Fungal Genetics and Biology 143 (2020) 103433

Contents lists available at ScienceDirect

Fungal Genetics and Biology

journal homepage: www.elsevier.com/locate/yfgbi

Genome comparisons suggest an association between *Ceratocystis* host adaptations and effector clusters in unique transposable element families

Arista Fourie[a], Ronnie de Jonge[b], Magriet A. van der Nest[a,c], Tuan A. Duong[a], Michael J. Wingfield[a], Brenda D. Wingfield[a], Irene Barnes[a,*]

[a] Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa
[b] Plant-Microbe Interactions, Department of Biology, Science4Life, Utrecht University, Utrecht 3584 CH, the Netherlands
[c] Biotechnology Platform, Agricultural Research Council, Private Bag X05, Onderstepoort 0110, 0002, South Africa

ARTICLE INFO

Keywords:
Ceratocystidaceae
Chromosomal rearrangement
RIP
Effectors
Fungi
Genome comparison

ABSTRACT

Ceratocystis fimbriata is a host specific fungal pathogen of sweet potato (*Ipomoea batatas*). The closely related species, *C. manginecans,* is an important pathogen of trees (e.g. *Acacia mangium* and *Mangifera indica*) but has never been isolated from tuber crops. The genetic factors that determine the host range and host specificity of these species have not been determined. The aim of this study was to compare the genomes of *C. fimbriata* and *C. manginecans* in order to identify species-specific genetic differences that could be associated with host specificity. This included whole-genome alignments as well as comparisons of gene content and transposable elements (TEs). The genomes of the two species were found to be very similar, sharing similar catalogues of CAZymes, peptidases and lipases. However, the genomes of the two species also varied, harbouring species-specific genes (e.g. small secreted effectors, nutrient processing proteins and stress response proteins). A portion of the TEs identified (17%) had a unique distribution in each species. Transposable elements appeared to have played a prominent role in the divergence of the two species because they were strongly associated with chromosomal translocations and inversions as well as with unique genomic regions containing species-specific genes. Two large effector clusters, with unique TEs in each species, were identified. These effectors displayed non-synonymous mutations and deletions, conserved within a species, and could serve as mutational hot-spots for the development of host specificity in the two species.

1. Introduction

Speciation in fungal plant pathogens is commonly associated with host jumps or host range expansions, followed by subsequent reproductive isolation (Restrepo et al. 2014; Stukenbrock 2013). The host range of a fungal species can be determined at different levels of pathogen-host interaction such as host penetration, nutrient acquisition and resistance to host immune defence responses (Borah et al. 2018; Dracatos et al. 2018). For example, the production of different secondary metabolites such as toxins (Walkowiak et al. 2016) or different transporter proteins that detoxify host defence products (Buiate et al. 2017) could determine host specificity.

Carbohydrate active enzymes (CAZymes), peptidases and lipases play a significant role in nutrient acquisition of fungal species. CAZymes are essential in plant cell wall degradation where they target cellulose, hemicellulose, lignin and pectin (Amselem et al. 2011; King et al. 2011). Differences in CAZymes in different fungal pathogens influence their ability to penetrate cell walls and obtain nutrients from a specific plant host (Buiate et al. 2017; King et al. 2011; Shirke et al. 2016). Peptidases hydrolyse protein peptide bonds and can provide amino acids as nutrients, break down signalling proteins or break down host products produced during defence response (Rawlings and Morton 2008). Lipases break down triglycerides into glycerol and fatty acids (Singh and Mukhopadhyay 2012), and they also form or break down ester bonds for lipid metabolism (Widmann and Pleiss 2016). Differences in these enzyme repertoires could influence the host range of different pathogens.

The interactions between fungal and plant immune systems are likely to play the most prominent role in host specificity. Host defence responses can either be activated when conserved pathogen-associated molecular patterns are detected (PAMP-triggered immunity) or when receptors in the host detect fungal secreted proteins in the apoplastic or cytoplasmic environment (effector-triggered immunity) (Jones and Dangl 2006; Presti et al. 2015). The differences in the presence/absence

* Corresponding author.
E-mail address: irene.barnes@fabi.up.ac.za (I. Barnes).

https://doi.org/10.1016/j.fgb.2020.103433
Received 25 January 2020; Received in revised form 18 June 2020; Accepted 30 June 2020
Available online 08 July 2020
1087-1845/

of small secreted effector proteins or variation in their amino acid sequences can influence the ability of a fungus to infect a plant host. Effectors allow a pathogen to avoid detection, to protect itself against host defence reactions and to interfere with host cellular processes (Inoue et al. 2017; Sánchez-Vallet et al. 2018; van Dam et al. 2016). Mutation or deletion of effectors and other host specific genes mostly occur in highly dynamic genomic compartments, such as transposable element (TE) dense regions (Shirke et al. 2016; Yoshida et al. 2016) and these regions are often being targeted by Repeat Induced Point mutations (RIP) (Grandaubert et al. 2014).

Ceratocystis species are important fungal pathogens of a wide range of woody plants (Barnes et al., 2018; Harrington, 2004; van Wyk et al., 2013; Roux et al., 2020). A few species are post-harvest agricultural pathogens that cause rot on tuber crops (Halsted and Fairchild 1891; Li et al. 2016; Liu et al. 2018). *Ceratocystis* species require wounds or natural openings to infect their hosts after which they move through the parenchymal cells to the xylem where spores are released and rapidly disseminated (Araujo et al. 2014). The fungal mycelia, in combination with tyloses and phenolics produced as host defence responses, block water transport in the xylem that can result in wilting and eventually, tree death (Araujo et al. 2014; Da Silva et al. 2018; Trang et al. 2018).

Ceratocystis fimbriata and *Ceratocystis manginecans* are important pathogens of root and tree crops, respectively. *Ceratocystis fimbriata* causes black rot on *Ipomoea batatas* (sweet potato). Large disease outbreaks have recently re-emerged in the USA (Scruggs et al. 2017) and infections have also been reported in China (Li et al. 2016), Japan (Paul et al. 2018) and other Asian countries (Fourie et al. 2018; Paul et al. 2018). In this study *C. fimbriata* refers to the *I. batatas*-specific species, recently epitypified by Marincowitz et al., 2020, and also referred to by some researchers as *C. fimbriata* f. sp. *ipomoea* (Valdetaro et al. 2019). *Ceratocystis manginecans* has resulted in up to 60% losses in mango orchards in the Middle East (Da Silva Galdino et al. 2016) and has contributed to an almost 50% yield loss in *A. mangium* plantations in South East Asian countries (Hardie et al. 2018; Harwood and Nambiar 2014). This species is also pathogenic to *Punica granatum* (pomegranate) (Harrington et al. 2014a; Harrington et al. 2014b; Huang et al. 2003), the leguminous trees *Prosopis cineraria* and *Dalbergia sissoo* (Al Adawi et al. 2013) and some *Eucalyptus* species (Chen et al. 2013). *Ceratocystis manginecans* has been referred to as *C. fimbriata* based on a biological species concept (Harrington et al. 2014b), but in this study, based on a phylogenetic species concept (Fourie et al. 2014), we consider it a distinct species.

Ceratocystis fimbriata and *C. manginecans* are closely related and can be induced to mate under laboratory conditions (Fourie et al. 2018; Oliveira et al. 2015). However, inoculations with each of these species onto *I. batatas* and different tree species have shown that *C. fimbriata* is highly specific to *I. batatas* while *C. manginecans* can infect several tree species but is not pathogenic to *I. batatas* (Baker et al. 2003; Fourie et al. 2018). Nothing is known regarding the genetic factors that determine the distinctly different host ranges of the two species; whether this is determined by their ability to penetrate the host cells, nutrient availability, or the PAMP- and effector-triggered immune interactions.

A small number of genomic studies on *Ceratocystis* species have focused on their pathogenicity and host specificity (Fourie et al. 2019; Molano et al. 2018). A recent study on an interspecific cross between *C. fimbriata* and *C. manginecans* investigated the inheritance of pathogenicity and host specificity traits (Fourie et al. 2019). This shed some light on the genetic factors associated with these traits. One QTL region was identified to be associated with aggressiveness on *A. mangium* and two QTLs for aggressiveness on *I. batatas*. Candidate genes in these regions included genes involved in oxidative stress response, protein processing and putative effector proteins that were often in regions with a high transposable element content (Fourie et al. 2019). These QTLs could explain roughly 20% of the variation observed in the phenotype, suggesting that additional genomic regions are involved in determining host specificity.

A recent genomic study, considering a small number of *Ceratocystis* species, identified several candidate genes involved in their pathogenicity, but not necessarily host specificity (Molano et al. 2018). This was mainly based on a comparative genomic study between the cacao (*Theobroma cacao*) pathogen, *Ceratocystis cacaofunesta*, and *C. fimbriata* from sweet potato. Pathogenicity factors included an expanded phosphatidylinositol-specific phospholipase-C (PI-PLC) gene family, CAZymes and candidate effector proteins. Both *C. cacaofunesta* and *C. fimbriata* had numerous unique genes, possibly associated with their host specificity. Unique protein families were also identified in *C. manginecans*, but these were not further investigated. It should be noted that the *C. fimbriata* and *C. manginecans* genome assemblies used in the Molano et al. (2018) study were versions that consisted of numerous contigs, and results would likely improve with the more recent and improved assemblies now available (Fourie et al. 2019; Wingfield et al. 2019).

The factors that influence host specificity of fungal pathogens can often be detected at a genomic level, based on differences in gene content and sequence variation (Borah et al. 2018). The aim of the present study was to consider genomic differences between *C. fimbriata* and *C. manginecans* that might underpin their host specificity. The genome sequences of three *C. manginecans* isolates from *P. cineraria* and *A. mangium,* from three different countries (Fourie et al. 2019; van der Nest et al. 2014), and one *C. fimbriata* isolate from *I. batatas* from the USA (Wingfield et al. 2019), are available in public databases and genome comparisons could thus be performed between and within species. This study provides one of few comparative genomic studies between a pathogen of woody trees versus a pathogen of a herbaceous plant, which could provide novel insights into the genetic adaptations of two closely related species to highly divergent host plants.

2. Materials and methods

2.1. Genome sequences utilised

Three *C. manginecans* isolates obtained from diseased *A. mangium* (CMW22563 from Indonesia, CMW46461 from Malaysia) and *P. cineraria* (CMW17570 from Oman) and one *C. fimbriata* isolate from infected *I. batatas* (CMW14799 from USA) were used for genome comparisons (Table 1). Since *C. fimbriata* from *I. batatas* (*C. fimbriata* f. sp. *ipomoea*) is globally a clonal population with minimal variation between isolates (Li et al. 2016), a single genome would be sufficient to represent this species. Three genomes were used for *C. manginecans* because the species has been shown to have high genetic diversity (Fourie et al. 2016) and also a relatively wide host range (Al Adawi et al. 2013; Tarigan et al. 2011).

The genomes of the *C. manginecans* (CMW46461) and *C. fimbriata* (CMW14799) isolates were sequenced previously using Illumina sequence technologies (Fourie et al. 2019; Wingfield et al. 2019) and were obtained, along with two additional *C. manginecans* genomes, from the GenBank database at the National Center for Biotechnology Information (NCBI) (CMW17570 (JJRZ00000000), CMW22563 (VIFZ01000000), CMW46461 (SGIO00000000), CMW14799 (APWK03000000)). Improved genome assemblies, consisting of significantly fewer contigs and a higher sequence quality (Table 1), were available for *C. fimbriata* (CMW14799) and *C. manginecans* (CMW46461) (https://doi.org/10.17632/v6txk25v58.1; Fourie et al. 2019). These genomes consisted of Illumina, as well as Oxford Nanopore, sequence data and the contigs were concatenated *in silico,* based on synteny alignments and linkage map data, to represent putative chromosomes (Fourie et al. 2019). The putative chromosomes represent 99.6% of the respective genome sequences and the unincorporated shorter contigs (0.4%) only contain a few TEs but no protein coding genes. The two genomes assembled with Nanopore data were used in all genome comparison analyses. Because the sequencing technology can influence the prediction of genome size and repetitive elements

Table 1
Summary of the genomes of the *Ceratocystis* isolates used for genome comparisons.

Isolate	Host	Sequencing Technology	Predicted size (Mb)	N50 (bp)	No. of contigs (> 500 bp)	BUSCO complete-ness	No. of genes predicted[a]	%TEs
C. fimbriata (CMW14799)	*Ipomoea batatas*	Illumina HiSeq	30.15	173,733	399	98%	7,316 (479)	13.3
		Nanopore	32.08	3,623,386	16*	97.4%	7,274 (477)	20.8
C. manginecans (CMW46461)	*Acacia mangium*	Illumina HiSeq	32.05	593,783	231	98%	7,260 (473)	18.6
		Nanopore	32.15	2,253,610	26*	98.3%	7,186 (472)	20.3
C. manginecans (CMW22563)	*Acacia mangium*	Illumina Genome Analyzer IIx	31.9	61,072	802	98%	7,337 (481)	18.2
C. manginecans (CMW17570)	*Prosopis cineraria*	Illumina Genome Analyzer IIx	31.8	77,070	981	98%	7,210 (418)	17.8

* Number of contigs refer to the initial Nanopore genome assembly, before contigs were joined *in silico* to produce supercontigs.
[a] The annotation of a small number of genes were not transferred from the Illumina genome to the Nanopore genome, hence the difference in gene number obtained for the same isolate. The number of secreted proteins is indicated in brackets.

(Thomma et al. 2016), the Illumina-sequenced genomes for isolates CMW14799 and CMW46461 were included only for comparisons between sequencing technologies in some of the analyses (e.g. TEs and RIP). The completeness of the four Illumina and the two Nanopore genome assemblies was determined with the BUSCO (Benchmarking Universal Single-Copy Orthologs) pipeline using the Ascomycota data set (Simão et al. 2015).

The presence of the telomeric repeat sequence (TTAGGG/CCCTAA) was counted in 10 kb "windows" across the genome using BEDtools count analysis (Quinlan and Hall 2010). The distribution of the repeats was investigated throughout the supercontigs in order to determine whether the ends of the Nanopore supercontigs represent telomeric ends (van Wyk et al. 2018). Because the genomic sequence in the centromere also contains various repetitive sequences (Grady et al. 1992; Okumura-Finato et al. 2000), a high density of the telomeric repeat sequence in the middle of a contig was considered indicative of putative centromeric regions. Where a higher number of repeats was observed in the last 100 kb of a contig, relative to the rest of the contig, this was considered as a putative telomere. A high density of transposable element repeat sequences at the ends of supercontigs were also investigated to confirm the presence of putative telomeres (see section 2.6).

2.2. Gene prediction and functional annotation

Gene prediction for the genomes of all four isolates was performed with the MAKER genome annotation pipeline (Cantarel et al. 2008). The *C. fimbriata* (CMW14799) and *C. manginecans* (CMW46461) Illumina genomes had been annotated previously and these annotations have also been transferred to the Nanopore genomes (Fourie et al. 2019; Wingfield et al. 2019). Since the *C. fimbriata* predicted gene models were supported by RNA-seq data, this annotation was used to train the gene prediction software in MAKER for the annotation of the two additional *C. manginecans* genomes using the parameters described previously (Wingfield et al. 2019). All predicted genes and proteins are available publicly (https://doi.org/10.17632/v6txk25v58.1). The GFF3 annotation files obtained were quality checked using GenomeTools v. 1.5.9 (Gremme et al. 2013), which also produced general statistics on the gene and sequence content for each genome. The genome-wide distribution of the genes in the *C. fimbriata* (CMW14799) and *C. manginecans* (CMW46461) Nanopore genomes were determined by counting the number of genes per 5 Kb "window", using BEDtools count (Quinlan and Hall 2010). These data were visualised, along with other genomic data, as described in section 2.6.

The functional annotation of the predicted genes was performed for both additional *C. manginecans* genomes in a manner similar to that for the *C. fimbriata* genome (Fourie et al. 2019; Wingfield et al. 2019). Predicted proteins were compared to the NCBI curated protein database (Swissprot, March 2017), using a tera-BLASTp analysis on Decypher® v. 2 with an E-value cut-off of 1e-5 and, to increase the stringency, a percent identity threshold of 50% was used, rather than the suggested 30%. Protein family domains were also predicted from the Pfam database using InterProScan v. 5.24 (Jones et al. 2014). The functional predictions of the proteins were incorporated in the GFF3 files using ANNIE (Tate et al. 2014) and GAG (Hall et al. 2014). In addition, the proteins were compared to the pathogen host-interaction (PHI) database (Urban et al. 2015) using a BLASTp analysis. Both polyketide synthase and non-ribosomal peptide synthase secondary metabolite gene clusters have previously been investigated in *Ceratocystis* species (Sayari et al. 2019; Sayari et al. 2018). To determine whether *C. fimbriata* and *C. manginecans* have the capacity to produce any other secondary metabolites, all secondary metabolite gene clusters were predicted using the Fungal AntiSMASH server (Weber et al. 2015).

2.3. Gene content comparisons

The protein differences between *C. fimbriata* and *C. manginecans* were identified by clustering homologous proteins into orthogroups, with the program OrthoFinder (Emms and Kelly 2015), using diamond for similarity searches and an MCL inflation parameter of 1. This program was chosen rather than a normal BLAST similarity approach, because BLAST scores are sensitive to the length of the proteins being searched (Emms and Kelly 2015). OrthoFinder transforms BLAST scores to take protein length into account and considers reciprocal best BLAST hits before constructing orthogroups. Orthogroups unique to one of the two species or orthogroups with higher copy numbers in one species were identified and further investigated.

An alternative approach to identify genes unique to a *Ceratocystis* species was performed by mapping the filtered sequence reads of *C. fimbriata* (CMW14799) to the genome of *C. manginecans* (CMW46461) and *vice versa*. Sequence data were mapped to the genome using Bowtie v.2 (Langmead and Salzberg 2012). The GFF3 annotation file and the mapping data were overlapped in BEDtools (Quinlan and Hall 2010) and genes with < 10% read coverage were identified as unique in the genome. The genomic regions surrounding unique genes were also investigated to determine the potential involvement of transposable elements or other mutational events.

2.4. Prediction of the secretome and the effector repertoire

The proteomes of all four isolates, obtained by the MAKER annotation, were used to identify different protein families involved in nutrient acquisition; including carbohydrate active enzymes (CAZymes), peptidases and lipases. The CAZymes, peptidases and lipases were predicted from the full proteome and the secretion signals of the proteins were also investigated.

CAZymes were predicted using dbCANv2 (Zhang et al. 2018). In dbCANv2, HMMER 6.0 is used to search against the dbCAN HMM (Hidden Markov Models) database (E-value < 1e-15, > 35% coverage), Diamond BLASTp analysis is performed against the CAZy database (E-value < 1e-201) and Hotpep (Homology to Peptide Pattern) is used to search for short conserved motifs in the PPR (Peptide Pattern Recognition) library (> 6 hits, > 2.6 frequency). Peptidases were predicted from the MEROPS v.11 database (Rawlings and Morton 2008) using a BLASTp search with an E-value threshold of < 1e-5. The lipase HMM database was searched using HMMER 3.1 to identify lipases present in the proteome.

The presence of a secretion signal in each protein was predicted using InterProScan v. 5.24–63 (Jones et al. 2014), including the Phobius v. 1.05 (Käll et al. 2004), and SignalP v. 4.1 algorithms (Petersen et al. 2011). Phobius considers both signal peptides and transmembrane domain predictions to provide an accurate prediction of true transmembrane domains (Käll et al. 2007). Secreted proteins were confirmed based on the presence of a secretion signal, predicted by SignalP, and with no transmembrane domain predicted by Phobius.

Candidate small secreted effector proteins were predicted from the secretome, based on a protein length < 310 amino acids (based on sizes summarised by Sperschneider et al. (2015)) and rich in cysteine (> 1%). However, not all effector proteins have a high cysteine content (Sperschneider et al. 2015) and therefore, an extensive prediction including all small secreted proteins was performed, irrespective of their cysteine content. All proteins with a functional prediction from the CAZy database were excluded as candidate effector proteins.

The abundance of effector proteins was investigated in each of the supercontigs of the Nanopore genomes to determine if some supercontigs have a higher/lower representation of effectors, relative to its overall gene content and relative to the rest of the genome. A one-tailed Fisher's exact test was performed in this regard ($\alpha = 0.05$). The distribution of effectors was also investigated in each supercontig in order to identify potential effector clusters, since it has been suggested that such a cluster could facilitate rapid modification or concerted expression of the clustered genes (Frantzeskakis et al. 2018; Meile et al. 2018). Clusters were classified as regions with two or more consecutive effector genes or if the effectors were separated by no more than one other gene. The genome-wide distribution of all effectors was determined by counting the number of effectors per 2 Kb "window", using BEDtools count (Quinlan and Hall 2010), and visualised in combination with other genomic data as described in section 2.6.

2.5. Genome synteny comparisons and identification of structural variations

To determine the overall genome synteny of the two *Ceratocystis* species, the Nanopore supercontigs for *C. fimbriata* (CMW14799) and *C. manginecans* (CMW46461) were aligned using the Nucmer package of MUMmer v. 3.23 (Kurtz et al. 2004) and average nucleotide identity was determined using Mash (Ondov et al. 2016). The resultant output was used to visualise genome synteny, using the RCircos package (Zhang et al. 2013) in R v. 3.2.5 (R Core Team 2016). The gene, TE and RIP distribution data, obtained from the BEDtools analyses, were also included in this figure (see section 2.2 and 2.6). Large chromosomal rearrangements, identified from the genome synteny maps, were further investigated in Geneious® v 7.0.6 (Biomatters Ltd.) for any genes and transposable elements that flank the chromosomal breakpoints. The chromosomal rearrangements were confirmed by aligning the assembled *C. manginecans* Illumina contigs to the *C. fimbriata* Nanopore genome; if a single contig was fragmented, it confirmed a dissimilarity in sequence order. Additionally, the raw paired-end sequence reads were mapped to the reference genome and the regions were investigated for read coverage and whether paired reads were disrupted or trimmed at these sites.

To obtain a more detailed, global view of all structural variations in the genomes, the Nanopore sequence reads of one species was aligned to the Nanopore genome of the other species, reciprocally. Alignments were performed using the long-read aligner NGMLR (Sedlazeck et al. 2018) and all the smaller inversions, translocations, insertions and deletions were identified, based on sequence alignment or split reads, using Sniffles (Sedlazeck et al. 2018). The distribution of the structural variations were determined by counting the number of occurrences of a specific structural variant per 2 Kb "window" (insertions and deletions) or 5 Kb "window" (inversions and translocations) across the genome, using BEDtools count (Quinlan and Hall 2010). The distribution was visualised, along with the TE distribution (as described in section 2.6), using Gviz (Hahne and Ivanek 2016) in R v. 3.2.5. The co-occurrence of the structural variants with TEs was further investigated by identifying all variants within a 1.5 Kb distance of a TE. This was determined using BEDtools flank and BEDtools intersect (Quinlan and Hall 2010).

2.6. Transposable elements and Repeat Induced Point (RIP) mutations

All four genomes sequenced with Illumina technology and the two Nanopore genomes (CMW14799, CMW46461), were analysed for repetitive content. The repetitive elements were identified using the REPET package v 2.2 (Flutre et al. 2011). The TEdenovo pipeline was used to predict and to classify transposable element families and consensus sequences. The consensus sequences were used as input TEs in the TEannot pipeline to annotate transposable elements in the genomes.

The repetitive content (excluding microsatellite repeats) was compared between the Illumina-sequenced genomes of all four isolates as well as between the two genomes of a single isolate where two different sequencing technologies had been used, as in the case of isolates CMW14799 and CMW46461. This made it possible to indicate whether the different sequencing technologies had influenced the detection of repetitive elements. Due to the fragmented assembly for isolates CMW17570 and CMW22563, the TE distribution and potential TE 'hotspots' on a chromosome level were only considered for the two Nanopore genomes.

Uniquely distributed TEs in each of the two Nanopore genomes were identified using BEDtools coverage analysis. The Nucmer alignment of the two genomes was overlapped with the TE annotation GFF file of one genome, and TE regions lacking alignment were identified. This was done reciprocally, in order to also consider non-syntenic regions. In addition, the genome-wide TE distribution was determined by dividing the genomes into 5 Kb "windows" and calculating the TE coverage per "window", using BEDtools (Quinlan and Hall 2010). The figures produced from these data could provide an indication as to whether the genome contains TE dense and TE sparse regions and also how this corresponds to the distribution of protein coding genes.

To confirm that the genome contained TE dense and TE sparse regions and that TEs were not randomly distributed, the inter-TE distances in each chromosome was compared with that of a randomly generated set of TEs (rTEs) equal to the number of actual TEs present in the genome. A Wilcoxon rank sum test was performed in R v. 3.2.5 (R Core Team 2016) to determine if the median of the inter-TE distances were significantly smaller than that of the inter-rTE distances. TE dense regions were further identified by means of a hypergeometric test in R v. 3.2.5 (R Core Team 2016). The genome was divided into 20 Kb windows with a 5 Kb sliding window and the TE and gene counts per window were compared to the TE and gene count outside the window. The p-values obtained per window were corrected for multiple testing, using the p.adjust function, to obtain adjusted p-values. Adjusted p-values < 0.05 confirmed TE enrichment in a specific window.

The Repeat Induced Point (RIP) profiles of all four Illumina and two Nanopore genomes were determined by analysing the genome sequences with the online web tool 'The RIPper' (http://theripper.hawk.rocks; van Wyk et al. 2019a). Large RIP-targeted regions (LRTR) are of specific interest in terms of genome evolution, hence the distribution of RIP-targeted regions > 4 Kb were investigated (Rouxel et al. 2011). The distributions of the RIP regions throughout the genomes were correlated with TE and gene distribution by incorporating it in the BEDtools analyses, based on percentage coverage per 5 Kb "window". The distribution of all genomic features (genes, effectors, TEs and RIP) was visualised using Gviz (Hahne and Ivanek 2016) in R v. 3.2.5.

In addition to The RIPper software, which uses an index-based approach to predict RIP regions, an alignment-based RIP analysis was performed on a few randomly selected TE families in both *C. fimbriata* and *C. manginecans*. This was to determine if RIP was active in TE regions not detected with the index-based approach. The genome-wide TE copies from each selected family were aligned using MAFFT online (Katoh et al. 2005) and analysed in RIPCAL (Hane 2015), using the alignment-based mode and the majority consensus comparative model.

3. Results

3.1. Genome sequences utilised

The GenBank NCBI genome sequences, constructed from Illumina sequence data, were similar in size in all three *C. manginecans* genomes (≈32 Mb) and approximately 2 Mb larger than the *C. fimbriata* genome (≈30 Mb) (Table 1). However, the genome size for a single isolate differed between the two sequencing technologies (Table 1). The Nanopore genome of *C. fimbriata* (CMW14799) was larger (32.08 Mb) than the Illumina genome (30.15 Mb) for the same isolate and differed by only 60 Kb from the *C. manginecans* (CMW46461) Nanopore genome. The Nanopore sequence data improved the genome assemblies, as the number of contigs were drastically reduced and resulted in a 20 × increase in the N50 value of CMW14799 and a 3 × increase in CMW46461 (Table 1). All four Illumina-sequenced and two Nanopore genomes were 97.4–98.3% complete, based on BUSCO analyses using the Ascomycota single-copy orthologs (Table 1).

Putative telomeric ends could be confirmed, based on the density of the telomeric repeat sequence, for either one or both sides of five supercontigs in the *C. manginecans* Nanopore genome (Fig. S1) and for seven of the *C. fimbriata* Nanopore supercontigs (Fig. S2). Although both telomeric ends were not identified in all supercontigs for both species, genome synteny alignments (section 3.6) suggested that the end of the supercontig in one species corresponded to the telomeric end of the other species. Thus, the last fragment of the telomeric repeat was likely lost during sequence assembly procedures. All nine of the supercontigs contained a high density of the repeat sequence in the middle or off-centre of the contig, which we predict to represent the centromeric region (Grady et al. 1992; Okumura-Finato et al. 2000). Collectively, these data support the assumption that each supercontig represents a separate chromosome. The high density of TEs at the ends of the contigs further supported the view that the supercontig ends represent telomeres (see section 3.7 and related figures).

3.2. Gene prediction and functional annotation

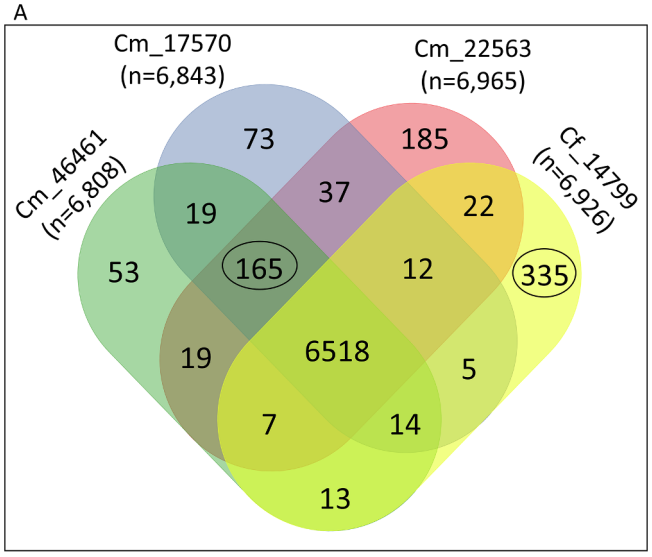
Gene prediction from the MAKER analyses identified around 7,200 genes in all four isolates, ranging from 7,210 – 7,337 (Table 1). The Illumina genome annotations that were transferred to the Nanopore genome assemblies differed slightly in gene number from that of the Illumina genomes, most likely due to the elimination of incomplete gene fragments present in the Illumina genomes. Approximately 62% (4,500) of the genes per species had a functional prediction. The predicted secretome ranged from 418 to 481 proteins among the four *Ceratocystis* isolates (Table 1). There was no significant difference in the number of genes or secreted proteins between *C. fimbriata* and *C. manginecans* and minor intraspecific variation was observed in *C. manginecans*.

3.3. Gene content comparisons

The OrthoFinder analysis identified between 6,808 and 6,965 orthogroups in *C. manginecans* (6,683 conserved in the species), and 6,926 in *C. fimbriata* (Fig. 1). The majority were conserved in both species (6,518) but 2.5% of the *C. manginecans* orthogroups were unique to the species and 5% of the *C. fimbriata* orthogroups were unique. The unique groups in *C. manginecans* and *C. fimbriata* were mostly single-copy genes and nearly 85% were hypothetical proteins (Table S1, S2). Of the functionally annotated unique genes, many aligned fully to the genome sequence (Nucmer analyses) of the other *Ceratocystis* species and differences were likely due to miss-annotation of the genes. These alignments did, however, confirm four unique genes with known functions in *C. fimbriata* (Table S1) and nine in *C. manginecans* (Table S2) as well as some hypothetical proteins. Some of the unique genes observed in both species also included effector proteins (reported in section 3.5.4). Although a large number of unique orthogroups (183) were present in isolate CMW22563 of *C. manginecans* the majority had no functional annotation. Fifty-two of these were smaller than 70 amino acids in size and the majority were at the end of small contigs, suggesting fragmented genes.

Two orthogroups had a very high copy number, as observed in a previous genome comparison study (Molano et al. 2018), which included the genome sequences of *C. cacaofunesta* and older assemblies of *C. fimbriata* and *C. manginecans*. These groups represent a family of vegetative incompatibility (HET-E) proteins and a family of 1-phosphatidylinositol phosphodiesterase (PI-PLC) proteins, respectively (Table S3). Other groups with high copy number included an effector protein family (located on supercontig 9 in *C. manginecans* and supercontig 6 and 9 in *C. fimbriata*) and a serine-threonine kinase, amongst others (Table S3). Some orthogroups were present in both species but had a higher copy number in one species. This was observed for 14 groups in *C. manginecans* including amongst others, two hypothetical protein families, four effector families and a putative amidase. In *C. fimbriata*, groups with higher copy numbers included: a calcium permeable stress-gated cation channel, two effector families, a transcriptional repressor and a serine/threonine kinase domain containing

A



B

	C. manginecans CMW46461	*C. manginecans* CMW22563	*C. manginecans* CMW17570	*C. fimbriata* CMW14799
C. manginecans CMW46461	1			
C. manginecans CMW22563	0.997	1		
C. manginecans CMW17570	0.999	0.998	1	
C. fimbriata CMW14799	0.993	0.993	0.993	1

Fig. 1. Genome similarities between *C. manginecans* and *C. fimbriata*. A) Venn diagram illustrating the number of orthogroups (n) identified in the genomes of *C. fimbriata* and three *C. manginecans* isolates. The black circles indicate the number of unique orthogroups conserved in *C. fimbriata* and *C. manginecans* respectively. B) Average nucleotide identity between the four genomes used in this study, showing a higher similarity between the *C. manginecans* genomes, compared to the *C. fimbriata* genome.

protein. Seven of the calcium permeable stress-gated cation channel genes were in close proximity on supercontig 7 amongst TE dense regions. Three of these did not fully align with the *C. manginecans* genome and are likely unique to *C. fimbriata*.

Sequence read mapping was used as an alternative approach to identify species-specific genes (genes with no read mapping from the alternate species). Based on the high overall genome similarity between the two species (section 2.5 and 3.6), unmapped regions in the reference genome could be truly missing from the other genome and not simply due to multiple polymorphisms and bad read mapping. In *C. manginecans* (CMW46461), 72 genes had no *C. fimbriata* read mapping (< 10% coverage) and, of these genes, 56 were conserved in all three *C. manginecans* genomes. This confirmed many of the unique orthogroups observed in *C. manginecans*. The unique genes were clustered to 22 unique genomic segments, ranging from 12 Kb to 291 Kb, distributed across all nine supercontigs (Table S4). No unique genes were observed in *C. fimbriata* using this approach. All the unique regions in the *C. manginecans* genome had a high density of transposable elements. Two-thirds of the unique genes could also be identified in the genomes of other *Ceratocystis* species, such as *C. eucalypticola* (GCA_001513815.1), *C. platani* (GCA_000978885.1) and *C. cacaofunesta* (GCA_002776505.1), suggesting these genes have been lost in the *C. fimbriata* genome (Table S4).

Of the 72 genes unique to *C. manginecans*, 32 had functional annotations or conserved functional domains. The unique genes included 13 putative effectors (8 conserved in all three *C. manginecans* genomes), 28 hypothetical proteins, neprilysin-11 peptidase (peptidase M13 family), putative sterigmatocystin biosynthesis peroxidase (*stcC*), DEAD/DEAH box helicase protein, NmrA-like family protein, an additional proteasome-associated protein (*ECM29*), Rio1 serine kinase, PKS-NRPS hybrid synthetase gene and multiple additional 1-phosphatidylinositol phosphodiesterase and HET-E (Vegetative incompatibility) genes. Five of the unique genes were also identified with the OrthoFinder analysis (SRSF protein kinase 3, indole-3-pyruvate monooxygenase, Ankyrin-2 protein, Ankyrin repeat-domain containing protein 50, diacylglycerol acyltransferase). The ankyrin repeat-domain is often associated with effector proteins (Mesarich et al. 2015). Several of the unique genomic intervals were at the putative sub-telomeric ends of the supercontigs (chromosomes). These regions had a high density of transposable elements and seven of the unique effectors were in these regions.

3.4. Secondary metabolites

The antiSMASH analyses allowed for the identification of nine secondary metabolite clusters in both *C. fimbriata* and *C. manginecans* (Table S5). The same nine clusters occurred in both species, which included two Type I and one Type III polyketide synthase (PKS) clusters, two non-ribosomal peptide synthase (NRPS) clusters, two clusters for terpene synthesis and two undefined clusters. The PKS and NRPS clusters corresponded with those previously identified for the two *Ceratocystis* species (Sayari et al. 2019; Sayari et al. 2018).

With the exceptions of one of the NRPS clusters (supercontig 1 of *C. manginecans*), all of the secondary metabolite clusters had identical gene content and order in *C. fimbriata* and *C. manginecans*. The NRPS cluster on supercontig 1 of *C. manginecans* was part of a ≈600 Kb subtelomeric fragment that was reciprocally translocated to supercontig 2 in *C. fimbriata* (see Section 3.6, Table S5, Fig. 2). The furthest end of this fragment had a high TE density and contained unique genes in *C. manginecans*. Two of these were part of the NRPS cluster, however these were not identified as core regulatory genes (Fig. 3). To determine whether these two unique genes are essential in the NRPS cluster, a search for homologues in other fungal secondary metabolite clusters was conducted using BLASTp analysis against the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database. No matches were found to this database. One of the Type I PKS clusters was also part of a ≈650 Kb translocated fragment moved from *C. manginecans* supercontig 9 to *C. fimbriata* supercontig 6. Such translocations could influence the expression and regulation of a gene cluster, especially if it was associated with nearby transcription factors.

3.5. Prediction of the secretome and the effector repertoire

3.5.1. Carbohydrate active enzymes

The carbohydrate active enzyme content was very similar in *C. fimbriata* and *C. manginecans*. A total of 211 – 216 enzymes (Table 2) in 81 different CAZyme families were identified but no unique or significantly expanded families were observed in either of the two species. Some CAZyme families have been associated with targeting specific substrates in plant and fungal cell wall components (Amselem et al. 2011) and CAZymes were categorised according to this annotation for comparison purposes (Table 2). Eight families of auxiliary activity enzymes (AA), known to be involved in the degradation of lignocellulose material (Morales-Cruz et al. 2015) were identified, of which AA3 and AA7 were the most abundant. The most abundant classes of CAZymes in the genomes were glycosyl hydrolases (GH) (38 families) and glycosyl transferases (GT) (26 families). The two most abundant GH and GT families were GH16 and GH18 and GT8. Three GH genes (GH13, 15, 72) also contained a carbohydrate binding module (CBM) which could facilitate their substrate specificity. The first two glycosyl hydrolases were associated with energy release and the last with fungal cell wall degradation, likely for protection against other invasive fungal species in a shared environment. The most prominent CAZy families in *C.*

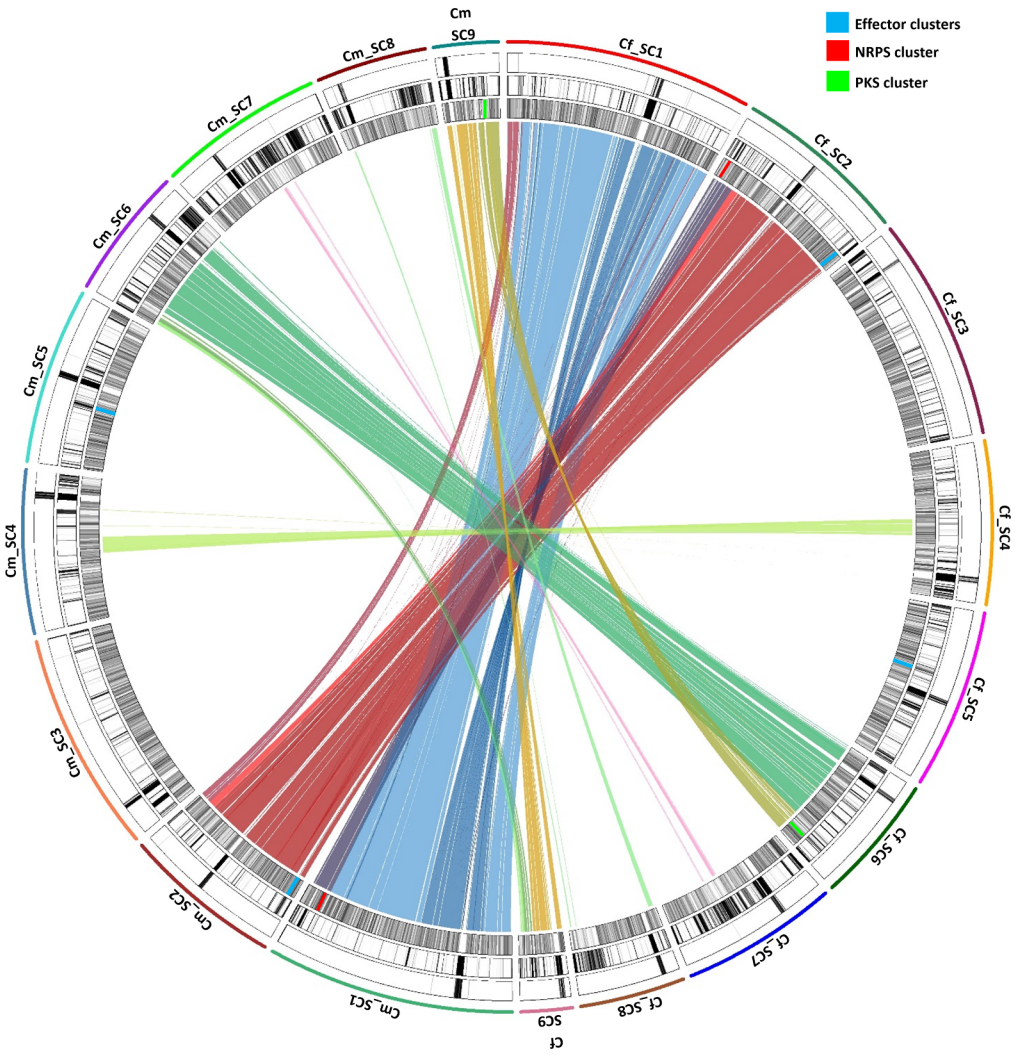


Fig. 2. Genomic inversions and translocations between the supercontigs (SC) of the *C. fimbriata* and *C. manginecans* Nanopore genomes. Corresponding supercontig numbers are syntenic in the two species and if two supercontigs were entirely syntenic, no alignment data was plotted. The coloured lines indicate regions of inversions and reciprocal translocations and each colour represents a different supercontig/chromosome. Supercontig 4, 7 and 8 each contained inverted sequences. A reciprocal translocation was observed between supercontig 1 and 2 as well as between supercontig 6 and 9 and for these contigs the full alignment is displayed for clarity purposes. The lighter blue and red colours represent inverted sequences in these contigs. The outer circles display the distribution of repeat induced point mutations (RIP) (outer track), transposable elements (TE's) (middle track) and genes (inner track). The locations of the two secondary metabolite clusters that were part of translocations (green and red), and the two large effector clusters (blue) are indicated in colour in the gene track. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fimbriata and *C. manginecans* were those families targeting plant hemicellulose and pectin (23 predicted families) and 43 families associated with fungal cell wall components.

3.5.2. Peptidases

The clades and families of peptidases identified in *C. fimbriata* and *C. manginecans* were very similar (Table S6). The peptidases identified represented 25 clans which contained between one and 12 families per clan. A total of 225–229 peptidases were identified per species (Fig. 4), five genes were present only in the *C. manginecans* isolates: a minor extracellular protease (S8, subtilisin family), ubiquitin carboxyl-terminal hydrolase creB (C19, ubiquitin peptidase family) and three peptidases from the M13 neprilysin family - a Neprilysin-11, Endothelin-converting enzyme 2 and Membrane metallo-endopeptidase-like 1. *Ceratocystis fimbriata* also contained a unique Neprilysin-11 peptidase. Proteins in the M13 neprilysin family are hypothesised to be involved in nutrient processing in bacteria; a similar function might be associated with these proteins in fungi (MEROPS; https://www.ebi.ac.uk/merops/cgi-bin/famsum?family = M13).

3.5.3. Lipases

Ceratocystis fimbriata and *C. manginecans* contained 57 and 58 lipases, respectively, that could be assigned to three super-families of the GGGX class, 12 of the GX class and one of the Y class (Table 3; Fig. 4). These classes differed in the signatures of their active sites, which is known to influence substrate specificity and are explained in more

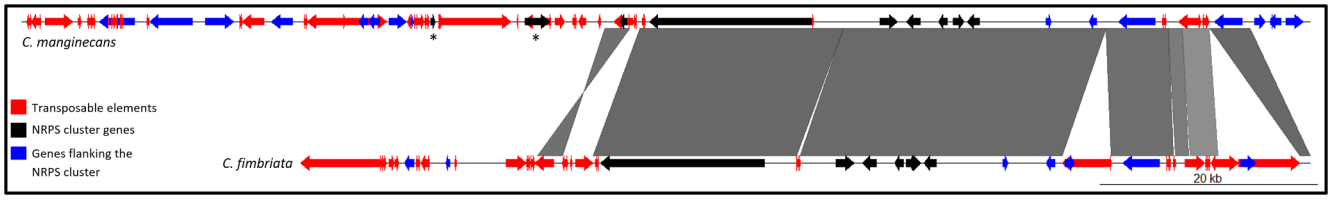


Fig. 3. A NRPS gene cluster with gene differences between *C. manginecans* and *C. fimbriata*. The sub-telomeric end of *C. manginecans* supercontig 1 (top) and *C. fimbriata* supercontig 2 (bottom) were aligned. The grey lines in the middle indicate genome synteny between the two species. The NRPS gene cluster (black arrows) as well as surrounding genes (blue) and repetitive elements (red) in this region are indicated at the top and bottom. The stars indicate two additional genes that are part of the *C. manginecans* NRPS cluster but absent in *C. fimbriata*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Summary of carbohydrate active enzymes (CAZymes) associated with the degradation of different plant and fungal cellular components, identified in the *C. fimbriata* (Cf) and C. *manginecans* (Cm) genomes. The values in brackets indicate the number of CAZymes with secretion signals.

Species		Plant cell wall					Fungal cell wall	Fungal or plant cell wall	Energy	Auxiliary activities	Unclassified		
	Total	Cellulose	Hemi-cellulose	Hemi-cellulose/Pectin	Pectin	H/HP/P					CE	GH	GT
CMW46461 (Cm)	216 (81)	3 (2)	9 (5)	7 (5)	7 (7)	1	49 (29)	15 (2)	11 (2)	29 (17)	11 (1)	16 (8)	58 (3)
CMW22563 (Cm)	211 (79)	3 (2)	9 (5)	7 (5)	7 (7)	1	47 (25)	15 (2)	11 (3)	29 (18)	11 (1)	16 (8)	55* (3)
CMW17570 (Cm)	216 (74)	3 (2)	9 (5)	7 (5)	7 (7)	1	50 (21)	15 (2)	11 (3)	29 (18)	11 (1)	16 (8)	57 (2)
CMW14799 (Cf)	216 (83)	3 (2)	9 (5)	7 (5)	7 (6)	1	49 (30)	15 (3)	11 (3)	29 (17)	11 (1)	16 (8)	58 (3)

*Difference likely due to fragmented genome assembly

depth by Borrelli and Trono (2015) and Widmann and Pleiss (2016). The lipases identified in all four genomes belonged to the same superfamilies and the same number of genes were present in each superfamily except one; *Ceratocystis fimbriata* contained one less gene in the abH03 superfamily, encoding a probable indole-3-pyruvate monooxygenase which is involved in auxin biosynthesis.

3.5.4. Effectors

The number of putative effectors identified in the four *Ceratocystis* isolates ranged from 179 to 209 (98 to 106 with > 1% cysteine content) (Fig. 4). A total of 152 effectors (83 with > 1% cysteine content) were conserved in both species, 21 (15 with > 1% cysteine) were unique to *C. manginecans* (conserved in at least two isolates) and 20 (14 with > 1% cysteine) were unique to *C. fimbriata* (Table S7a-d).

Some of the effectors (30 effectors, 20 with > 1% cysteine) had functional annotations or contained conserved Pfam domains (Table S7). Several effectors were also functionally annotated as extracellular serine-rich proteins. Effectors that are unique to a specific species were of particular interest because these could determine the host specificity of the pathogen. In *C. manginecans,* 21 effectors were unique, of which two had functional annotations - an altered inheritance of mitochondria protein 6 and an allergen 1. Most of the effectors unique in *C. fimbriata* had unknown functions and two proteins had functional domains – an alginate lyase 2 and a CVNH domain (CyanoVirin-N Homology domain). In bacteria, alginate is considered a virulence factor involved in biofilm formation, aiding bacterial fitness in stressful conditions such as desiccation. Alginate lyases in *Pseudomonas syringae pv. syringae*, have been implicated in virulence via their enzymatic activity on alginate (Preston et al. 2000). The role of alginate lyase domains in fungi is not yet known. The CVNH domain has also been observed in LysM domain-containing effectors of other fungi, although a specific role in fungal virulence hasn't been demonstrated to date (de Jonge and Thomma 2009).

The distribution of the effectors was investigated on a larger chromosomal scale in the two Nanopore genomes to identify supercontigs with a high representation of effectors and to identify potential effector clusters. All nine of the supercontigs in both species contained effectors (Table S7a-b; Fig. 5; Fig. S3). A significantly higher number of effectors was observed in supercontigs 6 (*C. fimbriata* only), 7 (both) and 9 (*C. manginecans* only) while supercontigs 3, 4 and 8 had a significantly lower number of effectors compared to the overall genome distribution (Fisher's exact; $p < 0.05$; Fig. S3). With regards to the genomic location of the effectors, nine effectors in *C. manginecans* and 13 in *C. fimbriata* were specifically near putative sub-telomeric ends (< 100 Kb from the end); these regions had a high TE density (Supplementary file 1). Six of these sub-telomeric effectors were unique in *C. manginecans* and three were unique to *C. fimbriata*. Of all the effectors in the *C. manginecans* and *C. fimbriata* Nanopore genomes 88 out of 185 and 103 out of 197, respectively, were near TEs (within 1.5 Kb) either on one side or flaked by TEs. Of specific interest is that half of the unique effectors, both in *C. manginecans* and *C. fimbriata,* were associated with TE's (Table S7a-b).

Effector clusters (≥ 2 consecutive genes) were observed in six of the *C. fimbriata* supercontigs and five of *C. manginecans* (Table S7). On average, clusters consisted of two to three effector genes. One large cluster was present on supercontig 2 consisting of 15 genes in *C. fimbriata* and 11 in *C. manginecans* (Fig. 6a). Five of the effectors in this cluster were extracellular serine-rich proteins. Another cluster was present on supercontig 5, with nine genes in *C. fimbriata* and five in *C. manginecans.* However, more candidate effectors were detected where the effector filters were less stringent (i.e. having a protein size between 310 and 350 aa or a secretion signal only detected by Phobius but not SignalP). If these additional, candidate effectors were included, *C. fimbriata* contained 15 effectors and *C. manginecans* 13 in this cluster (Fig. 6b, grey bars). The effectors in this cluster were all hypothetical proteins. Both cluster regions were in TE dense regions (hypergeometric

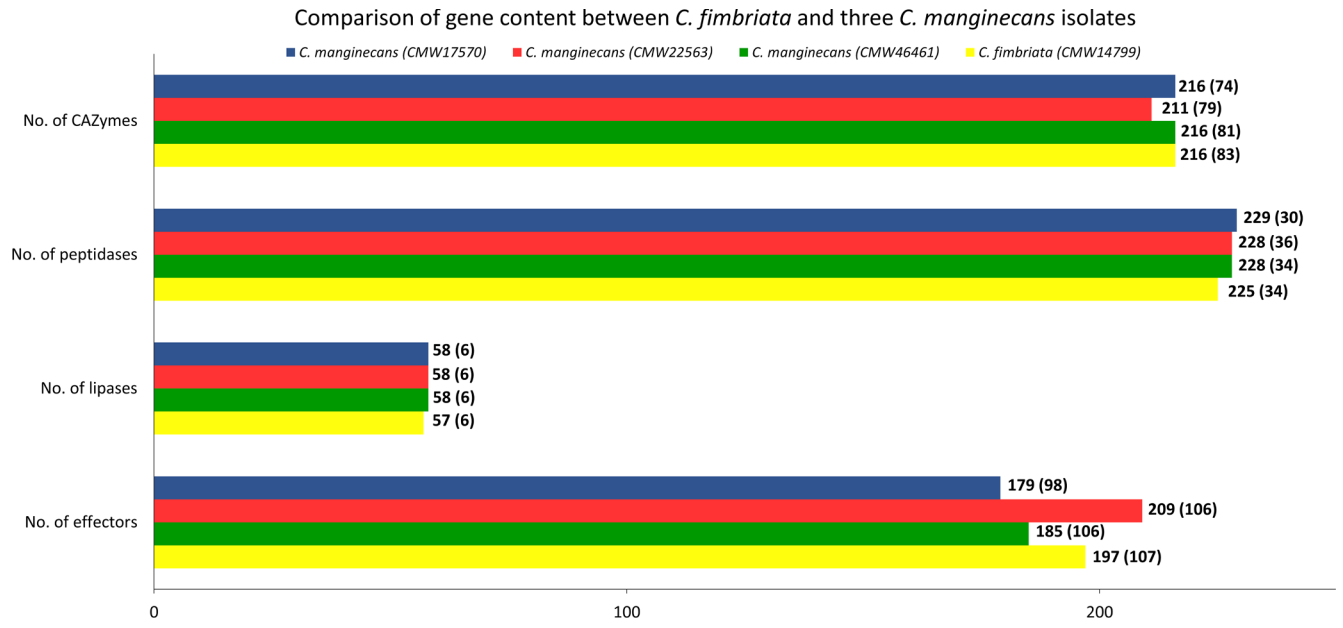


Fig. 4. Comparison of different families of secreted proteins in three *C. manginecans* and one *C. fimbriata* isolate. The numbers next to the bar indicate the number of proteins observed in a species and the number in brackets indicate those with secretion signals. For the effectors, the number indicates the total number of effectors and the number of effectors with > 1% cysteine content in brackets.

test $P < 0.05$; Supplementary file 1) and contained many TEs that were only present in one of the species, and were most likely younger, active TE families. Sequence alignment of these effector proteins indicated numerous non-synonymous mutations between the two species (ranging from two to 50 amino acid differences), but high conservation within *C. manginecans*. The effectors on supercontig 5 were more variable than those on supercontig 2 and some of the effector proteins had no secretion signal in one of the two species.

3.6. Genome synteny comparisons and identification of structural variations

Based on the Nucmer alignments, the *C. manginecans* genomes were very similar to each other with a lower level of similarity to the *C. fimbriata* genome. Using the *C. manginecans* (CMW46461) Illumina genome as reference, the genomes of the other two isolates (CMW17570 and CMW22563) of the species matched at a level of 98% and 98.5%, respectively, to the nucleotide sequence in the reference genome. The *C. fimbriata* (CMW14799) genome matched to only 93% of the *C. manginecans* reference genome. In a reciprocal search, the genome of isolate CMW46461 of *C. manginecans* matched at a level of 96% of the *C. fimbriata* reference genome. Reciprocal comparisons of the *C. fimbriata* and *C. manginecans* Nanopore genomes indicated an alignment that matched at a level of 93% to the respective genome of the other species. This similarity was also illustrated in the analysis of the average nucleotide identity, using Mash (Fig. 1B).

As a focus of this study, the effect of large structural variations on species differences were investigated in more detail. The genome synteny analysis between the two Nanopore genomes indicated a large terminal reciprocal translocation between supercontigs 1 and 2 and another between supercontigs 6 and 9 (Fig. 2). Both terminal ends of supercontig 8 were translocated to the opposite ends in *C. fimbriata*, compared to *C. manginecans*. A large inversion was present in

Table 3
Summary of lipase families present in the one *C. fimbriata* and three *C. manginecans* genomes investigated in this study.

Class[a]	Super-family	Superfamily description	CMW 14799[b]	CMW 46461[b]	CMW 22563[b]	CMW 17570[b]
GGGX	abH01	Carboxylesterases	1	1	1	1
GGGX	abH03	Candida rugosa lipase like	4 (1)	4 (1)	4 (1)	4 (1)
GGGX	abH04	Moraxella lipase 2 like	**9**	**10**	**10**	**10**
GX	abH07	Moraxella lipase 3 like	3	3	3	3
GX	abH08	Cytosolic hydrolases	9 (1)	9 (1)	9 (1)	9 (1)
GX	abH09	Microsomal Hydrolases	13	13	13	13
GX	abH13	Bacterial esterases	1 (1)	1 (1)	1 (1)	1 (1)
GX	abH14	Gastric lipases	2	2	2	2
GX	abH15	Burkholderia lipases	1	1	1	1
GX	abH19	Thioesterases	1 (1)	1 (1)	1 (1)	1 (1)
GX	abH22	Lysophospholispase	3	3	3	3
GX	abH23	Filamentous fungi lipases	3	3	3	3
Y	abH27	Dipeptidyl peptidase IV like	2	2	2	2
GX	abH31	Dienlactone hydrolases	1	1	1	1
GX	abH32	Xyalanase esterases	1 (1)	1 (1)	1 (1)	1 (1)
GX	abH34	Lysosomal protective protein	3 (1)	3 (1)	3 (1)	3 (1)
		Total	**57**	**58**	**58**	**58**

[a] Lipases are categorised into three classes based on the content of the oxyanion hole. Families in the GX class differ in the amino acid following the conserved glycine in the oxyanion hole, the GGGX class has a conserved hydrophobic residue following the glycine and in the Y class this residue is a tyrosine (Borrelli and Trono 2015).

[b] The total number of lipases are indicated with the number of secreted lipases in brackets. Numbers in bold indicate differences observed between the isolates.

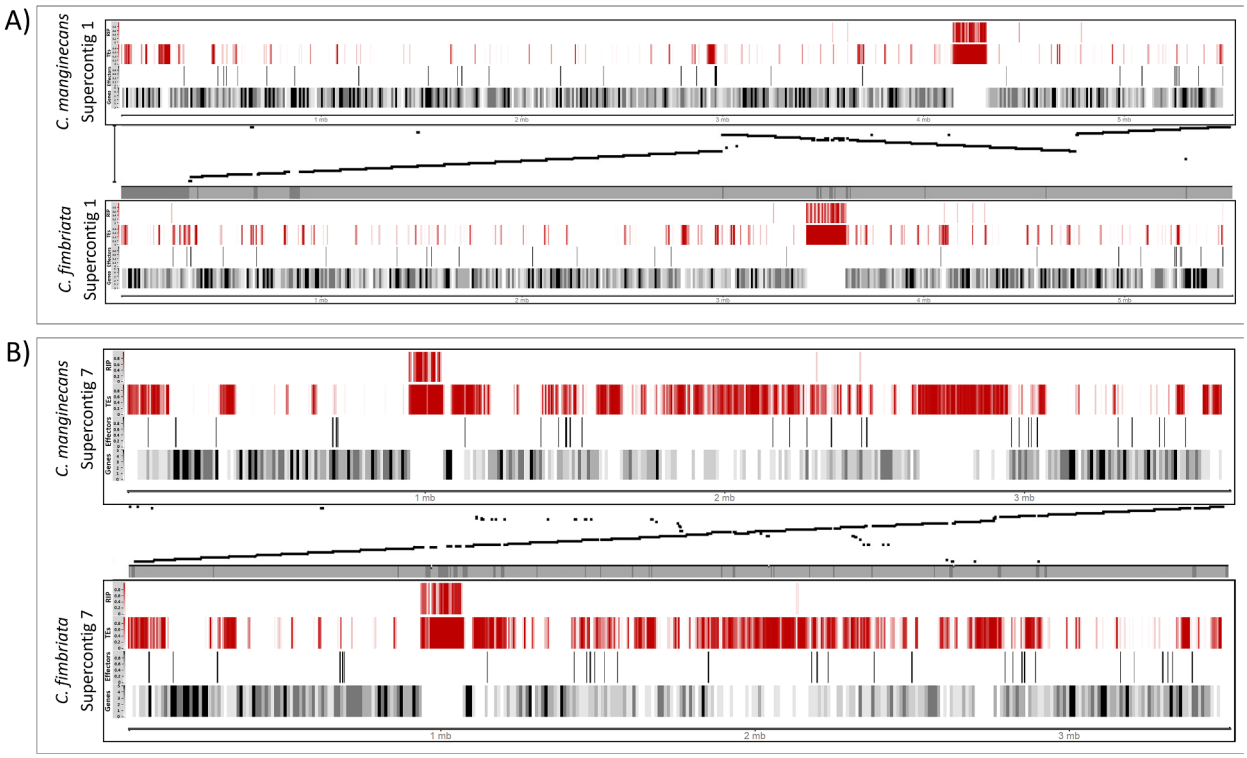


Fig. 5. Density distribution of the total gene content (bottom bar), effectors (second from bottom), transposable element content (second from top) and large RIP targeted regions (top) across two of the syntenic *C. manginecans* and *C. fimbriata* Nanopore supercontigs, SC 1 (A) and SC 7 (B). A LASTZ alignment is displayed between two supercontigs to show sequence similarity between the two species. This figure illustrates, among other things, the difference in TE density between supercontigs 1 (top two) and 7 (bottom two).

supercontig 1 (1.7 Mb) and smaller inversions on supercontigs 4 (400 Kb) and 7 (90 Kb) (Table S8; Fig. 2). Most of the genomic rearrangements were conserved in all three *C. manginecans* genomes (Table S8). The three translocations and most of the inversions had TEs at the breakpoints. Based on alignment comparisons, all of these breakpoints contained a large proportion of unique TEs that were not present in the other species at that breakpoint (Table S8). These repeat-rich regions likely facilitated chromosomal rearrangements; especially sites that contained unique TEs that could represent younger TE invasions.

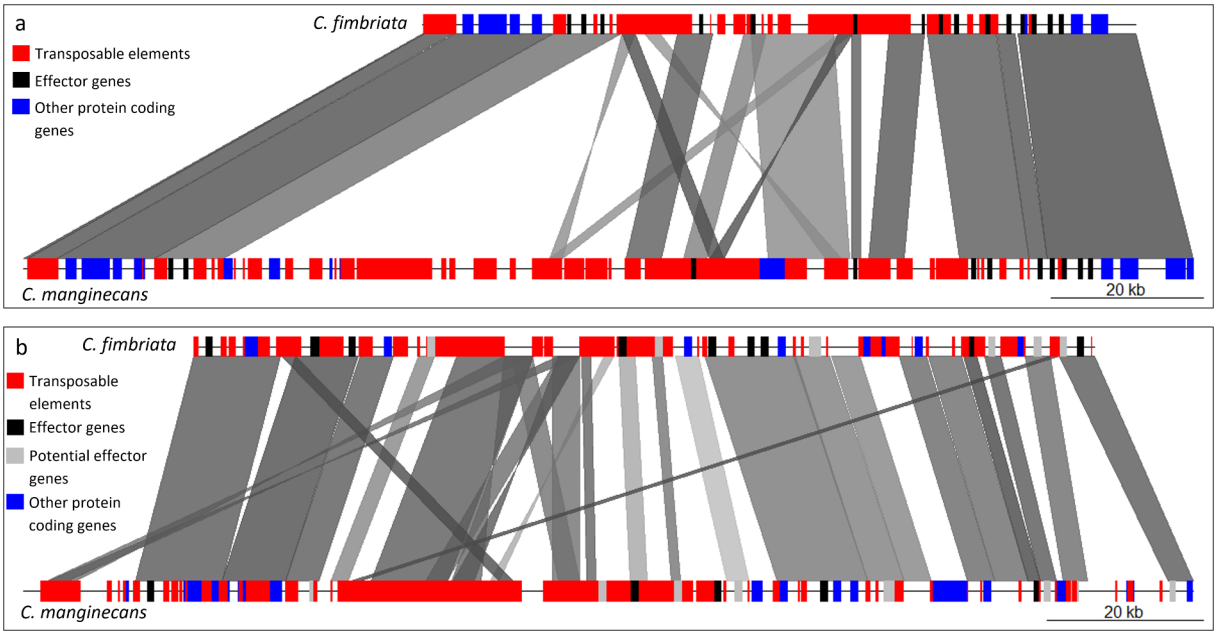


Fig. 6. TE dense effector clusters in the *C. fimbriata* and *C. manginecans* genomes. The alignment of the effector clusters between *C. fimbriata* (top) and *C. manginecans* (bottom) in (a) supercontig 2 and (b) supercontig 5 are displayed. Transposable elements are indicated in red blocks, effectors in black and other conserved genes in blue. Some genes did not pass all effector selection filters (secretion signals by SignalP and amino acid length <310) but are considered as potential effectors based on other characteristics (grey). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The genomic regions flanking the translocation and inversion breakpoints were compared between the two species to determine gene synteny around these regions. For most structural differences, the genes flanking these breakpoints were conserved in both species, but some unique genes were observed (Table S8). In both supercontigs 1 and 9, the translocation was alongside an effector protein (Table S8). In supercontig 6, *C. fimbriata* contained an additional unique effector, close to the breakpoint and another two conserved effector proteins were found further downstream. The latter were translocated to supercontig 9 of *C. manginecans*. These translocations could influence the regulation and expression of the effector proteins. In supercontig 6 of *C. manginecans* the TE dense region, next to the translocation breakpoint, contained two additional hypothetical proteins (with a metallo-beta-lactamase and NmrA-like domain) that were lost in *C. fimbriata*, possibly due to the translocation event.

Supercontig 4 contained an inversion that consisted of two consecutive regions, each inverted. The inversion site was alongside the NRPS secondary metabolite cluster, relocating all adjacent genes further downstream in the chromosome. The second (downstream) breakpoint of the inversion disrupted the coding region of a conserved hypothetical gene (containing a SPX superfamily domain) resulting in a protein that is 30 amino acids shorter in *C. fimbriata*. A 5.7-Kb transposable element was present at this site and could also have contributed to the mutation. The full-length protein was conserved in all three *C. manginecans* isolates.

A small inversion was present in supercontig 7. This inversion was TE dense and contained only 6–7 genes in each species. Of these genes, *C. manginecans* contained two unique hypothetical protein genes and *C. fimbriata* contained one unique effector. The effector was alongside the inversion breakpoint and was likely lost in *C. manginecans* during the inversion event. Supercontig 8 also contained an inversion but the gene content was conserved between the two species.

Systematic genome-wide analysis of all structural variants (insertions, deletions, inversions and translocations) was conducted to obtain an overview of all smaller and larger genomic differences. Most insertion events were observed in the *C. manginecans* genome (765) and, correspondingly, most deletion events in *C. fimbriata* (400) (Table S9). This corresponds with the slightly larger genome of *C. manginecans*. Significantly more translocation events were detected in *C. manginecans* (1148 vs 258), however, based on synteny alignments, many of these were short translocated fragments that corresponded to TE annotations. They most likely represent TEs that have relocated in *C. fimbriata* or occur in additional locations/chromosomes in *C. manginecans*. Insertion and deletion events were spread throughout the genome and did not appear to cluster at any specific sites. However, between 60 and 70% of insertions and > 85% of deletions were associated with TEs (within 1.5 Kb distance) (Figure S4). Importantly, most of the inversions and translocations were associated with TEs. This corroborates the involvement of TEs in the modification of the genomic landscape of both species, as has been observed in *Leptosphaeria* species (Grandaubert et al., 2014).

3.7. Transposable elements and Repeat Induced Point (RIP) mutations

The TE content of the four isolates (six genomes) ranged between 13 and 20% and was very similar in both *C. fimbriata* and *C. manginecans*. Differences were observed mostly between data for the different sequencing technologies. Repetitive regions of the genome are commonly miss-assembled/collapsed when short read sequencing technologies are used, compared to longer reads that can flank the entire repetitive region (Thomma et al. 2016). This was observed in the Illumina-sequenced *C. fimbriata* genome that was 1.93 Mb smaller in size and contained 7.5% less repetitive elements compared to the Nanopore genome (Table 4). Nonetheless, based on the Nanopore sequence data, 20% the *C. fimbriata* and *C. manginecans* (CMW46461) genomes consisted of TEs (Table 4) but not all regions were identical. Based on Nucmer alignments and BEDtools coverage analysis of Nanopore genomes, 17% of the TEs in each of the two species had a unique distribution and were thus absent in the corresponding genomic region of the other species. Of specific interest was that all the *C. manginecans* telomeric ends (last 16–66 Kb) and more than half of *C. fimbriata* telomeric ends (last 15–59 Kb) primarily consisted of these unique TEs. As mentioned previously, large structural variations and the two large effector clusters also had a strong representation of unique TEs.

In both *C. fimbriata* and *C. manginecans*, the majority of TEs were Class I (retrotransposons) transposable elements. These primarily consisted of large retrotransposon derivatives and a small number of terminal repeat retrotransposons in miniature and LINE retrotransposons (Table 4). Only about 1% of the TEs were Class II (DNA transposons) and these primarily consisted of terminal inverted repeat transposons and small numbers of miniature inverted repeats. Another 3% of transposable elements were unclassified.

The Nanopore genomes were further analysed for the distribution of TEs across the putative chromosomes. The TE density clearly varied between supercontigs (Fig. S3). In both species, supercontig 1 had the lowest percentage of repetitive elements (9% in *C. manginecans* and 12% in *C. fimbriata*) and supercontig 7 had the highest percentage (38% and 37%; Fig. 5). *Ceratocystis manginecans* had a lower TE content than *C. fimbriata* in supercontigs 1, 2, 3, 5 and 9 and a higher content in supercontig 6 and 8. The TEs were also not evenly distributed throughout the genomes (Wilcoxon rank sum test, $p < 0.05$) but displayed TE dense, gene sparse regions as well as regions with low TE density (hypergeometric test, $p < 0.05$; Fig. S3; Supplementary file 1).

RIP analysis using The RIPper indicated a slightly higher percentage of RIP targeted regions in all three *C. manginecans* isolates (2.6–3%) than in *C. fimbriata* (2.15%) (Table S10). Large RIP-targeted regions (LRTR) were also larger, on average, in *C. manginecans* (7,300 – 9,400 bp) than in *C. fimbriata* (6,500 bp), suggesting lower RIP activity in the latter genome. The more fragmented, Illumina sequenced genomes tended to display less LRTR's. In the Nanopore genomes, *C. manginecans* had 771 Kb of LRTR's in total and *C. fimbriata* only 388 Kb.

At a supercontig/chromosome level, only large TE dense regions were RIP targeted in the genomes based on The RIPper analysis. A few smaller sites were RIP targeted but only a single cluster of LRTRs was present in each supercontig (Fig. 5; Fig. S3). These LRTRs were 118 Kb, on average, in *C. manginecans* and 102 Kb in *C. fimbriata*. Because the RIP process can have a 'leaky' effect and target regions outside, adjacent to the repetitive regions, the sequence similarity of effectors near RIP targeted regions (based on RIPper analysis) were compared between the two species. The smallest distance between effectors and RIP targeted regions ranged from 36 Kb to 200 Kb and effector modifications were limited. In supercontig 1, a unique *C. manginecans* and *C. fimbriata* effector was within 100 Kb of a TE dense RIP targeted region.

The RIPCAL analysis suggested that the TE regions investigated are targeted by RIP, although these sites were not detected using the RIPper tool. Both the investigated TE families in *C. fimbriata* and one of the two investigated TE families in *C. manginecans* displayed RIP activity. This suggests that RIP machinery is active in both species and therefore, repetitive regions as well as their neighbouring regions could be affected by these types of mutations.

4. Discussion

Genome comparisons for *C. fimbriata,* which is specific to sweet potato, and *C. manginecans,* which has a broader host range, showed that the two species have very similar genome sequences and gene content, even though they are pathogens of different hosts. We, however, identified multiple unique genomic segments in the *C. manginecans* genome and various unique genes in each of the two species, including several effector proteins, and various effectors with polymorphisms between the species. Large structural rearrangements were also evident in the genomes of both species and these included

Table 4
Summary of transposable elements (TEs) in the *C. fimbriata* and *C. manginecans* genomes investigated in this study.

	%TE in the genome					
	C. f CMW14799 (Illumina)	*C. f* CMW14799 (Nanopore)	*C. m* CMW46461 (Illumina)	*C. m* CMW46461 (Nanopore)	*C. m* CMW22563 (Illumina)	*C. m* CMW17570 (Illumina)
Total TEs	13.38	20.85	18.63	20.58	18.28	17.8
Total Class I TEs	10.49	16.37	16.38	16.62	16.03	14.76
Large retrotransposon derivatives	10.21	15.51	16.05	16.38	15.76	14.31
Terminal repeat retrotransposons in miniature	0.22	0.03	0.32	0.15	0.27	0.34
Retrotransposon, LINE	0.004	0.83	0	0.09	0.006	0.11
Retrotransposon, SINE	0.06	0	0	0	0	0
Total Class II TEs	1.69	1.48	1.22	1.06	0.74	1.24
DNA transposon, terminal inverted repeat	1.32	1.26	0.94	0.91	0.45	0.98
Miniature inverted repeats	0.37	0.22	0.28	0.16	0.29	0.25
No category	2.68	2.99	3.19	2.9	3.4	3.2

reciprocal translocations between different chromosomes and inversions within chromosomes. The telomeric ends of the chromosomes were of particular interest because they commonly included unique genomic regions and protein coding genes. The genomes of both species had a high content of transposable elements and these appeared to have played a prominent role in the insertions/deletions of the unique genomic segments, the modification of effectors and the structural rearrangements of the chromosomes.

The repertoire of CAZymes, peptidases and lipases was very similar in *C. fimbriata* and *C. manginecans.* This suggests similar capabilities for cell wall degradation and nutrient processing. Minor differences included genes in three families of peptidases (S8, M13 and C19) and one lipase superfamily (abH04). The function of the unique genes in the S8 and M13 families have rarely been investigated in fungi and the ubiquitin carboxyl-terminal hydrolase (C19) is involved in the utilisation of carbon sources and plays a role during carbon starvation (https://www.uniprot.org). The additional C19 gene in *C. manginecans* could thus contribute to nutrient utilisation in its host environment.

Only one lipase (indole-3-pyruvate monooxygenase) was unique to *C. manginecans*. This protein forms part of the YUC pathway for the biosynthesis of the plant hormone auxin (Sardar and Kempken 2018), which induces growth and inhibits tannin biosynthesis (Hemingway and Laks, 1992). Plant hormones produced by fungi can play a role in virulence and fungal-host interactions (Chanclud and Morel 2016; Stringlis et al. 2018). The absence of any one of the genes in the auxin production pathway can impair the production of this hormone, as seen in four *Fusarium* species, where only *F. proliferatum* contained the fully functional genes to produce auxin (Tsavkelova et al. 2012). Similarly, in *C. fimbriata*, the loss of the indole-3-pyruvate monooxygenase could have de-activated a potentially functional pathway for auxin production in *Ceratocystis*. The functionality of this pathway can be confirmed with RNA expression data of the other genes involved in this pathway.

Unique genomic segments were identified in the *C. manginecans* genome, each encoding between one and seven species-specific genes. Transposable elements likely facilitated the insertion/deletion and sequence variation of these segments since the genes in each of the unique segments were interspersed with a high density of TEs. The role of TEs in the evolution of unique genomic regions has been confirmed in the *L. maculans 'brassicae'* genome, which contained AT-rich isochores with species-specific genes and it had a high TE density (Grandaubert et al. 2014). Similarly, in *Aspergillus fumigatus* and related species, the unique, species-specific, genomic clusters also had a high TE-density (Fedorova et al. 2008). The unique genes in each of the *Ceratocystis* species could be involved in their host specificity.

Fifty percent of the unique genomic regions in *C. manginecans* were located at sub-telomeric ends. Telomeres are considered dynamic regions with high frequencies of recombination and gene modification (Cuomo et al. 2007; Seidl and Thomma 2017) and they have contributed to species-specific diversity in a number of fungal species (Fedorova et al. 2008; Plissonneau et al. 2016). Since telomeres serve as mutational sites for gene addition and modification (Farman 2007; Walkowiak et al. 2016), we hypothesise that this is also the case for the two *Ceratocystis* species considered in the present study and that dynamic telomeres could play a role in its host adaptation.

Various species-specific genes were identified from the genome comparisons in this study. Nearly 33% of these have no known biological function (hypothetical proteins) and could be unique to *Ceratocystis*. However, some functionally annotated genes of interest were associated with metabolic processes, which could influence nutrient processing in the host environment and thus contribute to host specificity. In *C. manginecans*, for example, this included an additional copy of a phospholipid:diacylglycerol acyltransferase that is involved in the synthesis of triacylglycerol for fatty acid accumulation (Dahlqvist et al. 2000), and additional amidase proteins that hydrolyse amides, acting on carbon–nitrogen bonds (Fournand and Arnaud 2001). In *C. fimbriata,* this included additional copies of the quinic acid utilization activator that catabolises quinic acid as a carbon source, and glycosyltransferase HI involved in polysaccharide synthesis (https://www.uniprot.org). In addition, the high number of genes in the calcium permeable stress-gated cation channel family in *C. fimbriata* is of interest, as these channels can assist in the adaption of a fungus to environmental stress conditions, especially osmotic stress (Hou et al. 2014).

The effector repertoires predicted in *C. fimbriata* and *C. manginecans* varied both within and between species, but nearly 10% were species-specific. The species-specific effectors could be an important contributor to the host specificity of these species as the presence/absence of a single effector can determine if a pathogen is detected and if it can protect itself against host defences (de Jonge et al. 2012; Dracatos et al. 2018; Meile et al. 2018; Yuan et al. 2018). Unique effectors have been found to influence host specificity in different *forma speciales* (f. sp.) of other fungal pathogens. In *Fusarium oxysporum*, all cucurbit-infecting f. sp. contain unique effectors, as does the tomato infecting f. sp. (van Dam et al. 2016). The specificity of *Leptosphaeria maculans 'lepidii'* to *Lepidium* weeds, as opposed to *Brassica* plants, was attributed to an expanded set of effector proteins (Grandaubert et al. 2014). In the powdery mildew pathogen, *Blumeria graminis,* each f. sp. infects different grass species (wheat, barley or oats) and the presence/absence and copy number variation of secreted proteins and effectors were associated with host specificity (Frantzeskakis et al. 2018). The *in planta* expression of the species-specific *Ceratocystis* effectors should be determined in future studies to determine their possible role in host specificity.

Some of the effectors observed in both *Ceratocystis* species were of

specific interest because these had conserved annotations, known to be associated with pathogenicity in other fungi. This included six effectors with a CFEM domain (Kulkarni et al. 2003), three allergen 1 effectors (Cramer and Lawrence 2003), a NPP1 domain (necrosis inducing protein) and a LysM domain (de Jonge and Thomma 2009) effector. These effectors might not influence the host range of either of the species but could contribute to the general pathogenicity of both species.

The effector genes in the *Ceratocystis* genomes were not concentrated on specific chromosomes but were often found near TEs in the genome. Nearly half of the effectors were near to or flanked by TEs and 50% of the unique effectors were associated with TEs. Many of the effectors at TE-dense sub-telomeric regions were also unique to a species. This suggests that the two *Ceratocystis* species investigated conformed to a 'two-speed' genome model, similar to that observed for *C. albifundus* (van der Nest et al. 2019). A 'two-speed' genome is compartmentalised into gene dense and gene sparse regions, the latter having a high TE content and often containing effector gene clusters that undergo more rapid diversification (Dong et al. 2015; Raffaele and Kamoun 2012). This genomic structure has been described as an adaptive evolutionary model in many other fungi (Peng et al. 2018; Presti et al. 2015; Seidl and Thomma 2017; van Dam et al. 2016). The high prevalence of effectors in TE dense regions suggest the TE's were likely involved in effector diversification of *Ceratocystis*, as has been observed in other fungal pathogens (Faino et al. 2016; Grandaubert et al. 2014; Shirke et al. 2016; Yoshida et al. 2016).

Although many effectors were not located to genomic clusters in the two *Ceratocystis* species, two large effector clusters were observed in supercontigs 2 and 5, respectively. These were of specific interest, because they contain various non-synonymous mutations and presence/absence variations that are species-specific. These two clusters could serve as diversity hot-spots of *Ceratocystis* effectors contributing to host adaptation, similar to the TE dense effector cluster identified in *Zymoseptoria tritici* populations (Meile et al. 2018).

Fungi have developed a protective mechanism against gene family expansion and TE activity through permanent base modification of highly similar sequences, known as Repeat Induced Point mutations (RIP) (Hane et al. 2015). During this process, transition mutations convert C/G nucleotides to T/A nucleotides over successive sexual reproduction cycles. This study provides the first genome-wide investigation of RIP targeted regions in *Ceratocystis*. Although a previous study on *C. cacaofunesta* confirmed the activity of RIP in transposable elements, the chromosomal distribution and extent of RIP targeted regions was not determined (Molano et al. 2018). LRTR regions appear to be less prominent in *Ceratocystis* compared to the Sordariomycete *Fusarium* (van Wyk et al. 2019b) or the Dothideomycete *L. maculans* (Rouxel et al. 2011), that have a genome-wide distribution of LRTRs.

The LRTR regions in *Ceratocystis* were located on a single region of 102–118 Kb, either at the centre or off-centre in each supercontig, which corresponds to the size of fungal centromeres (Smith et al. 2012). A strong association has been observed between centromeres and RIP in, for example, *N. crassa* (Cambareri et al. 1998; Smith et al. 2012), *L. maculans* (Attard et al. 2005) and *Fusarium* species (Smith et al. 2012), suggesting these LRTR regions in *Ceratocystis* represent centromeres. The fact that RIP activity was also detected in additional TEs using RIPCAL suggests that smaller RIP target sites in, and around TEs, could be a source of mutation of the effector proteins found near TEs, as observed in other fungi (Fudal et al. 2009).

Two large reciprocal terminal translocations and various large inversion events have occurred in the respective genomes of *C. fimbriata* and *C. manginecans*. Some inversions were not conserved in all three *C. manginecans* isolates investigated, suggesting those events occurred only in some populations of *C. manginecans*. The breakpoints of two of the inversions (supercontigs 1 and 4) contained TEs only in *C. fimbriata*, which could have facilitated the events in this genome, and likely occurred after its divergence from *C. manginecans*. Similarly, some of the translocations and inversions contained many unique TE's in *C. fimbriata*, suggesting the TEs expanded in this species and facilitated the modification events. Chromosomal rearrangements can skew recombination; as is evidenced by the segregation distortion in three chromosomes of the interspecific cross between *C. fimbriata* and *C. manginecans* (Fourie et al. 2019). This could have contributed to the diversification of the chromosomes in the two species, similar to the numerous chromosomal polymorphisms observed in populations of *Z. tritici* (Plissonneau et al. 2016). Genome inversions and translocations can lead to mutations, gene loss and change in gene regulation (de Jonge et al. 2013; Plissonneau et al. 2016; Stukenbrock et al. 2010). In the *Ceratocystis* genomes, these events contributed to the loss of an effector and other hypothetical protein genes as well as disruption of a gene's coding sequence. Numerous smaller structural variants were also detected across the genomes and these still require detailed investigation to determine their impact on gene modification, gene expression and speciation. The extent to which the structural variations contribute to the host specificity of the species will need to be confirmed by functional studies, such as gene knock-outs.

Host specificity in fungi can be determined at different levels of plant interaction. When two related fungal species are specifically pathogenic to different hosts, the phylogenetic distance of the host species can be indicative of whether the pathogen has adapted to overcome the physical barriers of the host, the PAMP-triggered immune responses (PTI) or effector-triggered immune responses (Dracatos et al. 2018; Schulze-Lefert and Panstruga 2011). Because effector repertoires and corresponding host receptors are rapidly evolving, when the two plant hosts being compared are very distantly related, the more conserved resistance factors such as physical barriers and PAMP receptors are expected to play a more prominent role in host resistance. The results of this study, however, suggest that despite the large evolutionary distance between the *Ceratocystis* hosts investigated, host specificity is strongly influenced by the effector repertoires of the species and the effector triggered susceptibility in the host, with limited influence from the initial stages of host entrance, nutrient acquisition or PAMP molecules of the fungi. This is supported by the variation observed in the sequence and content of the effector catalogues of *C. fimbriata* and *C. manginecans*.

5. Conclusions

This study provided the first in-depth genome comparison between two closely related *Ceratocystis* species, pathogenic to very distantly related plant hosts. Furthermore, it provided insights into the gene- and repetitive content and structural differences between these pathogens. Various candidate genes and processes, likely associated with host specificity, were identified and these include species-specific enzymes associated with metabolism and nutrient processing, as well as effector proteins. These candidate genes provide a foundation for further functional studies. It is important to recognise that the comparisons in this study were based on gene sequences but other differences between the two species, such as epigenetic modifications and gene expression, could also result in phenotypic differences. These factors should also be considered when seeking to determine the basis of host specificity.

Transposable elements are amongst the genomic factors considered in this study that deserve further investigation in *Ceratocystis* species. These factors appear to have played a prominent role in many of the evolutionary processes of *Ceratocystis*, leading to the divergence and host adaptation in *C. fimbriata* and *C. manginecans*. The breakpoints of all the large chromosomal structural modifications and most of the smaller structural variants could all be linked to TE regions, as could unique genomic clusters in the genomes. TEs also appear to have contributed to the diversification of genes, especially effectors, at the subtelomeric ends as well as in the effector clusters observed in supercontigs 2 and 5. Most of the aggressively pathogenic *Ceratocystis* species, apart from *C. fimbriata*, are adapted to tree hosts (Marin-Felix et al. 2017). This could suggest that evolutionary events such as gene losses,

chromosomal inversions and translocations as well as effector modifications in *C. fimbriata* provided an adaptation that has enabled it to infect only sweet potato while mitigating detection by host defence responses.

Acknowledgements

Financial support for this research was provided by members of the Tree Protection Cooperative Programme, based at the Forestry and Agricultural Biotechnology Institute at the University of Pretoria, the National Research Foundation, South Africa (ZA) provided a student bursary and additional funding (Grant UID: 106309, 89619 and 95875) and the University of Pretoria a travel grant. The authors thank Prof. Yves van de Peer, Department of Plant Biotechnology and Bioinformatics, Ghent University, for hosting an exchange visit at his research facilities and providing the appropriate infrastructure and resources for data analysis. We also thank Dr. Nanette Christie, University of Pretoria, for assistance with statistical analyses of the transposable element data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fgb.2020.103433.

References

Al Adawi, A.O., Barnes, I., Khan, I.A., Al Subhi, A.M., Al Jahwari, A.A., Deadman, M.L., Wingfield, B.D., Wingfield, M.J., 2013. *Ceratocystis manginecans* associated with a serious wilt disease of two native legume trees in Oman and Pakistan. Australas. Plant. Pathol. 42, 179–193.

Amselem, J., Cuomo, C.A., van Kan, J.A.L., Viaud, M., Benito, E.P., Couloux, A., Coutinho, P.M., de Vries, R.P., Dyer, P.S., Fillinger, S., Fournier, E., Gout, L., Hahn, M., Kohn, L., Lapalu, N., Plummer, K.M., Pradier, J.-M., Quévillon, E., Sharon, A., Simon, A., ten Have, A., Tudzynski, B., Tudzynski, P., Wincker, P., Andrew, M., Anthouard, V., Beever, R.E., Beffa, R., Benoit, I., Bouzid, O., Brault, B., Chen, Z., Choquer, M., Collémare, J., Cotton, P., Danchin, E.G., Da Silva, C., Gautier, A., Giraud, C., Giraud, T., Gonzalez, C., Grossetete, S., Güldener, U., Henrissat, B., Howlett, B.J., Kodira, C., Kretschmer, M., Lappartient, A., Leroch, M., Levis, C., Mauceli, E., Neuvéglise, C., Oeser, B., Pearson, M., Poulain, J., Poussereau, N., Quesneville, H., Rascle, C., Schumacher, J., Ségurens, B., Sexton, A., Silva, E., Sirven, C., Soanes, D.M., Talbot, N.J., Templeton, M., Yandava, C., Yarden, O., Zeng, Q., Rollins, J.A., Lebrun, M.-H., Dickman, M., 2011. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLOS Genet. 7, e1002230.

Araujo, L., Silva Bispo, W.M., Cacique, I.S., Cruz, M.F.A., Rodrigues, F.A., 2014. Histopathological aspects of mango resistance to the infection process of *Ceratocystis fimbriata*. Plant Pathol. 63, 1282–1295.

Attard, A., Gout, L., Ross, S., Parlange, F., Cattolico, L., Balesdent, M.-H., Rouxel, T., 2005. Truncated and RIP-degenerated copies of the LTR retrotransposon Pholy are clustered in a pericentromeric region of the *Leptosphaeria maculans* genome. Fungal Genet. Biol. 42, 30–41.

Baker, C.J., Harrington, T.C., Krauss, U., Alfenas, A.C., 2003. Genetic variability and host specialization in the Latin American clade of *Ceratocystis fimbriata*. Phytopathol. 93, 1274–1284.

Barnes, I., Fourie, A., Wingfield, M.J., Harrington, T.C., McNew, D.L., Sugiyama, L.S., Luiz, B.C., Heller, W.P., Keith, L.M., 2018. New *Ceratocystis* species associated with rapid death of *Metrosideros polymorpha* in Hawai'i. Persoonia. 40, 154–181.

Borah, N., Albarouki, E., Schirawski, J., 2018. Comparative methods for molecular determination of host-specificity factors in plant-pathogenic fungi. Int. J. Mol. Sci. 19, 863.

Borrelli, G., Trono, D., 2015. Recombinant lipases and phospholipases and their use as biocatalysts for industrial applications. Int. J. Mol. Sci. 16, 20774.

Buiate, E.A.S., Xavier, K.V., Moore, N., Torres, M.F., Farman, M.L., Schardl, C.L., Vaillancourt, L.J., 2017. A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species *Colletotrichum graminicola* and *Colletotrichum sublineola*. BMC Genomics. 18, 67.

Cambareri, E.B., Aisner, R., Carbon, J., 1998. Structure of the chromosome VII centromere region in *Neurospora crassa*: Degenerate transposons and simple repeats. Mol. Cell. Biol. 18, 5465.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., Yandell, M., 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18, 188–196.

Chanclud, E., Morel, J.B., 2016. Plant hormones: a fungal point of view. Mol. Plant Pathol. 17, 1289–1297.

Chen, S., van Wyk, M., Roux, J., Wingfield, M.J., Xie, Y., Zhou, X., 2013. Taxonomy and pathogenicity of *Ceratocystis* species on *Eucalyptus* trees in South China, including *C. chinaeucensis* sp. nov. Fungal Divers. 58, 267–279.

Cramer, R.A., Lawrence, C.B., 2003. Cloning of a gene encoding an Alt a 1 isoallergen differentially expressed by the necrotrophic fungus *Alternaria brassicicola* during *Arabidopsis* infection. Appl. Environ. Microbiol. 69, 2361–2364.

Cuomo, C.A., Güldener, U., Xu, J.-R., Trail, F., Turgeon, B.G., Di Pietro, A., Walton, J.D., Ma, L.-J., Baker, S.E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y.-L., DeCaprio, D., Gale, L.R., Gnerre, S., Goswami, R.S., Hammond-Kosack, K., Harris, L.J., Hilburn, K., Kennell, J.C., Kroken, S., Magnuson, J.K., Mannhaupt, G., Mauceli, E., Mewes, H.-W., Mitterbauer, R., Muehlbauer, G., Münsterkötter, M., Nelson, D., Donnell, K., Ouellet, T., Qi, W., Quesneville, H., Roncero, M.I.G., Seong, K.-Y., Tetko, I.V., Urban, M., Waalwijk, C., Ward, T.J., Yao, J., Birren, B.W., Kistler, H.C., 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science. 317, 1400.

Da Silva, A.C., De Oliveira Silva, F.M., Milagre, J.C., Omena-Garcia, R.P., Abreu, M.C., Mafia, R.G., Nunes-Nesi, A., Alfenas, A.C., 2018. Eucalypt plants are physiologically and metabolically affected by infection with *Ceratocystis fimbriata*. Plant Physiol. Biochem. 123, 170–179.

Da Silva Galdino, T.V., Kumar, S., Oliveira, L.S.S., Alfenas, A.C., Neven, L.G., Al-Sadi, A.M., Picanço, M.C., 2016. Mapping global potential risk of mango sudden decline disease caused by *Ceratocystis fimbriata*. PLOS ONE. 11, e0159450.

Dahlqvist, A., Stahl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., Ronne, H., Stymne, S., 2000. Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. Proc. Natl. Acad. Sci. U. S. A. 97, 6487–6492.

de Jonge, R., Bolton, M.D., Kombrink, A., van den Berg, G.C., Yadeta, K.A., Thomma, B.P., 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 23, 1271–1282.

de Jonge, R., van Esse, P.H., Maruthachalam, K., Bolton, M.D., Santhanam, P., Saber, M.K., Zhang, Z., Usami, T., Lievens, B., Subbarao, K.V., Thomma, B.P.H.J., 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc. Natl. Acad. Sci. 109, 5110–5115.

de Jonge, R., Thomma, B.P.H.J., 2009. Fungal LysM effectors: extinguishers of host immunity? Trends Microbiol. 17, 151–157.

Dong, S., Raffaele, S., Kamoun, S., 2015. The two-speed genomes of filamentous pathogens: waltz with plants. Curr. Opin. Genet. Dev. 35, 57–65.

Dracatos, P.M., Haghdoust, R., Singh, D., Park, R.F., 2018. Exploring and exploiting the boundaries of host specificity using the cereal rust and mildew models. New Phytol. 218, 453–462.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157.

Faino, L., Seidl, M.F., Shi-Kunne, X., Pauper, M., van den Berg, G.C.M., Wittenberg, A.H.J., Thomma, B.P.H.J., 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res. 26, 1091–1100.

Farman, M.L., 2007. Telomeres in the rice blast fungus *Magnaporthe oryzae* : the world of the end as we know it. FEMS Microbiol. Lett. 273, 125–132.

Fedorova, N.D., Khaldi, N., Joardar, V.S., Maiti, R., Amedeo, P., Anderson, M.J., Crabtree, J., Silva, J.C., Badger, J.H., Albarraq, A., Angiuoli, S., Bussey, H., Bowyer, P., Cotty, P.J., Dyer, P.S., Egan, A., Galens, K., Fraser-Liggett, C.M., Haas, B.J., Inman, J.M., Kent, R., Lemieux, S., Malavazi, I., Orvis, J., Roemer, T., Ronning, C.M., Sundaram, J.P., Sutton, G., Turner, G., Venter, J.C., White, O.R., Whitty, B.R., Youngman, P., Wolfe, K.H., Goldman, G.H., Wortman, J.R., Jiang, B., Denning, D.W., Nierman, W.C., 2008. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. PLOS Genet. 4, e1000046.

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

Fourie, A., van der Nest, M.A., De Vos, L., Wingfield, M.J., Wingfield, B.D., Barnes, I., 2019. QTL mapping of mycelial growth and aggressiveness to distinct hosts in *Ceratocystis* pathogens. Fungal Genet. Biol. 131, 103242. https://doi.org/10.1016/j.fgb.2019.103242.

Fourie, A., Wingfield, M.J., Wingfield, B.D., Barnes, I., 2014. Molecular markers delimit cryptic species in *Ceratocystis sensu stricto*. Mycol. Prog. 14, 1–18.

Fourie, A., Wingfield, M.J., Wingfield, B.D., Thu, P.Q., Barnes, I., 2016. A possible centre of diversity in South East Asia for the tree pathogen, *Ceratocystis manginecans*. Infect. Genet. Evol. 41, 73–83.

Fourie, A., Wingfield, M.J., Wingfield, B.D., van der Nest, M.A., Loots, M.T., Barnes, I., 2018. Inheritance of phenotypic traits in the progeny of a *Ceratocystis* interspecific cross. Fungal Biol. 122, 717–729.

Fournand, D., Arnaud, A., 2001. Aliphatic and enantioselective amidases: from hydrolysis to acyl transfer activity. J. Appl. Microbiol. 91, 381–393.

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., Spanu, P.D., Maekawa, T., Schulze-Lefert, P., Panstruga, R., 2018. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. BMC Genomics. 19, 381.

Fudal, I., Ross, S., Brun, H., Besnard, A.-L., Ermel, M., Kuhn, M.-L., Balesdent, M.-H., Rouxel, T., 2009. Repeat-Induced Point Mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*. Mol. Plant Microbe Interact. 22, 932–941.

Grady, D.L., Ratliff, R.L., Robinson, D.L., McCanlies, E.C., Meyne, J., Moyzis, R.K., 1992. Highly conserved repetitive DNA sequences are present at human centromeres. Proc. Natl. Acad. Sci. 89, 1695–1699.

Grandaubert, J., Lowe, R.G., Soyer, J.L., Schoch, C.L., Van de Wouw, A.P., Fudal, I., Robbertse, B., Lapalu, N., Links, M.G., Ollivier, B., Linglin, J., Barbe, V., Mangenot, S., Cruaud, C., Borhan, H., Howlett, B.J., Balesdent, M.-H., Rouxel, T., 2014. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans-Leptosphaeria biglobosa* species complex of fungal pathogens.

BMC Genomics. 15, 891.
Gremme, G., Steinbiss, S., Kurtz, S., 2013. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM Trans. Comput. Biol. Bioinform. 10, 645–656.
Hahne, F., Ivanek, R., 2016. Statistical Genomics: Methods and Protocols. In: Mathé, E., Davis, S. (Eds.), Visualizing genomic data using Gviz and Bioconductor. Springer, New York, New York, pp. 335–351.
Hall, B., DeRego, T., Geib, S., GAG: the Genome Annotation Generator (Version 1.0) [Software]. Available from http://genomeannotation.github.io/GAG., 2014.
Halsted, B.D., Fairchild, D.G., 1891. Sweet-Potato black rot. J. Mycol. 7, 1–11.
Hane, J.K., 2015. Calculating RIP mutation in fungal genomes using RIPCAL. In: van den Berg, M., Maruthachalam, K. (Eds.), Genetic transformation systems in fungi. Springer, Cham, pp. 69–78.
Hane, J.K., Williams, A.H., Taranto, A.P., Solomon, P.S., Oliver, R.P., 2015. Repeat-induced point mutation: a fungal-specific, endogenous mutagenesis process. In: In: van den Berg, M., Maruthachalam, K. (Eds.), Genetic Transformation Systems in Fungi Volume 2. Springer, Cham, pp. 55–68.
Hardie, M., Akhmad, N., Mohammed, C., Mendham, D., Corkrey, R., Gafur, A., Siregar, S., 2018. Role of site in the mortality and production of *Acacia mangium* plantations in Indonesia. South. For. 80, 37–50.
Harrington, T.C., 2004. Ceratocystis fimbriata. CABI Crop protection Compendium. CAB International, Wallingford, UK.
Harrington, T.C., Huang, Q., Ferreira, M.A., Alfenas, A.C., 2014a. Genetic analyses trace the Yunnan, China population of *Ceratocystis fimbriata* on pomegranate and taro to populations on *Eucalyptus* in Brazil. Plant Dis. 99, 106–111.
Harrington, T.C., Kazmi, M.R., Al-Sadi, A.M., Ismail, S.I., 2014b. Intraspecific and intragenomic variability of ITS rDNA sequences reveals taxonomic problems in *Ceratocystis fimbirata sensu stricto*. Mycologia. 106, 224–242.
Harwood, C., Nambiar, E., 2014. Productivity of acacia and eucalypt plantations in Southeast Asia. 2. Trends and variations. International Forestry Review. 16, 249–260.
Hemingway, R.W., Laks, P.E. (Eds.), 1992. Plant polyphenols: synthesis, properties, significance. Plenum Press, New York, pp. 111–131.
Hou, C., Tian, W., Kleist, T., He, K., Garcia, V., Bai, F., Hao, Y., Luan, S., Li, L., 2014. DUF221 proteins are a family of osmosensitive calcium-permeable cation channels conserved across eukaryotes. Cell Res. 24, 632–635.
Huang, Q., Zhu, Y.Y., Chen, H.R., Wang, Y.Y., Liu, Y.L., Lu, W.J., Ruan, X.Y., 2003. First report of pomegranate wilt caused by *Ceratocystis fimbriata* in Yunnan. China. Plant Dis. 87, 1150.
Inoue, Y., Vy, T.T.P., Yoshida, K., Asano, H., Mitsuoka, C., Asuke, S., Anh, V.L., Cumagun, C.J.R., Chuma, I., Terauchi, R., Kato, K., Mitchell, T., Valent, B., Farman, M., Tosa, Y., 2017. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. Science. 357, 80–83.
Jones, J.D.G., Dangl, J.L., 2006. The plant immune system. Nature. 444, 323–329.
Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30, 1236–1240.
Käll, L., Krogh, A., Sonnhammer, E.L.L., 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338, 1027–1036.
Käll, L., Krogh, A., Sonnhammer, E.L.L., 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res. 35, W429–W432.
Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518.
King, B., Waxman, K., Nenni, N., Walker, L., Bergstrom, G., Gibson, D., 2011. Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. Biotechnol. Biofuels. 4, 1–14.
Kulkarni, R.D., Kelkar, H.S., Dean, R.A., 2003. An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. Trends Biochem. Sci. 28, 118–121.
Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. Genome Biol. 5, R12.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. methods. 9, 357–359.
Li, Q., Harrington, T.C., McNew, D., Li, J., Huang, Q., Somasekhara, Y.M., Alfenas, A.C., 2016. Genetic bottlenecks for two populations of *Ceratocystis fimbriata* on sweet potato and pomegranate in China. Plant Dis. 100, 2266–2274.
Liu, F.F., Barnes, I., Roux, J., Wingfield, M.J., Chen, S.F., 2018. Molecular phylogenetics and microsatellite analysis reveal a new pathogenic *Ceratocystis* species in the Asian-Australian clade. Plant Pathol. 67, 1097–1113.
Marincowitz, S., Barnes, I., de Beer, Z.W., Wingfield, M.J., 2020. Epitypification of *Ceratocystis fimbriata*. Fungal Systemat. Evol. https://doi.org/10.3114/fuse.2020.06.14.
Marin-Felix, Y., Groenewald, J.Z., Cai, L., Chen, Q., Marincowitz, S., Barnes, I., Bensch, K., Braun, U., Camporesi, E., Damm, U., De Beer, Z.W., Dissanayake, A., Edwards, J., Giraldo, A., Hernández-Restrepo, M., Hyde, K.D., Jayawardena, R.S., Lombard, L., Luangsa-ard, J., McTaggart, A.R., Rossman, A.Y., Sandoval-Denis, M., Shen, M., Shivas, R.G., Tan, Y.P., van der Linde, E.J., Wingfield, M.J., Wood, A.R., Zhang, J.Q., Zhang, Y., Crous, P.W., 2017. Genera of phytopathogenic fungi: GOPHY 1. Stud. Mycol. 86, 99–216.
Meile, L., Croll, D., Brunner, P.C., Plissonneau, C., Hartmann, F.E., McDonald, B.A., Sánchez-Vallet, A., 2018. A fungal avirulence factor encoded in a highly plastic genomic region triggers partial resistance to septoria tritici blotch. New Phytol. 219, 1048–1061.
Mesarich, C.H., Bowen, J.K., Hamiaux, C., Templeton, M.D., 2015. Repeat-containing protein effectors of plant-associated organisms. Front. Plant Sci. 6, 872.
Molano, E.P.L., Cabrera, O.G., Jose, J., Do Nascimento, L.C., Carazzolle, M.F., Teixeira, P.J.P.L., Alvarez, J.C., Tiburcio, R.A., Tokimatu Filho, P.M., De Lima, G.M.A., Guido, R.V.C., Corrêa, T.L.R., Leme, A.F.P., Mieczkowski, P., Pereira, G.A.G., 2018. *Ceratocystis cacaofunesta* genome analysis reveals a large expansion of extracellular phosphatidylinositol-specific phospholipase-C genes (PI-PLC). BMC Genomics. 19, 58.
Morales-Cruz, A., Amrine, K.C.H., Blanco-Ulate, B., Lawrence, D.P., Travadon, R., Rolshausen, P.E., Baumgartner, K., Cantu, D., 2015. Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. BMC Genomics. 16, 469.
Okumura-Finato, A., Varella-Garcia, M., Tajara, E.H., Taddei, V.A., Morielle-Versute, E., 2000. Intrachromosomal distribution of telomeric repeats in *Eumops glaucinus* and *Eumops perotis* (Molossidae, Chiroptera). Chromosome Res. 8, 563–569.
Oliveira, L.S.S., Harrington, T.C., Ferreira, M.A., Damacena, M.B., Al-Sadi, A.M., Al-Mahmooli, I.H.S., Alfenas, A.C., 2015. Species or genotypes? Reassessment of four recently described species of the Ceratocystis wilt pathogen, *Ceratocystis fimbriata*, on *Mangifera indica*. Phytopathol. 105, 1229–1244.
Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., Phillippy, A.M., 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 17, 132.
Paul, N.C., Nam, S.-S., Kachroo, A., Kim, Y.-H., Yang, J.-W., 2018. Characterization and pathogenicity of sweet potato (*Ipomoea batatas*) black rot caused by *Ceratocystis fimbriata* in Korea. Eur. J. Plant Pathol. 152, 1–8.
Peng, Z., Oliveira-Garcia, E., Lin, G., Hu, Y., Dalby, M., Migeon, P., Tang, H., Farman, M., Cook, D., White, F.F., Valent, B., Liu, S., 2018. Effector gene reshuffling involves dispensable mini-chromosomes in the wheat blast fungus. bioRxiv.
Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods. 8, 785.
Plissonneau, C., Stürchler, A., Croll, D., 2016. The evolution of orphan regions in genomes of a fungal pathogen of wheat. mBio. 7, e01231–16.
Presti, L.L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., Kahmann, R., 2015. Fungal effectors and plant susceptibility. Annu. Rev. Plant Biol. 66, 513–545.
Preston, L.A., Wong, T.Y., Bender, C.L., Schiller, N.L., 2000. Characterization of alginate lyase from *Pseudomonas syringae* pv. syringae. J. Bacteriol. 182, 6268–6271.
Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 841–842.
R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
Raffaele, S., Kamoun, S., 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat. Rev. Microbiol. 10, 417.
Rawlings, N.D., Morton, F.R., 2008. The MEROPS batch BLAST: A tool to detect peptidases and their non-peptidase homologues in a genome. Biochimie. 90, 243–259.
Restrepo, S., Tabima, J.F., Mideros, M.F., Grünwald, N.J., Matute, D.R., 2014. Speciation in fungal and oomycete plant pathogens. Annu. Rev. Phytopathol. 52, 289–316.
Roux, J., WIngfield, M.J., Fourie, A., Noeth, K., Barnes, I., 2020. Ceratocystis wilt on Eucalyptus: first record from South Africa. Southern Forests: J. For. Sci. https://doi.org/10.2989/20702620.2019.1686687.
Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., van de Wouw, A.P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A.J., Ciuffetti, L.M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S.B., Gout, L., Glaser, N., Linglin, J., Kema, G.H.J., Lapalu, N., Lawrence, C.B., May, K., Meyer, M., Ollivier, B., Poulain, J., Schoch, C.L., Simon, A., Spatafora, J.W., Stachowiak, A., Turgeon, B.G., Tyler, B.M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville, H., Oliver, R.P., Wincker, P., Balesdent, M.-H., Howlett, B.J., 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat. Commun. 2, 202.
Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F.E., Soyer, J.L., Tellier, A., Croll, D., 2018. The genome biology of effector gene evolution in filamentous plant pathogens. Annu. Rev. Phytopathol. 56, 21–40.
Sardar, P., Kempken, F., 2018. Characterization of indole-3-pyruvic acid pathway-mediated biosynthesis of auxin in *Neurospora crassa*. PLOS ONE. 13, e0192293.
Sayari, M., van der Nest, M.A., Steenkamp, E.T., Soal, N.C., Wilken, P.M., Wingfield, B.D., 2019. Distribution and evolution of nonribosomal peptide synthase gene clusters in the *Ceratocystidaceae*. Genes. 10, 328.
Sayari, M., Steenkamp, E.T., van der Nest, M.A., Wingfield, B.D., 2018. Diversity and evolution of polyketide biosynthesis gene clusters in the *Ceratocystidaceae*. Fungal Biology. 122, 856–866.
Schulze-Lefert, P., Panstruga, R., 2011. A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends Plant Sci. 16, 117–125.
Scruggs, A.C., Basaiah, T., Adams, M.L., Quesada-Ocampo, L.M., 2017. Genetic diversity, fungicide sensitivity, and host resistance to *Ceratocystis fimbriata* infecting sweet potato in North Carolina. Plant Dis. 101, 994–1001.
Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., Michael, C., Schatz, M.C., 2018. Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods. 15, 461–468.
Seidl, M.F., Thomma, B.P.H.J., 2017. Transposable elements direct the coevolution between plants and microbes. Trends Genet. 33, 842–851.
Shirke, M.D., Mahesh, H.B., Gowda, M., 2016. Genome-wide comparison of *Magnaporthe* species reveals a host-specific pattern of secretory proteins and transposable elements. PLoS ONE. 11, e0162458.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015.

BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31, 3210–3212.
Singh, A.K., Mukhopadhyay, M., 2012. Overview of fungal lipase: A review. Appl. Biochem. Biotechnol. 166, 486–520.
Smith, K.M., Galazka, J.M., Phatale, P.A., Connolly, L.R., Freitag, M., 2012. Centromeres of filamentous fungi. Chromosome Res. 20, 635–656.
Sperschneider, J., Dodds, P.N., Gardiner, D.M., Manners, J.M., Singh, K.B., Taylor, J.M., 2015. Advances and challenges in computational prediction of effectors from plant pathogenic fungi. PLOS Pathog. 11, e1004806.
Stringlis, I.A., Zhang, H., Pieterse, C.M.J., Bolton, M.D., de Jonge, R., 2018. Microbial small molecules – weapons of plant subversion. Nat. Prod. Rep. 35, 410–433.
Stukenbrock, E.H., 2013. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. New Phytol. 199, 895–907.
Stukenbrock, E.H., Jørgensen, F.G., Zala, M., Hansen, T.T., McDonald, B.A., Schierup, M.H., 2010. Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola*. PLOS Genet. 6, e1001189.
Tarigan, M., Roux, J., van Wyk, M., Tjahjono, B., Wingfield, M.J., 2011. A new wilt and die-back disease of *Acacia mangium* associated with *Ceratocystis manginecans* and *C. acaciivora sp*. nov. in Indonesia. SA. J. Bot. 77, 292–304.
Tate, R., Hall, B., DeRego, T., Geib, S., 2014. Annie: the ANNotation Information Extractor (Version 1.0) [Software]. Available from http://genomeannotation.github.io/annie.
Thomma, B.P.H.J., Seidl, M.F., Shi-Kunne, X., Cook, D.E., Bolton, M.D., van Kan, J.A.L., Faino, L., 2016. Mind the gap; seven reasons to close fragmented genome assemblies. Fungal Genet. Biol. 90, 24–30.
Trang, T.T., Eyles, A., Davies, N., Glen, M., Ratkowsky, D., Mohammed, C., 2018. Screening for host responses in *Acacia* to a canker and wilt pathogen. *Ceratocystis manginecans*. Forest Pathol. 48, e12390.
Tsavkelova, E., Oeser, B., Oren-Young, L., Israeli, M., Sasson, Y., Tudzynski, B., Sharon, A., 2012. Identification and functional characterization of indole-3-acetamide-mediated IAA biosynthesis in plant-associated *Fusarium* species. Fungal Genet. Biol. 49, 48–57.
Urban, M., Irvine, A., Cuzick, A., Hammond-Kosack, K., 2015. Using the pathogen-host interactions database (PHI-base) to investigate plant pathogen genomes and genes implicated in virulence. Front. Plant Sci. 6, 605.
Valdetaro, D.C.O.F., Harrington, T.C., Oliveira, L.S.S., Guimarães, L.M.S., McNew, D.L., Pimenta, L.V.A., Gonçalves, R.C., Schurt, D.A., Alfenas, A.C., 2019. A host specialized form of *Ceratocystis fimbriata* causes seed and seedling blight on native *Carapa guianensis* (andiroba) in Amazonian rainforests. Fungal Biol. 123, 170–182.
van Dam, P., Fokkens, L., Schmidt, S.M., Linmans, J.H.J., Kistler, H.C., Ma, L.-J., Rep, M., 2016. Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*. Environ. Microbiol. 18, 4087–4102.
van der Nest, M.A., Bihon, W., de Vos, L., Naidoo, K., Roodt, D., Rubagotti, E., Slippers, B., Steenkamp, E.T., Wilken, P.M., Wilson, A., Wingfield, M.J., Wingfield, B.D., 2014. IMA Genome-F 2: *Ceratocystis manginecans*, *Ceratocystis moniliformis*. *Diplodia sapinea*. IMA Fungus. 5, 135–140.
van der Nest, M.A., Steenkamp, E.T., Roodt, D., Soal, N.C., Palmer, M., Chan, W.-Y., Wilken, P.M., Duong, T.A., Naidoo, K., Santana, Q.C., Trollip, C., de Vos, L., van Wyk, S., McTaggart, A.R., Wingfield, M.J., Wingfield, B.D., 2019. Genomic analysis of the aggressive tree pathogen *Ceratocystis albifundus*. Fungal Biol. https://doi.org/10.1016/j.funbio.2019.02.002.
van Wyk, M., Wingfield, B.D., Wingfield, M.J., 2013. *Ceratocystis* species in the *Ceratocystis fimbriata* complex. In: Seifert, K.A., de Beer, Z.W., Wingfield, M.J. (Eds.), Ophiostomatoid Fungi: Expanding Frontiers. CBS biodiversity Series, CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands, pp. 65–73.
van Wyk, S., Harrison, C., Wingfield, B.D., De Vos, L., van der Merwe, N.A., Steenkamp, E.T., 2019a. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. PeerJ. 7, e7447.
van Wyk, S., Wingfield, B.D., de Vos, L., Santana, Q.C., van der Merwe, N.A., Steenkamp, E.T., 2018. Multiple independent origins for a subtelomeric locus associated with growth rate in *Fusarium circinatum*. IMA Fungus. 9, 27–61.
van Wyk, S., Wingfield, B.D., de Vos, L., van der Merwe, N.A., Santana, Q.C., Steenkamp, E.T., 2019b. Repeat-induced point mutations drive divergence between *Fusarium circinatum* and its close relatives. Pathogens. 8, 1–21.
Walkowiak, S., Rowland, O., Rodrigue, N., Subramaniam, R., 2016. Whole genome sequencing and comparative genomics of closely related Fusarium head blight fungi: *Fusarium graminearum, F. meridionale* and *F. asiaticum*. BMC Genomics. 17, 1014.
Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., Medema, M.H., 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43, W237–W243.
Widmann, M., Pleiss, J., 2016. Sequence, structure, function: What we learn from analyzing protein families. In: Svendsen, A. (Ed.), Understanding enzymes: Function, design, engineering, and analysis. Pan Stanford Publishing, pp. 321–350.
Wingfield, B.D., Fourie, A., Simpson, M., Bushula-Njah, V., Aylward, J., Barnes, I., Coetzee, M., Dreyer, L., Duong, T.A., Geiser, D., Roets, F., Steenkamp, E.T., van der Nest, M.A., van Heerden, C., Wingfield, M.J., 2019. IMA Genome-F 11 draft genome sequences of *Fusarium xylarioides, Teratosphaeria gauchensis* and *T. zuluensis* and genome annotation for *Ceratocystis fimbriata*. IMA Fungus 10. https://doi.org/10.1186/s43008-019-0013-7.
Yoshida, K., Saunders, D.G.O., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., Inoue, Y., Chuma, I., Tosa, Y., Cano, L.M., Kamoun, S., Terauchi, R., 2016. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics. 17, 1–18.
Yuan, C., Wang, M., Skinner, D.Z., See, D.R., Xia, C., Guo, X., Chen, X., 2018. Inheritance of virulence, construction of a linkage map, and mapping dominant virulence genes in *Puccinia striiformis f. sp. tritici* through characterization of a sexual population with genotyping-by-sequencing. Phytopathol. 108, 133–141.
Zhang, H., Meltzer, P., Davis, S., 2013. RCircos: an R package for Circos 2D track plots. BMC Bioinformatics. 14, 244.
Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y., Yin, Y., 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46, W95–W101.