genes which are numbered and colored are shown to be syntenic and conserved through the majority of the species, with some unique genes also included. The genes without numbers or colored gray were not found to be syntenic across genera. From the left: 1–40S ribosomal protein S19; 2—hypothetical protein; 3—fluoride export protein 1; 4—hypothetical protein containing DUFF 4449 domain; 5—DNA ligase 4; 6—*CDO3*; 7—hypothetical protein; 8—COP9 signalosome complex subunit 5; 9—hypothetical protein; 10—putative RNA-binding protein sce3; 11—hypothetical protein; 12—universal stress protein A family protein C25B2.10; 13—GPCR-type G protein 2; 14—putative RNA-binding protein sce3; 15—hypothetical protein; 16—putative vacuolar membrane protein; 17—glycerol kinase; 18—mannose-1-phosphate guanyltransferase; 19—hypothetical protein; 20—protein farnesyltransferase subunit beta.

Discussion

Throughout the fungal kingdom, closely related species often differ significantly in their lifestyles (Soanes *et al.* 2008). This is also observed in the Ceratocystidaceae, a family which includes fungi that are associated with plants in different ways, ranging from saprophytic to highly pathogenic. Different lifestyles in closely related fungi are often determined by different gene inventories and not by evolutionary distances (Schäfer *et al.* 1989; Yoder and Turgeon 2001).

Plants synthesize a wide range of phenolic compounds, usually in high concentrations (Levin 1971). These compounds have antifungal activity, through binding to and precipitating extracellular proteins (Charlton *et al.* 2002; Ndhlala *et al.* 2015), interfering with membrane potentials (Rao *et al.* 2010), disrupting membrane structure (Sang Sung and Gun Lee 2010) and the oxidative respiratory electron chain (Shalaby and Horwitz 2015). In order to survive in association with plants, as epiphytes, pathogens or endophytes, fungi have adapted to these compounds, and intricate mechanisms to circumvent their toxicity have evolved (Anderson *et al.* 2010; Frantzeskakis *et al.* 2020).

An adaptation to phenolic plant defenses are catabolic degradation pathways, where the phenolic ring is cleaved by a catechol dioxygenase enzyme and the linearized carbon chain is subsequently used as an energy source (Broderick 1999). In previous studies, CDO enzymes were identified as pathogenicity and virulence factors in the bark-beetle associated blue-stain fungus, *E. polonica*, a member of the Ceratocystidaceae (Hammerbacher *et al.* 2013; Wadke *et al.* 2016). To expand on previous research, CDO genes from different genera of the Ceratocystidaceae with widely different ecological lifestyles were investigated in this study with a focus on genomic differences using a comparative approach.

CDO genes in the Ceratocystidaceae are more abundant in pathogenic species

In previous studies, 4 CDOs were described in *E. polonica*, (CDO1-CDO4) (Wadke *et al.* 2016). In this study, all 4 homologous CDOs were identified in the genera *Ceratocystis*, *Davidsoniella*, *Berkeleyomyces*, *Endoconidiophora*, and *Thielaviopsis*. Species in these genera are considered highly pathogenic on diverse hosts. For example, *Ce. fimbriata* causes black rot of sweet potatoes, *Ce. manginecans* and *Ce. albifundus* cause wilt in Acacia trees and *Ce. platani* causes canker stains in plane trees (Halsted and Fairchild 1891; Johnson *et al.* 2005; Ocasio-morales *et al.* 2007; Roux *et al.* 2007; Tarigan *et al.* 2011; Adawi *et al.* 2013). *Thielaviopsis* species cause disease on many species of palm trees around the world (Paulin-

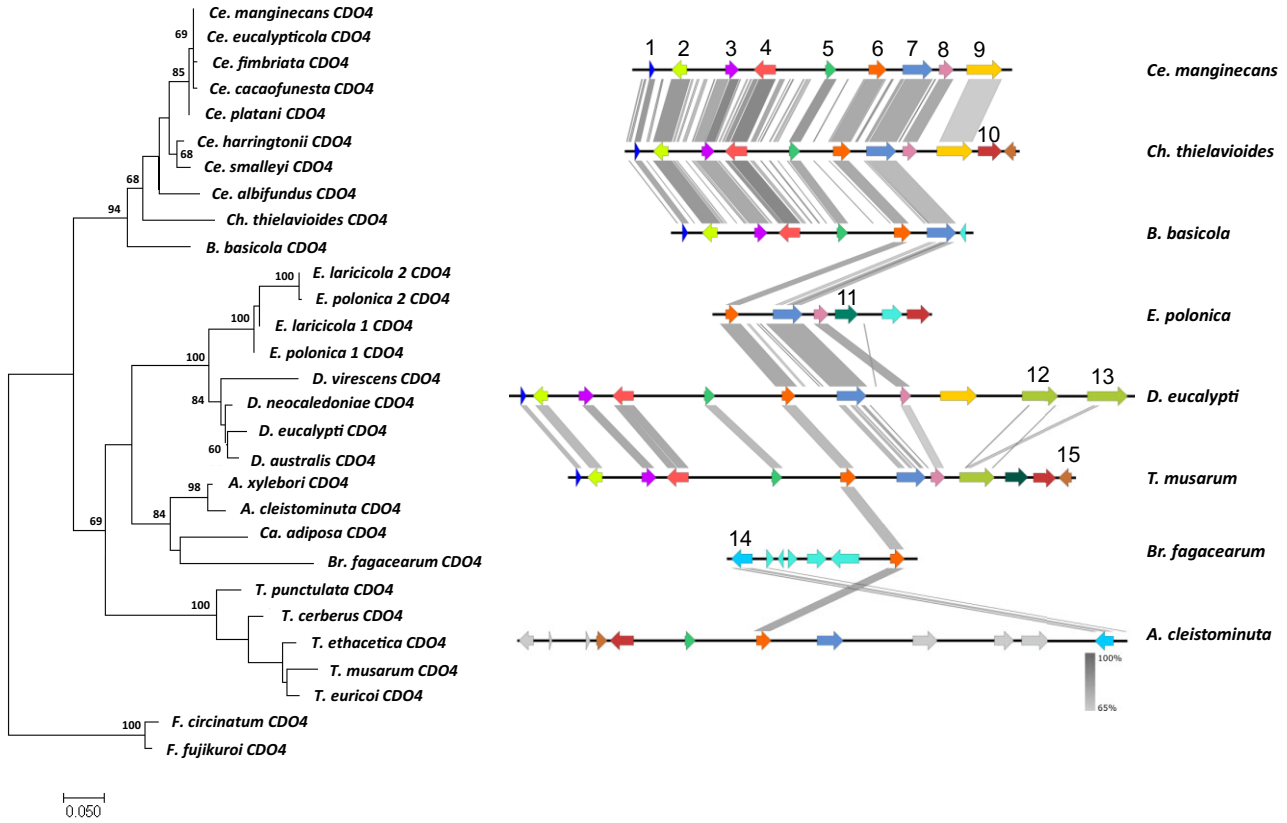

Fig. 5. Maximum likelihood gene tree (left) for CDO4 amino acid sequences (1000 bootstrap replicates) with outgroup species *F. circinatum* and *F. fujikuroi*. Synteny map (right) for representative species of each of the genera present in the gene tree representing the sequence similarity of the genomic regions flanking *CDO4*. **CDO4* (6, orange) including 5 genes upstream and downstream of the *CDO*. Genes which are numbered and colored are shown to be syntenic and conserved through the majority of the species, with some unique genes also included. The genes without numbers or colored gray were not found to be syntenic across genera. From the left: 1—catabolic 3-dehydroquinase; 2–3-dehydroshikimate dehydratase; 3—quinate dehydrogenase; 4—quinate permease; 5—hypothetical protein; 6—*CDO4*; 7—Hypothetical protein part of fungal TF MHR superfamily, containing a GAL4 zinc finger DNA-binding domain; 8—phosphopantetheine adenylyltransferase; 9—subtilase family protein; 10—DNA polymerase kappa; 11—hypothetical protein part of beta elim lyase superfamily; 12—subtilase/peptidase family; 13—subtilase/peptidase family; 14—histone-lysine N-methyltransferase SET9; 15—FK506-binding protein 2.

Mahady *et al.* 2002; Melo *et al.* 2016; Saeed *et al.* 2016) and *Be. basicola* causes root rot in many hosts such as cotton, chicory, carrots, groundnuts and tobacco plants (Punja and Sun 1999; Nel *et al.* 2018). *Endoconidiophora* species are vectored by bark beetles to coniferous hosts such as *Picea abies* and *Larix decidua* causing blue-stain of the bark and wood, and kill the trees under high inoculation densities (Krokene and Solheim 1997, 1998). The genus *Davidsoniella* mostly contains pathogenic species. *Davidsoniella virescens* is a virulent pathogen, causing sap streak in maple trees, but is also weakly pathogenic on other hardwood hosts (Shigo 1962; Richter 2012; Bal *et al.* 2013). In contrast, *D. eucalypti* is weakly pathogenic to certain species of *Eucalyptus* (Kile *et al.* 1996; Richter 2012; Wingfield *et al.* 2018), however, as it was only isolated from artificial stem wounds its primary host is not known.

In contrast to the pathogenic genera included in this study, only 1 CDO was identified in our study from the genus *Huntiella*. Species of this genus are saprophytic and colonize fresh wounds in tree bark, without causing disease (Roux *et al.* 2004; Van Wyk *et al.* 2006). This dramatic gene loss in the genus *Huntiella* supports previous findings, that the degradation of phenolic compounds is more common in pathogenic fungi compared to their saprophytic counterparts (Gluck-Thaler and Slot 2018). This pattern of gene loss in the saprophytic species of the Ceratocystidaceae has also been observed in the genes of the glycoside hydrolase protein family, which are involved in the catabolism of sucrose (Van der Nest *et al.* 2015).

Ambrosiella species, which are the mycangial symbionts of ambrosia beetles (von Arx and Hennebert 1965; Mayers *et al.* 2015, 2020) only retained 2 of the 4 *CDOs* found in the Ceratocystidaceae. As ambrosia beetle symbionts, *Ambrosiella* species are also only weakly pathogenic, but are required to assimilate sufficient nutrients from the host to sustain the beetles' development (Hulcr and Dunn 2011; Jankowiak 2011). *CDO1* and *CDO2*, therefore, do not seem to play an important role in nutrient acquisition for these fungi. Furthermore, the genome of *Ca. adiposa*, formerly *Ceratocystis adiposa* (Mayers *et al.* 2020), also contained only 2 *CDOs* (*CDO3* and *CDO4*) and is an opportunistic pathogen, which causes root rot in sugarcane cuttings but is considered a weak pathogen and does not attack established plants (Sartoris 1927). Furthermore, *CDO1* was lost from the genome of *Ch. thielavioides*. This species has a similar lifestyle to *Ca. adiposa* and is an opportunistic pathogen causing black mold in weakened plants, such as rose grafts (Longree 1940).

CDO1 was also lost from the genome of *Br. fagacearum*, a highly virulent pathogen, causing red oak wilt (Juzwik *et al.* 2008; de Beer *et al.* 2017). However, this fungus, which enters the tree through wounds and root grafts, restricts itself to the xylem vessels of the outer layers of the sapwood and avoids contact with phenolic-

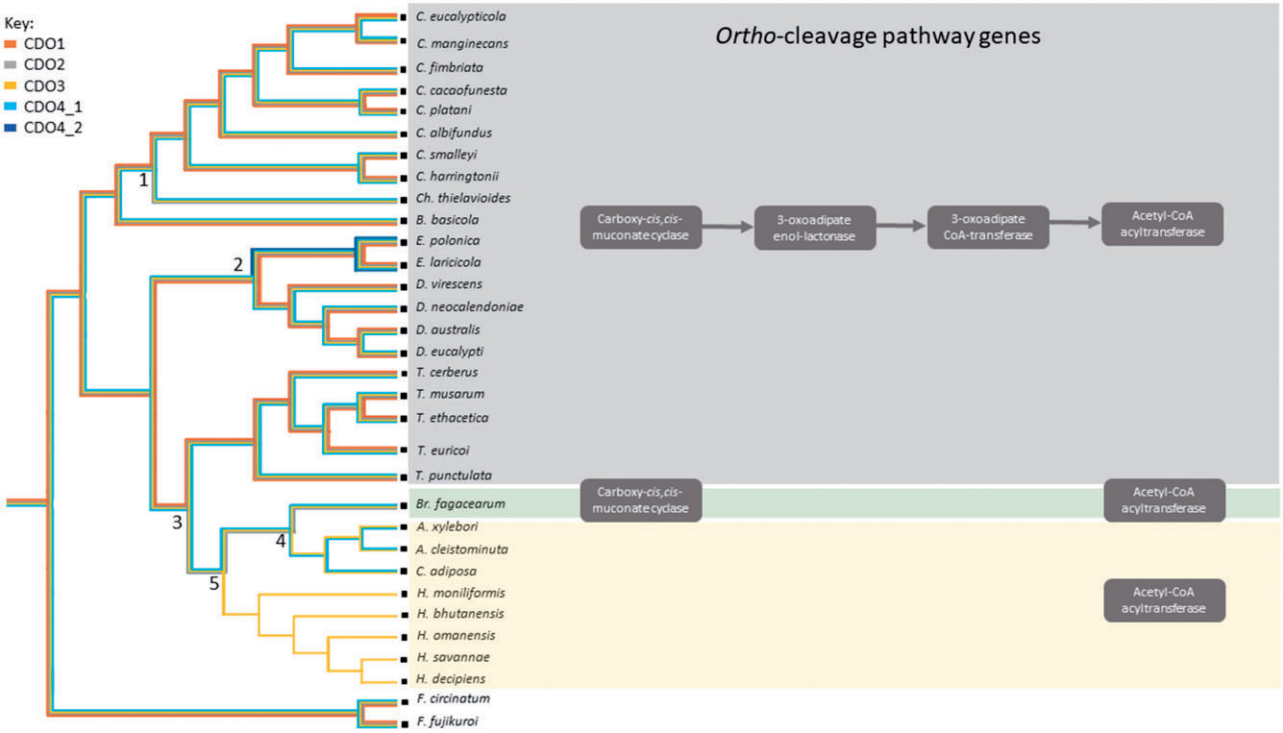

Fig. 6. Ancestral history of the *CDO* 1-4 genes (left image) based on a multi-locus species tree. The presence of a gene is indicated by the colored line (*CDO1*—orange line, *CDO2*—gray line, *CDO3*—yellow line, *CDO4*—light blue line, *CDO4* duplication—dark blue line), and the loss of a gene is indicated by the absence of that colored line from the node. Numbers on graph represent gene loss/gain events (1—loss of *CDO1*; 2—gain of *CD04*; 3—loss of *CDO1*; 4—loss of *CDO2*; 5—loss of *CDO4*). Graphical representation (right image) of *ortho*-cleavage pathway genes present in each species.

producing parenchyma cells (Juzwik and French 1983; Appel *et al.* 1987). Therefore, CDO copy number and type in the Ceratocystidaceae is not only determined by pathogenicity, but also by the pathogen's mode of infection, its preferred niche within the host and the resilience of the host tissue.

Although there does appear to be a trend in *CDO* copy number and ecological lifestyle, it is clear that the gene loss is largely restricted to a single clade with fewer pathogens (Fig. 6). As such the variation of the distribution of *CDOs* observed could be due to phylogenetic relatedness. To further investigate this, future studies will need to include more species, providing multiple independent pathogen and nonpathogen clades that could be compared for *CDO* copy number distribution.

The loss of the *ortho*-cleavage pathway in the Ceratocystidaceae

Another contributing factor to the differences in mode of infection among the Ceratocystidaceae representatives may be attributed to the complete absence of the *ortho*-cleavage pathway in the genera that have lost some of their intradiol CDOs. The species which have lost this pathway are all grouped in one clade suggesting that the loss of this pathway is in some way related to their evolution over time. The clade includes the species of *Huntiella*, *Bretziella*, *Ambrosiella*, and *Catunica*.

The absence of the *ortho*-cleavage pathway in *Huntiella*, *Catunica*, *Ambrosiella*, and *Bretziella* will require further research to understand if the gene loss in these genera provides a fitness disadvantage for these fungi and limits them to their specific ecological niches. The *ortho*-cleavage pathway is predominantly used by fungi to degrade phenolic compounds (Leatham *et al.* 1983; Camarero *et al.* 1994; Fountoulakis *et al.* 2002; Shanmugam *et al.* 2010). As the *ortho*-cleavage pathway produces intermediates which enter the β-ketoadipate pathway, it also provides a nutritional advantage to the fungus (Hammerbacher *et al.* 2013; Wadke *et al.* 2016).

The single *CDO* present in *Huntiella* provides interesting insight into some differences between the pathogenic and saprophytic species. The lack of synteny observed between the flanking regions surrounding *CDO3* of *Huntiella* and other genera of the Ceratocystidaceae provides further evidence of differences in the evolutionary histories of the gene in the different species. Discordance in local synteny has been found to indicate nonorthologous gene relationships (Jun *et al.* 2009) and may contribute to the lifestyle adaptations of fungi (Yoder and Turgeon 2001). A lack of local synteny surrounding genes involved in pathogenicity in the Ceratocystidaceae has also been observed for genes involved in the utilization of plant-derived sucrose (Van der Nest *et al.* 2015), where the gene loss and lack of synteny were attributed to the action of transposable elements (TEs) found in the Ceratocystidaceae genomes. TEs have been shown to have an important impact on fungal lifestyles (Grandaubert *et al.* 2014; Van der Nest *et al.* 2015; Muszewska *et al.* 2019).

Although the *meta*-cleavage pathway has not been characterized in fungi, the presence of the extradiol dioxygenase (*CDO3*) in the *Huntiella* genomes and the increase in expression of CDO3 in *H. moniliformis* when grown in the presence of caffeic acid indicates that it does utilize the enzyme (data not shown). In *E. polonica* expression of CDO3, which was shown to have ring-cleaving activity *in vitro*, also increased during spruce infection (Wadke *et al.* 2016). As the complete *meta*-degradation pathway has not

Table 2. Identification of *ortho*-cleavage genes involved in the degradation of catecholic substrates in the genomes of the Ceratocystidaceae.

Species	*Ortho*-cleavage pathway genes			
	Carboxy-*cis*, *cis*-muconate cyclase	3-oxoadipate enol-lactonase	3-oxoadipate CoA-transferase	Acetyl-CoA acyltransferase
A. cleistominuta	—	—	—	√
A. xylebori	—	—	—	√
Be. basicola	√	√	√	√
Br. fagacearum	√	—	—	√
Ca. adiposa	—	—	—	√
Ceratocystis albifundus	√	√	√	√
Ce. eucalypticola	√	√	√	√
Ce. cacaofunesta	√	√	√	√
Ce. fimbriata	√	√	√	√
Ce. harringtonii	√	√	√	√
Ce. manginecans	√	√	√	√
Ce. platani	√	√	√	√
Ce. smalleyi	√	√	√	√
Ch. thielavioides	√	√	√	√
Davidsoniella australis	√	√	√	√
D. eucalypti	√	√	√	√
D. neocaledoniae	√	√	√	√
D. virescens	√	√	√	√
E. laricicola	√	√	√	√
E. polonica	√	√	√	√
H. bhutanensis	—	—	—	√
H. decipiens	—	—	—	√
H. moniliformis	—	—	—	√
H. omanensis	—	—	—	√
H. savannae	—	—	—	√
Thielaviopsis cerberus	√	√	√	√
T. ethacetica	√	√	√	√
T. euricoi	√	√	√	√
T. musarum	√	√	√	√
T. punctulata	√	√	√	√

Reference genes used in BLAST query: Carboxy-*cis*, *cis*-muconate cyclase (*Ce. fimbriata* PHH5432.1); 3-oxoadipate enol-lactonase (*Ce. platani* KKF95260.1); 3-oxoadipate CoA-transferase (*Ce. platani* KKF92536.1); acetyl-CoA acyltransferase (*Ce. fimbriata* PHH52801.1).

been resolved in fungi, there are 2 possible scenarios. The first is that all the fungi in this study can fully metabolize the phenolics using an as yet uncharacterized pathway involving extradiol cleavage. The second scenario is that the extradiol enzyme detoxifies the plant secondary metabolites through cleavage, promoting the survival of the fungus, but does not fully metabolize them (Yu and Keller 2005). Such a scenario was observed in the saprophyte *Ophiostoma piceae* which was able to initially degrade, but not metabolize, monoterpenoids to ensure its continued survival (Yu and Keller 2005; Haridas *et al.* 2013). Further investigation into the importance of extradiol dioxygenases in fungal plant infection should be undertaken to determine how they are utilized, as well as their role during plant infection. Due to their presence across all the species of the Ceratocystidaceae, it is likely that the extradiol dioxygenases are important for the metabolism of both the pathogenic and saprophytic species.

Conclusions

In this study, we considered the evolution of *CDOs* in a subset of species of the Ceratocystidaceae and their potential links with fungal lifestyle. Publicly available genomes were used to identify the full complement of genes encoding catechol dioxygenase enzymes in multiple species of the Ceratocystidaceae. While the genomes of the necrotrophic pathogens, with the exception of *Be. fagacearum*, contained 4 different genes encoding *CDOs*, mildly pathogenic species only contained 2 to 3 genes and saprophytic species only contained a single gene. The loss of the genes and the associated metabolic pathway appears to have occurred in a lineage-specific manner.

Further investigation into the ecological importance of the CDOs in a wider range of species from the Ceratocystidaceae needs to be conducted to determine which CDOs are favored by species that occupy different ecological niches. Future studies should also include the analysis of *CDO* gene expression while the fungi are occupying their specific niches and enzyme activity assays. This information will further aid in understanding the importance of CDOs in plant-pathogen fitness and evolution.

Data availability

The datasets used and/or analyzed in this study are available in the NCBI Database (https://www.ncbi.nlm.nih.gov/nucleotide/). Accession numbers are detailed in Table 1.

Supplementary material is available at *G3* online.

Acknowledgments

The authors would like to acknowledge the University of Pretoria and the Forestry and Agricultural Biotechnology Institute for use of infrastructure and the National Research Foundation for funds to BDW (DSI/NRF SARChI Chair for Fungal Genetics) and the Max Planck Society for funds to AH.

Funding

The funding for this research was provided by the Forestry and Agricultural Biotechnology Institute (FABI) and the DSI/NRF SARChI Chair for Fungal Genetics and Genomics.

Conflicts of Interest

None declared.

Literature cited

Adawi AA, Khan IA, Barnes I, Subhi AA, Jahwari AAA, Deadman ML, Wingfield BD, Wingfield MJ. *Ceratocystis manginecans* associated with a serious wilt disease of two native legume trees in Oman and Pakistan. Australasian Plant Pathol. 2013;42(2):179–193.

Anderson JP, Gleason CA, Foley RC, Thrall PH, Burdon JB, Singh KB. Plants versus pathogens: an evolutionary arms race. Funct Plant Biol. 2010;37(6):499–512. doi:10.1071/FP09304.

Appel DN, Peters R, Lewis R. Tree susceptibility, inoculum availability, and potential vectors in a Texas oak wilt center. J Arboric. 1987;13(7):169–173.

Arx JA, Hennebert GL. Deux champignons ambrosia. Mycopathol Mycol Appl. 1965;25(3–4):309–315.

Auclair SM, Bhanu MK, Kendall DA. Signal peptidase I: cleaving the way to mature proteins. Protein Sci. 2012;21(1):13–25. doi: 10.1002/pro.757.

Bal TL, Richter DL, Storer AJ, Jurgensen MF. The relationship of the sapstreak fungus, *Ceratocystis virescens*, to sugar maple dieback and decay in Northern Michigan. Am J Plant Sci. 2013;4(2): 436–443. doi:10.4236/ajps.2013.42a056.

Broderick JB. Catechol dioxygenases. Essays Biochem. 1999;34: 173–189.

Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. Pathways for degradation of lignin in bacteria and fungi. Nat Prod Rep. 2011; 28(12):1883–1896. doi:10.1039/c1np00042j. http://xlink.rsc.org/?DOI=c1np00042j.

Camarero S, Galletti GC, Martinez AT. Preferential degradation of phenolic lignin units by two white rot fungi. Appl Environ Microbiol. 1994;60(12):4509–4516.

Charlton AJ, Baxter NJ, Khan ML, Moir AJG, Haslam E, Davies AP, Williamson MP. Polyphenol/peptide binding and precipitation. J Agric Food Chem. 2002;50(6):1593–1601. doi:10.1021/jf010897z.

Csűrös M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics. 2010;26(15): 1910–1912. doi:10.1093/bioinformatics/btq315.

Csűrös M, Miklós I. A probabilistic model for gene content evolution with duplication, loss, and horizontal transfer. Res Comput Mol Biol. 2006;3909:206–220.

Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 2011;27(8): 1164–1165. doi:10.1093/bioinformatics/btr088.ProtTest.

Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and high-performance computing Europe PMC Funders Group. Nat Methods. 2012;9(8):772.doi: 10.1038/nmeth.2109.jModelTest.

de Beer ZW, Marincowitz S, Duong TA, Wingfield MJ. *Bretziella*, a new genus to accommodate the oak wilt fungus, *Ceratocystis fagacearum* (Microascales, Ascomycota). MC. 2017;27:1–19. doi: 10.3897/mycokeys.27.20657.

Eltis LD, Bolin JT. Evolutionary relationship among extradiol dioxygenases. J Bacteriol. 1996;178(20):5930–5937.

Ferreira R, Monteiro S, Freitas R, Santos C, Chen Z, Batista L, Duarte J, Borges A, Teixeira A. The role of plant defence proteins in fungal pathogenesis. Mol Plant Pathol. 2007;8(5):677–700. doi: 10.1111/J.1364–3703.2007.00419.X.

Fountoulakis MS, Dokianakis SN, Kornaros ME, Aggelis GG, Lyberatos G. Removal of phenolics in olive mill wastewaters using the white-rot fungus *Pleurotus ostreatus*. Water Res. 2002; 36(19):4735–4744. doi:10.1016/S0043-1354(02)00184-7.

Frantzeskakis L, Di PA, Rep M, Schirawski J, Wu C, Panstruga R. Rapid evolution in plant – microbe interactions – a molecular genomics perspective. New Phytol. 2020;225(3):1134–1142. doi: 10.1111/nph.15966.

Gluck-Thaler E, Slot J. Specialized plant biochemistry drives gene clustering in fungi. ISME J. 2018;12(7):1694–1705. doi:10.1038/s41396-018–0075-3.

Gluck-Thaler E, Vijayakumar V, Slot JC. Fungal adaptation to plant defences through convergent assembly of metabolic modules. Mol Ecol. 2018;27(24):5120–5136. doi:10.1111/mec.14943.

Grandaubert J, Balesdent M, Rouxel T. Evolutionary and adaptive role of transposable elements in fungal genomes. Adv Bot Res. 2014;70:79–107. http://dx.doi.org/10.1016/B978-0–12-397940-7.00003-3.

Halsted B, Fairchild D. Sweet-potato black rot (*Ceratocystis fimbriata*, Ell. & Hals.). J Mycol. 1891;7(1):1–11.

Hammerbacher A, Schmidt A, Wadke N, Wright LP, Schneider B, Bohlmann J, Brand WA, Fenning TM, Gershenzon J, Paetz C. A common fungal associate of the spruce bark beetle metabolizes the stilbene defenses of Norway spruce. Plant Physiol. 2013; 162(3):1324–1336. doi:10.1104/pp.113.218610.

Harayama S, Rekik M. Bacterial aromatic ring-cleavage enzymes are classified into two different gene families. J Biol Chem. 1989; 264(26):15328–15333.

Haridas S, Wang Y, Lim L, Alamouti SM, Jackman S, Docking R, Robertson G, Birol I, Bohlmann J, Breuil C. The genome and transcriptome of the pine saprophyte *Ophiostoma piceae*, and a comparison with the bark beetle-associated pine pathogen *Grosmannia clavigera*. BMC Genomics. 2013;14:373.doi:10.1186/1471–2164-14–373.

Harwood CS, Parales RE. The β-ketoadipate pathway and the biology of self-identity. Annu Rev Microbiol. 1996;50(1):553–590. doi: 10.1146/annurev.micro.50.1.553.

Hayaishi O. Crystalline oxygenases of pseudomonads. Bacteriol Rev. 1966;30(4):720–731.

Hoff KJ, Stanke M. WebAUGUSTUS–a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res. 2013;41:123–128. doi:10.1093/nar/gkt418.

Hoff KJ, Stanke M, Contact G. TrainAUGUSTUS—a web server application for parameter training and gene prediction in eukaryotes. In: International Plant & Animal XX Conference, 2012, U.S.A. (Poster).

Hulcr J, Dunn RR. The sudden emergence of pathogenicity in insect-fungus symbioses threatens naive forest ecosystems. Proc Biol Sci. 2011;278(1720):2866–2873. doi:10.1098/rspb.2011.1130.

Jankowiak R. Assessing the pathogenicity of three sap-stain fungi associated with *Tomicus minor* to Scots pine seedlings. Phytopathologia. 2011;61:9–15.

Jeandet P, Douillet-Breuil AC, Bessis R, Debord S, Sbaghi M, Adrian M. Phytoalexins from the vitaceae: biosynthesis, phytoalexin gene expression in transgenic plants, antifungal activity, and metabolism. J Agric Food Chem. 2002;50(10):2731–2741. doi:10.1021/jf011429s.

Johnson JA, Harrington TC, Engelbrecht CJB. Phylogeny and taxonomy of the North American clade of the *Ceratocystis fimbriata*

complex. Mycologia. 2005;97(5):1067–1092. doi:10.3852/mycologia.97.5.1067.

Jones JDG, Dangl JL. The plant immune system. Nature. 2006; 444(7117):323–330. doi:10.1038/nature05286.

Jones P, Binns D, Chang H, Fraser M, Li W, Mcanulla C, Mcwilliam H, Maslen J, Mitchell A, Nuka G, *et al.* Sequence analysis InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–1240. doi:10.1093/bioinformatics/btu031.

Jun J, Mandoiu II, Nelson CE. Identification of mammalian orthologs using local synteny. BMC Genomics. 2009;10(1):630.doi:10.1186/1471–2164-10–630.

Juzwik J, French D. *Ceratocystis fagacearum* and *C. piceae* on the surfaces of free-flying and fungus-mat-inhabiting nitidulids. Phytopathology. 1983;73(8):1164–1168.

Juzwik J, Harrington TC, Macdonald WL, Appel DN. The origin of *Ceratocystis fagacearum*, the oak wilt fungus. Annu Rev Phytopathol. 2008;46:13–26. doi:10.1146/annurev.phyto.45.062806.094406.

Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 2008;9(4):286–298. doi:10.1093/bib/bbn013.

Kile G, Harrington T, Yuan Z, Dudzinski M, Old K. *Ceratocystis eucalypti* sp. nov., a vascular stain fungus from eucalypts in Australia. Mycol Res. 1996;100(5):571–579. doi:10.1016/S0953-7562(96)80010-4.

Kocaçalişkan I, Talan I, Terzi I. Antimicrobial activity of catechol and pyrogallol as allelochemicals. Zeitschrift fur Naturforsch—Sect C. Z Naturforsch C J Biosci. 2006;61(9–10):639–642. doi:10.1515/znc-2006-9-1004.

Krogh A, Larsson È, Heijne GV, Sonnhammer ELL. Predicting transmembrane protein topology with a Hidden Markov Model: application to complete genomes. J Mol Biol. 2001;305(3):567–580. doi:10.1006/jmbi.2000.4315.

Krokene P, Solheim H. Growth of four bark-beetle-associated bluestain fungi in relation to the induced wound response in Norway spruce. Can J Bot. 1997;75(4):618–625.

Krokene P, Solheim H. Pathogenicity of four blue-stain fungi associated with aggressive and nonaggressive bark beetles. Phytopathology. 1998;88(1):39–44.

Kück P, Longo GC. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front Zool. 2014;11(1):81.doi:10.1186/s12983-014–0081-x.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–1549. doi:10.1093/molbev/msy096.

Lattanzio V, Kroon P, Quideau S, Treutter D. Plant phenolics—secondary metabolites with diverse functions. In: Daayf F, Lattanzio V, editors. Recent Advances in Polyphenol Research. Oxford, UK: Wiley-Blackwell. Vol. 1. 2008. p. 1–35. [Database]

Lattanzio V, Lattanzio VMT, Cardinali A, Amendola V. Role of phenolics in the resistance mechanisms of plants against fungal pathogens and insects. Phytochem Adv Res. 2006;661(2):23–67.

Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25(7):1307–1320. doi:10.1093/molbev/msn067.

Leatham GF, Crawford RL, Kirk TK. Degradation of phenolic compounds and ring cleavage of catechol by *Phanerochaete chrysosporium*. Appl Environ Microbiol. 1983;46(1):191–197.

Levin DA. Plant phenolics: an ecological perspective. Am Nat. 1971; 105(942):157–181.

Li L, Stoeckert CJJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13(9):2178–2189. doi:10.1101/gr.1224503.

Liu Y, Hassan S, Kidd BN, Garg G, Mathesius U, Singh KB, Anderson JP. Ethylene signaling is important for isoflavonoid-mediated resistance to *Rhizoctonia solani* in roots of *Medicago truncatula*. Mol Plant Microbe Interact. 2017;30(9):691–700. doi:10.1094/MPMI-03–17-0057-R.

Longree K. *Chalaropsis thielavioides*, cause of "black mold" of rose grafts. Phytopathology. 1940;30(10):793–807.

Maddison WP, Maddison DR. *Mesquite: a modular system for evolutionary analysis*. 2019. http://www.mesquiteproject.org.

Martins TM, Hartmann DO, Planchon S, Martins I, Renaut J, Silva Pereira C. The old 3-oxoadipate pathway revisited: new insights in the catabolism of aromatics in the saprophytic fungus *Aspergillus nidulans*. Fungal Genet Biol. 2015;74:32–44. doi:10.1016/j.fgb.2014.11.002.

Maruyama K, Shibayama T, Ichikawa A, Sakou Y, Yamada S, Sugisaki H. Cloning and characterization of the genes encoding enzymes for the protocatechuate *meta*- degradation pathway of *Pseudomonas ochraceae*. Biosci Biotechnol Biochem. 2004;68(7):1434–1441. doi:10.1271/bbb.68.1434.

Mayers CG, Harrington TC, Masuya H, Jordal BH, Mcnew DL, Shih H, Roets F, Kietzka GJ. Patterns of coevolution between ambrosia beetle mycangia and the Ceratocystidaceae, with five new fungal genera and seven new species. Persoonia. 2020;44:41–66.

Mayers CG, Mcnew DL, Harrington TC, Roeper RA, Fraedrich SW, Biedermann PHW, Castrillo LA, Reed SE, Peterson SW. Three genera in the Ceratocystidaceae are the respective symbionts of three independent lineages of ambrosia beetles with large, complex mycangia. Fungal Biol. 2015;119(11):1075–1092. doi:10.1016/j.funbio.2015.08.002.

Melo MD, Matos S, Evando J, Beserra A, Moreira SI, Silva FF, Neto FDA, Alves E, Pereira OL. *Thielaviopsis musarum* causes postharvest crown and fruit rot of banana in Northeastern Brazil. Trop Plant Pathol. 2016;41(4):258–263. doi:10.1007/s40858-016–0094-4.

Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K. Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci Rep. 2019;9(1):4307–4310. doi:10.1038/s41598-019–40965-0.

Ndhlala AR, Ghebrehiwot HM, Ncube B, Aremu AO, Gruz J, Šubrtová M, Doležal K, Du Plooy CP, Abdelgadir HA, Van Staden J. Antimicrobial, anthelmintic activities and characterisation of functional phenolic acids of *Achyranthes aspera* Linn.: a medicinal plant used for the treatment of wounds and ringworm in east Africa. Front Pharmacol. 2015;6:1–8. doi:10.3389/fphar.2015.00274.

Nel WJ, Duong TA, Wingfield BD, Wingfield MJ, de Beer ZW. A new genus and species for the globally important, multihost root pathogen *Thielaviopsis basicola*. Plant Pathol. 2018;67(4):871–812. doi:10.1111/ppa.12803.

Ocasio-Morales R, Panaghiotis T, Harrington TC. Origin of *Ceratocystis platani* on native *Platanus orientalis* in Greece and its impact on natural forests. Plant Dis. 2007;91(7):901–904.

Ornston LN, Stanier RY. The conversion of catechol and protocatechuate to β-ketoadipate. J Biol Chem. 1966;241(16):3776–3786.

Paulin-Mahady AE, Harrington TC, McNew D. Phylogenetic and taxonomic evaluation of *Chalara, Chalaropsis*, and *Thielaviopsis* anamorphs associated with *Ceratocystis*. Mycologia. 2002;94(1):62–72. doi:10.1080/15572536.2003.11833249.

Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8(10):785–786. doi:10.1038/nmeth.1701.

Punja ZK, Sun LJ. Morphological and molecular characterization of *Chalara elegans* (*Thielaviopsis basicola*), cause of black root rot on diverse plant species. Can J Bot. 1999;77(12):1801–1812. doi:10.1139/cjb-77-12-1801.

Rao A, Zhang Y, Muend S, Rao R. Mechanism of antifungal activity of terpenoid phenols resembles calcium stress and inhibition of the TOR pathway. Antimicrob Agents Chemother. 2010;54(12):5062–5069. doi:10.1128/AAC.01050-10.

Richter DL. The sugar maple sapstreak fungus (*Ceratocystis virescens* (Davidson) Moreau, Ascomycota) in the Huron mountains, Marquette county, Michigan. Michigan Bot. 2012;51:73–81.

Roux J, Heath RN, Labuschagne L, Nkuekam GK, Wingfield MJ. Occurrence of the wattle wilt pathogen, *Ceratocystis albifundus* on native South African trees. Forest Pathol. 2007;37(5):292–302. doi:10.1111/j.1439-0329.2007.00507.x.

Roux J, Van Wyk M, Hatting H, Wingfield MJ. *Ceratocystis* species infecting stem wounds on *Eucalyptus grandis* in South Africa. Plant Pathol. 2004;53(4):414–421. doi:10.1046/j.1365–3059.2004.01014.x.

Saeed EE, Sham A, El-Tarabily K, Abu Elsamen F, Iratni R, AbuQamar SF. Chemical control of black scorch disease on date palm caused by the fungal pathogen *Thielaviopsis punctulata* in United Arab Emirates. Plant Dis. 2016;100(12):2370–2376. doi:10.1094/pdis-05–16-0645-re.

Sang Sung W, Gun Lee D. Antifungal action of chlorogenic acid against pathogenic fungi, mediated by membrane disruption. Pure Appl Chem. 2010;82(1):219–226. doi:10.1351/PAC-CON-09-01-08.

Sartoris GB. A cytological study of *Ceratostomella adiposum* (Butl.) comb. nov., the black-rot fungus of sugar cane. J Agric Res. 1927;35(7):577–585.

Schäfer W, Straney D, Ciuffetti L, Van Etten HD, Yoder OC. One enzyme makes a fungal pathogen, but not a saprophyte, virulent on a new host plant. Science. 1989;246(4927):247–249. doi:10.1126/science.246.4927.247.

Shalaby S, Horwitz BA. Plant phenolic compounds and oxidative stress: integrated signals in fungal–plant interactions. Curr Genet. 2015;61(3):347–357. doi:10.1007/s00294-014–0458-6.

Shanmugam V, Ronen M, Shalaby S, Larkov O, Rachamim Y, Hadar R, Rose MS, Carmeli S, Horwitz BA, Lev S. The fungal pathogen *Cochliobolus heterostrophus* responds to maize phenolics: novel small molecule signals in a plant-fungal interaction. Cell Microbiol. 2010;12(10):1421–1434. doi:10.1111/j.1462–5822.2010.01479.x.

Shigo AL. Observations on the succession of fungi on hardwood pulpwood bolts. Plant Dis Report. 1962;46(5):379–380.

Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Norman W, Rattray M, Hubbard SJ, Oliver SG, Talbot NJ. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS One. 2008;3(6):e2300.doi:10.1371/journal.pone.0002300.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–1313. doi:10.1093/bioinformatics/btu033.

Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27(7):1009–1010. doi:10.1093/bioinformatics/btr039.

Swofford DL. 2002. *PAUP version 4.0 b10. Phylogenetic Analysis Using Parsimony (and Other Methods)*. Sunderland, Massachusetts: Sinauer Associates.

Tarigan M, Roux J, Van Wyk M, Tjahjono B, Wingfield MJ. A new wilt and die-back disease of *Acacia mangium* associated with *Ceratocystis manginecans* and *C. acaciivora* sp. nov. in Indonesia. South African J Bot. 2011;77(2):292–304. doi:10.1016/j.sajb.2010.08.006.

Ullah C, Unsicker SB, Fellenberg C, Constabel CP, Schmidt A, Gershenzon J, Hammerbacher A. Flavan-3-ols are an effective chemical defense against rust infection. Plant Physiol. 2017;175(4):1560–1578. doi:10.1104/pp.17.00842.

Vaillancourt F, Bolin J, Lindsay D. The ins and outs of ring-cleaving dioxygenases. Crit Rev Biochem Mol Biol. 2006;41(4):241–267. doi:10.1016/j.tcb.2004.07.007.

van der Nest MA, Beirn LA, Crouch JA, Demers JE, de Beer ZW, De Vos L, Gordon TR, Moncalvo J-M, Naidoo K, Sanchez-Ramirez S, *et al.* IMA Genome-F 3: draft genomes of *Amanita jacksonii, Ceratocystis albifundus, Fusarium circinatum, Huntiella omanensis, Leptographium procerum, Rutstroemia sydowiana*, and *Sclerotinia echinophila*. IMA Fungus. 2014;5(2):472–485. doi:10.5598/imafungus.2014.05.02.11.

van der Nest MA, Steenkamp ET, Mctaggart AR, Trollip C, Godlonton T, Sauerman E, Roodt D, Naidoo K, Coetzee MPA, Wilken PM, *et al.* Saprophytic and pathogenic fungi in the Ceratocystidaceae differ in their ability to metabolize plant-derived sucrose. BMC Evol Biol. 2015;15(1):273. doi:10.1186/s12862-015–0550-7.

van Wyk M, Roux J, Barnes I, Wingfield BD, Wingfield MJ. Molecular phylogeny of the *Ceratocystis moniliformis* complex and description of *C. tribiliformis* sp. nov. Fungal Divers. 2006;21:181–201.

Wadke N, Kandasamy D, Vogel H, Lah L, Wingfield BD, Paetz C, Wright LP, Gershenzon J, Hammerbacher A. The bark-beetle-associated fungus, *Endoconidiophora polonica*, utilizes the phenolic defense compounds of its host as a carbon source. Plant Physiol. 2016;171(2):914–931. doi:10.1104/pp.15.01916.

Westrick NM, Smith DL, Kabbage M. Disarming the host: detoxification of plant defense compounds during fungal necrotrophy. Front Plant Sci. 2021;12:651716–651718. doi:10.3389/fpls.2021.651716.

Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 2001;18(5):691–699.

Wilken PM, Steenkamp ET, Wingfield MJ, de Beer ZW, Wingfield BD. Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. IMA Fungus. 2013;4(2):357–358. doi:10.5598/imafungus.2013.04.02.14.

Wilson AM, Godlonton T, van der Nest MA, Wilken PM, Wingfield MJ, Wingfield BD. Unisexual reproduction in *Huntiella moniliformis*. Fungal Genet Biol. 2015;80:1–9. doi:10.1016/j.fgb.2015.04.008.

Wingfield BD, Ambler JM, Coetzee MPA, de Beer ZW, Duong TA, Joubert F, Hammerbacher A, McTaggart AR, Naidoo K, Nguyen HDT, *et al.* Draft genome sequences of *Armillaria fuscipes, Ceratocystiopsis minuta, Ceratocystis adiposa, Endoconidiophora*

laricicola, *E. polonica* and *Penicillium freii* DAOMC 242723. IMA Fungus. 2016a;7(1):217–227. doi:10.5598/imafungus.2016.07.01.11.

Wingfield BD, Barnes I, de Beer ZW, De Vos L, Duong TA, Kanzi AM, Naidoo K, Nguyen HDT, Santana QC, Sayari M, *et al.* 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. 2015;6(2):493–506. doi:10.5598/imafungus.2015.06.02.13.

Wingfield BD, Berger DK, Steenkamp ET, Lim H-J, Duong TA, Bluhm BH, de Beer ZW, De Vos L, Fourie G, Naidoo K, *et al.* Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus, Huntiella decipiens* and *Ophiostoma ips*. IMA Fungus. 2017;8(2):385–396. doi:10.5598/imafungus.2017.08.02.10.

Wingfield BD, Duong TA, Hammerbacher A, van der Nest MA, Wilson A, Chang R, Wilhelm de Beer Z, Steenkamp ET, Markus Wilken P, Naidoo K, *et al.* IMA Genome-F 7 Draft genome sequences for *Ceratocystis fagacearum, C. harringtonii, Grosmannia penicillata*, and *Huntiella bhutanensis*. IMA Fungus. 2016b;7(2):317–323. doi:10.5598/imafungus.2016.07.02.11.

Wingfield BD, Liu M, Nguyen HDT, Lane FA, Morgan SW, de Vos L, Markus P, Duong TA, Aylward J, Coetzee MPA, *et al.* Nine draft genome sequences of *Claviceps purpurea* s.lat., including *C. humidiphila*, and *C.* cf. *spartinae*, pseudomolecules for the pitch canker pathogen *Fusarium circinatum*, draft genome of *Davidsoniella eucalypti*, *Grosmannia galeiformis, Quambalaria eucalypti*, and *Teratosphaeria destructans*. IMA Fungus. 2018;9(2):401–418. doi:10.5598/imafungus.2018.09.02.10.

Yoder OC, Turgeon BG. Fungal genomics and pathogenicity. Curr Opin Plant Biol. 2001;4(4):315–321.

Yu J-H, Keller N. Regulation of secondary metabolism in filamentous fungi. Annu Rev Phytopathol. 2005;43:437–458. doi:10.1146/annurev.phyto.43.040204.140214.

Communicating editor: A. Rokas