QTL mapping of mycelial growth and aggressiveness to distinct hosts in *Ceratocystis* pathogens

Arista Fourie[a], Magriet A. van der Nest[b], Lieschen de Vos[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, South Africa, 0002

[b]Biotechnology Platform, Agricultural Research Council, Private Bag X05, Onderstepoort, 0110, South Africa, 0002

*Corresponding author: email: irene.barnes@fabi.up.ac.za

Highlights

- The first linkage map for a *Ceratocystis* species was constructed.
- Two genes potentially involved in mycelial growth rate of *Ceratocystis* was identified.
- Candidate genes associated with host aggressiveness of two *Ceratocystis* species were identified.
- Four of the linkage groups of *C. fimbriata* and *C. manginecans* have reciprocal terminal translocations.

Abstract

Some species of *Ceratocystis* display strong host specificity, such as *C. fimbriata sensu stricto* that is restricted to sweet potato (*Ipomoea batatas*) as host. In contrast, the closely related *C. manginecans,* infects *Acacia mangium* and *Mangifera indica* but is not pathogenic to *I. batatas*. Despite the economic importance of these fungi, knowledge regarding the genetic factors that influence their pathogenicity and host specificity is limited. A recent inheritance study, based on an interspecific cross between *C. fimbriata* and *C. manginecans* and 70 F_1 progeny, confirmed that traits such as mycelial growth rate, spore production and aggressiveness on *A. mangium* and *I. batatas* are regulated by multiple genes. In the present study, a quantitative trait locus (QTL) analysis was performed to determine the genomic loci associated with these traits. All 70 progeny isolates were genotyped with SNP markers and a linkage map was constructed. The map contained 467 SNPs, distributed across nine linkage groups, with a total length of 1,203 cM. Using the progeny genotypes and phenotypes, one QTL was identified on the linkage map for mycelial growth rate, one for aggressiveness to *A. mangium* and two for aggressiveness to *I. batatas* ($P < 0.05$). Two candidate genes, likely associated with mycelial growth rate, were identified in the QTL region. The three QTLs associated with aggressiveness to different hosts contained candidate genes involved in protein processing, detoxification and regions with effector genes and high transposable element density. The results provide a foundation for studies considering the function of genes regulating various quantitative traits in *Ceratocystis*.

Keywords: Ion Torrent, SNP markers, Sordariomycetes, reciprocal translocation, host specificity

1. Introduction

Fungal pathogens contribute significantly to yield loss in agriculture and forestry (Dean et al. 2012; Wingfield et al. 2015). Some fungal pathogens have broad host ranges and have the ability to expand their host ranges while others remain highly host-specific. Disease management of economically important plants requires an understanding of the infection biology of fungal pathogens as well as the molecular pathways that determine their host specificity (Gao et al. 2016; Vleeshouwers and Oliver 2014).

During the first stages of infection, fungi can encounter challenges such as rigid physical barriers, difficulties in acquiring nutrients from the host and the Pathogen-Associated Molecular Pattern (PAMP) immune response (PTI), triggered by highly conserved fungal target molecules (Divon and Fluhr 2007; Niks and Marcel 2009; Schulze-Lefert and Panstruga 2011). Fungal pathogens can, as a protection response, release small secreted effector proteins and secondary metabolites such as toxins which can induce an effector-triggered immune response (ETI) in the host (Jones and Dangl 2006). The host range of fungal pathogens can be determined during the initial stages of host entry, or later, during ETI (Buiate et al. 2017; Deng et al. 2017; Möller and Stukenbrock 2017).

Most life history traits that make fungi successful pathogens are regulated by multiple genes and biochemical pathways and are thus considered quantitative traits. Examples include melanisation (Lendenmann et al. 2014), growth rate (de Vos et al. 2011; Heinzelmann et al. 2017), fungicide sensitivity (Lendenmann et al. 2015), spore production, aggressiveness to hosts (Lind et al. 2007; Stewart et al. 2016) and host specificity (Borah et al. 2018; Cozijnsen et al. 2000; Lind et al. 2007). The genes associated with a quantitative trait can be positioned at multiple locations in a fungal genome, as in the case of the effector gene compartments of *Zymoseptoria tritici* (Meile et al. 2018; Möller and Stukenbrock 2017). Alternatively, they can be clustered and regulated coherently, such as pathogenicity genes on lineage specific chromosomes of *Fusarium oxysporum* species (Ma et al. 2013). The genomic locations associated with quantitative traits can be determined by investigating the inheritance of the trait, either in progeny of controlled crosses or in sexually reproducing natural populations, and subsequently correlating the phenotypes to the genotypes of the individuals (Borah et al. 2018; Plissonneau et al. 2017).

Quantitative studies, by means of inheritance studies, have identified candidate pathogenicity genes in fungal pathogens of agronomically valuable plants (Lendenmann et al. 2015; Mideros et al. 2017) but ironically, few studies have focused on tree pathogens (Lind et al. 2007); even less so on ascomycetes (de Vos et al. 2011). The ascomycete genus *Ceratocystis* includes 41 species (Barnes et al. 2018; Holland et al. 2019; Liu et al. 2018; Marin-Felix et al. 2017) and most species in the Latin

American clade (LAC) of the genus are pathogens of agronomic and tree crops. These species provide interesting models to fill the knowledge gap on fungal tree pathogens.

Two closely related pathogenic species, *C. fimbriata sensu stricto* (*s.s.*) and *C. manginecans,* are of particular interest when considering infection biology and host specificity. *Ceratocystis fimbriata sensu stricto* (*s.s.*), which has recently been informally designated as *C. fimbriata* f. sp. *ipomoea,* only infects sweet potato (*Ipomoea batatas*) (Valdetaro et al. 2019), whereas *C. manginecans* infects various tree species but not *I. batatas* (Baker et al. 2003; Fourie et al. 2018). *Ceratocystis fimbriata* s.s. causes a post-harvest black rot disease which has had a great impact for sweet potato growers (Harter and Weimer 1929; Scruggs et al. 2017). This clonal pathogen on sweet potato has likely been distributed along with sweet potato tubers (Li et al. 2016) from the USA (Halsted 1890) to New Zealand, Australia and numerous Asian countries (Harter and Weimer 1929; Li et al. 2016; Paul et al. 2018).

Ceratocystis manginecans is an aggressive pathogen of a broad range of tree hosts including *Acacia mangium* (Tarigan et al. 2011), *Mangifera indica* (Oliveira et al. 2015; van Wyk et al. 2007), *Punica granatum* (Harrington et al. 2014; Huang et al. 2003), *Eucalyptus* spp. (Chen et al. 2013) as well as other leguminous trees (Al Adawi et al. 2013). It infects trees through open wounds to then occupy parenchymal cells and xylem transport vessels (Araujo et al. 2014; Da Silva et al. 2018). This reduces water transport that subsequently leads to leaf wilt and eventually tree death (Al-Sadi et al. 2010; Araujo et al. 2014; Da Silva et al. 2018). *Ceratocystis manginecans* (considered by some researchers to be synonymous with *C. eucalypticola* (Harrington et al. 2015)), has had huge impacts on the mango industry of Brazil, Oman and Pakistan; causing up to 60% losses in some countries (Da Silva Galdino et al. 2016). In South East Asia, *C. manginecans* infection has changed the face of forestry where most of the *A. mangium* plantations have been replaced with *Eucalyptus spp.* (Harwood and Nambiar 2014). Despite the economic impact of *C. fimbriata* and *C. manginecans*, limited information is available on the molecular factors and pathways involved in determining the aggressiveness and host range of these fungal species.

An interspecific cross was recently performed between the species *C. fimbriata* (CMW14799, NC, USA) and *C. manginecans* (CMW46461, Malaysia) to investigate the inheritance of traits associated with host specificity and aggressiveness in 70 F_1 progeny (Fourie et al. 2018). Traits that displayed quantitative characteristics included lesion length induced when the pathogen was inoculated onto *A. mangium* and *I. batatas*, growth rate on malt extract agar (MEA) and *A. mangium* extract agar (AEA) and asexual spore production (Fourie et al. 2018). Based on the inheritance results, it was evident that a small number of genes likely regulate aggressiveness to distinct hosts and spore production and that

a greater number of genes are involved in fungal growth rate. The genomic loci associated with these traits have, however, not been identified.

The overall aim of this study was to determine the quantitative trait loci associated with the phenotypic traits previously investigated in *C. fimbriata* and *C. manginecans*. As a first step, a genetic linkage map was constructed by genotyping the 70 progeny isolates obtained from the interspecific cross. Subsequently, the linkage map and available phenotypic data were used to perform QTL analyses on the investigated traits including aggressiveness on a specific host, mycelial growth rate and spore production. The data were further used to investigate the inheritance of nuclear and mitochondrial genomic material in the two *Ceratocystis* species. The QTL loci identified in this study made it possible to identify candidate genes associated with aggressiveness to distinct hosts in *Ceratocystis* species. These can then be targeted in future functional studies to further an understanding of essential pathways in *Ceratocystis* infection.

2. Material and methods

2.1. Phenotyping of the F_1 progeny population

The mating experiment and phenotypic results for the parents and 70 progeny isolates were presented in detail in a previous study (Fourie et al. 2018). Briefly, cultures representing MAT1 and MAT2 self-sterile mating types were produced for both the *C. fimbriata* (CMW14799) and *C. manginecans* (CMW46461) isolates and crosses were performed reciprocally, using the one species as the MAT1 parent and the other species as the MAT2 parent, and *vice versa*. Two to three ascospore drops were collected from the apices of the ascomata formed in the hybrid zone of both crosses and these were dispersed on an agar plate to obtain single ascospores. Ascospores were randomly selected and 35 confirmed hybrids were chosen from each cross to make up the F_1 population (70 in total).

The phenotypic traits investigated in the inheritance study included mycelial growth rate on malt extract agar (MEA) and on *A. mangium* extract agar (AEA), conidial spore production and lesion length induced on *A. mangium* and on *I. batatas,* after inoculation (Fourie et al. 2018). Most of the progeny grew significantly faster than the *C. manginecans* parent on both media types and faster than the *C. fimbriata* parent on the AEA medium. Most of the progeny either produced the same amount or fewer conidia than the parental isolates and only two produced more conidia than the parents. Sixteen of the progeny isolates were pathogenic to *I. batatas,* none were more aggressive than the *C. fimbriata* parent but four of the isolates were as aggressive as the parent. On *A. mangium,* none of the progeny isolates were as aggressive as the *C. manginecans* parent but 11 produced a significant lesion and were considered pathogenic. Three isolates were pathogenic on both *I. batatas* and *A. mangium*.

There were, however, no significant associations found between aggressiveness, growth rate or mating type.

All the phenotypic data discussed above, and published by Fourie et al. (2018), were used for QTL analyses in the present study. In the current study however, analyses of lesion size on *A. mangium* were conducted using the average lesion size per isolate (Fourie et al. 2018) as well as using an average where outliers (values more than 1.5 times smaller or larger than the inter-quartile range) were removed from the data. This was found to improve the accuracy of the results.

2.2. Reference genomes of the parental isolates

The first version of the genome sequences of the two parental isolates, *C. fimbriata* (CMW14799) and *C. manginecans* (CMW46461) were generated using Illumina HiSeq technology. The annotated genome of *C. fimbriata* was obtained from the National Center for Biotechnology Information database (NCBI accession number APWK03000000; (Fourie et al. 2019) and the *C. manginecans* genome was generated in this study. DNA was extracted using a phenol/chloroform method (Goodwin et al. 1992). The DNA sequence libraries for Illumina sequencing of isolate CMW46461 consisted of a 300 bp and 500 bp paired-end and a 5 Kb mate-pair library. The Illumina sequence reads were assembled using VelvetOptimiser v. 2.2.5 (Zerbino and Birney 2008), SSPACE v.2 (Boetzer et al. 2011) and GapFiller v.1.10 (Boetzer and Pirovano 2012). These genomes served as a reference for SNP genotyping in the progeny isolates.

Gene prediction was performed using the MAKER annotation pipeline (Cantarel et al. 2008; Holt and Yandell 2011), by including gene models from the gene predictors SNAP (Korf 2004), AUGUSTUS (Stanke et al. 2004) and GeneMark (Besemer and Borodovsky 2005) as well as FgeneSH (Solovyev et al. 2006), optimised with *Neurospora crassa* gene models. The *C. fimbriata* annotation, available at NCBI, was performed with RNA evidence for the gene models (Fourie et al. 2019), thus its annotation was also included to train MAKER for the annotation of *C. manginecans*. Functional annotation was performed using BLASTp analysis against the Swissprot database. InterproScan v. 5.24 (Jones et al. 2014) was used to identify conserved protein domains (Pfam) as well as secretion signals (SignalP v. 4.1 and Phobius v. 1.05). Effectors were classified based on a protein sequence length < 310 amino acids and >1% cysteine content, the presence of a secretion signal (SignalP) and the absence of transmembrane domains (Phobius)(Sperschneider et al. 2015).

Additional genome sequences of both isolates, *C. manginecans* (CMW46461) and *C. fimbriata* (CMW14799), became available during a later stage of this study. This included MinION sequencing data (Oxford Nanopore technologies), assembled in combination with the Illumina HiSeq sequence reads, providing a reduced number of scaffolds with high confidence base calls (Mendeley Data). The

C. fimbriata genome consisted of 16 scaffolds and the *C. manginecans* genome consisted of 26 scaffolds. The SNP sites and gene annotation from the Illumina genomes were transferred to the MinION genomes, using Geneious Prime (https://www.geneious.com; (Kearse et al. 2012). These two genomes were used for further downstream analyses. To determine the completeness of both the Illumina and MinION genome assemblies, the presence of a conserved set of Ascomycete single-copy orthologs were determined using the BUSCO pipeline (Benchmarking Universal Single-Copy Orthologs; (Simão et al. 2015).

To investigate repetitive elements in the QTL regions, transposable elements (TEs) were predicted in both the *C. fimbriata* and *C. manginecans* genomes using the REPET package (Flutre et al. 2011). Both the TEdenovo and TEannot pipelines of the package were used. TE families were first identified to obtain consensus sequences (TEdenovo) and these were then used as the input for the final TE prediction and annotation (TEannot).

2.3. Whole genome re-sequencing of the progeny isolates

2.3.1. DNA extraction

In order to genotype all 70 progeny isolates (Fourie et al. 2018) a low-coverage whole genome re-sequencing approach was used. To obtain genomic DNA, isolates were grown at 22 °C on 2% malt extract agar (MEA) medium, containing 150mg/L streptomycin (Sigma-Aldrich, Germany) and 1mg/L thiamine (Sigma-Aldrich, Germany) with autoclaved cellophane membrane (Bio-Rad Laboratories, Johannesburg, South Africa) placed on the agar surface. After 21 days, mycelium was scraped from the surface of the cellophane sheets and freeze-dried. Using metal beads, the material was macerated in a mixer mill type MM 301 Retsch® tissue grinder for 3 minutes at 30 oscillations/second (Retsch, Germany). DNA extraction was performed using a phenol/chloroform method (Barnes et al. 2001) with minor modifications. No initial heating step was performed and an RNaseA treatment step, including 1 hour incubation at 37 °C and 1 hour at room temperature, was included after DNA extraction. RNase was removed with an over-night ethanol precipitation step. The DNA pellet was re-suspended in distilled, autoclaved water.

2.3.2. DNA quality control

DNA quality for the 70 progeny isolates was evaluated using a Nanodrop ND_1000 (Nanodrop, Wilmington, DE) and quantity was determined using a Qubit® 2.0 Fluorometer (ThermoFisher Scientific). To confirm that all isolates were clean fungal cultures, the 5.8S rRNA and internal transcribed spacer regions 1 and 2 was amplified using the ITS1 and ITS4 primers (White et al. 1990). Where isolates had the ITS sequence of both parents, and thus mixed chromatogram peaks, two additional gene regions were amplified to confirm that these isolates are haploid homokaryons. These

included Beta-tubulin 1 (βt1a and βt1b; (Glass and Donaldson 1995) and *MS204* (MS204F.ceratoB and MS204R.ceratoB; Fourie et al. 2014). Reaction and cycler conditions were identical to those previously described (Fourie et al. 2014). The sequences of the amplicons were determined with Sanger sequencing using the ABI BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied BioSystems, Thermo Fisher, California, USA). The sequence of each isolate and each gene region was then aligned against previously generated sequences of *C. fimbriata* and *C. manginecans*. As a standard procedure, to confirm the absence of bacterial contamination, the 16S rRNA region was PCR amplified using the DNA from the isolates and a positive control bacterial sample, using primers 27F and 1492R (Beukes et al. 2013; de Long 1992). Where no band was observed, the sample was considered free of bacteria.

2.3.3. DNA library preparation, whole genome sequencing and data processing

The progeny isolates were sequenced with a low coverage whole genome sequencing approach, using Ion Torrent sequencing technology. Library building for each DNA sample was done at the DNA Ion Torrent sequencing facility, University of Pretoria, with the Ion Xpress™ Plus kit using the Ion Xpress™ Plus gDNA fragment library preparation protocol (Thermo Fisher Scientific, Carlsbad, USA). Size selection was done using a 2% E-gel (Invitrogen). Ion Express Barcode adaptors 1-19 (Thermo Fisher Scientific, Carlsbad, USA) were used to uniquely label the DNA of each individual, multiplexed on a single chip. Template amplification for the Ion Proton was done using the Ion OneTouch™ 2 System (OT2) Ion PI™ Hi-Q™ OT2 200 Kit (Thermo Fisher Scientific, Carlsbad, USA). DNA sequencing was performed on the Ion Proton apparatus, at the Central Analytical Facilities at the University of Stellenbosch, using the Ion PI™ Hi-Q™ Sequencing 200 Kit (Thermo Fisher Scientific, Carlsbad, USA) and the Ion PI™ Chip Kit v3. Up to 19 individuals were sequenced per chip.

Adapters were removed from the sequence reads and data were separated, based on barcodes, by the Ion Torrent sequencing facility. The quality of the sequence reads of each library was determined using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Low quality reads (phred score <20) and reads shorter than 80 bp were removed from the data using Trimmomatic (Bolger et al. 2014). Some samples displayed an overrepresented k-mer at the 3' end of the reads and 10-20 bp were trimmed from the 3' edges using seqtk (https://github.com/lh3/seqtk). After read trimming, reads were reanalysed, using FastQC, to confirm the quality of the final reads.

2.4. SNP genotyping of parents and progeny isolates

The genome of *C. manginecans* isolate CMW46461, sequenced using Illumina technology, was used as the reference sequence for SNP identification. The quality trimmed Illumina sequence reads of the *C. fimbriata s.s.* parental isolate (CMW14799) and Ion Torrent reads of all 70 progeny isolates were

separately mapped to the reference sequence (creating a bam file), using bowtie v. 2.2.6 (Langmead and Salzberg 2012). Read group information, indicating the sample identity and sequencing run of each sample, were added to the bam files, using Picardtools v. 2.6.0 (https://github.com/broadinstitute/picard). The same software was used to mark duplicate reads, obtained from PCR amplification bias, and to realign regions around indels in the mapping file.

SNP identification was performed using Haplotypecaller in the Genome Analysis ToolKit (GATK) v. 3.7.0 (van der Auwera et al. 2013). A minimum base score of 20 was selected, a ploidy of 1 and a maximum of 1 alternate alleles. SNPs were called using the BP_RESOLUTION option that provides data for every basepair in the progeny genome and not only the variable SNP sites. This provided confidence in the progeny SNPs that were similar to the *C. fimbriata* alternate alleles as well as those that were identical to the reference genome alleles. The individual SNP files of each isolate were combined to a single SNP file, using the GenotypeGVCFs tool. SNPs and indels were separated and only SNPs were used in downstream analyses. SNPs were filtered for quality by removing those SNPs with a quality depth (QD) < 20.0, a Fisher Strand bias (FS) > 40.0 and a mapping quality (MQ) < 40.0. Vcftools v. 0.1.15 was used to remove novel SNPs identified in the progeny but not present in the parents, by setting the parental SNP sites as reference positions (Danecek et al. 2011). A final genotype filter was applied to the data using maxNOCALLfraction in GATK, removing all SNP sites absent in more than 10% of the population. Data were converted to binary code using a custom python script (Mendeley Data).

The SNP calling were quality checked, and, to help reduce the number of SNP markers, some of those in high linkage disequilibrium (LD) were filtered out. This was done in R v. 3.2.5 (http://www.r-project.org/; R core team 2016), using the Rqtl (Broman et al. 2003) and qtltools (https://github.com/jtlovell/qtlTools) packages. SNPs displaying variation in only one or two isolates were likely sequence errors. Consequently, markers with strong segregation distortion were identified, based on a P-value < 10^{-8}, and removed. Markers with a segregation distortion between the P-value of <0.05 and > 10^{-8} were considered as true segregation distortion and these were retained for further analyses and included in linkage map construction. SNP markers were considered to be in complete linkage if the recombination frequency was lower than the minimum possible value of 0.014. One representative SNP was thus selected, using the qtltools argument DropSimilarMarkers, with a block size of 70 SNPs and a recombination frequency <0.014. Markers with strong segregation distortion or a high number of missing data were considered as co-factors when selecting SNPs for removal. The final set of SNPs was used for linkage map construction.

2.5. Linkage map construction

The linkage map was constructed in JoinMap v. 4.1 (van Ooijen 2006). Markers were assigned to linkage groups using an independence LOD statistic, evaluating LOD thresholds between 2.0-10.0. This

approach is based on the G^2 statistic for independence and is unaffected by segregation distortion. Linkages were not considered where the recombination frequency between markers was > 0.6. The final grouping of markers and linkage map construction was based on a LOD score of 6.0 or higher. Marker order and distances were determined using the maximum likelihood mapping algorithm and the default parameters. Recombination frequencies were estimated using Gibbs sampling, based on a maximum likelihood multipoint estimation. Recombination frequencies were converted to genetic distances (cM) based on the Haldane mapping function.

After the first round of linkage map construction, markers that resulted in an inflation in cM distance were manually inspected for SNP marker quality. Where the mapping quality was low, or if the SNP allele was inconsistent with the linkage phase of numerous surrounding markers, it was considered as a sequence error and removed from the map. If a single individual displayed a SNP marker with inconsistent linkage phase, the allele was changed to missing data for that individual. Markers that indicated a high fit/stress value, and likely an incorrect placement in the map were also removed from the final map. Based on the number of individuals included in the linkage map (70), if a minimum of two individuals supported a recombination event, the minimum reliable cM distance between two markers was calculated as 3cM ($2^{Recombinant}/70^{Total}$ *100). The majority of markers with a distance of less than 3cM in the linkage map were removed and the linkage map was reconstructed with the remaining markers to obtain a final, more robust linkage map. The final illustration of the linkage map was constructed using MapChart v. 2.3 (Voorrips 2002).

Recombination maps for each of the 70 progeny isolates were constructed and visualised using GGT 2.0. This was done in order to summarise the inheritance of parental genetic material in each isolate and indicate the sites and number of recombination events. To determine the relationship between the cM distance and the physical distance in the genomes of both species, the average Kb per recombination unit (cM) was calculated for each linkage group by dividing the predicted super-contig size (MinION genomes) of a linkage group by its cM length (de Vos et al. 2007).

The MinION sequenced genomes were used to validate the construction and marker order of the linkage map. The linkage map could also be used to support the joining of scaffolds in the genome to the level of putative chromosomes for each species. LG1 of the linkage map contained a large size inflation (23 cM) in the middle of the linkage map and the MinION genomes of the two species were aligned, using LASTZ Alignment v. 1.02 (Harris 2007) in Geneious version 7.0.6. (https://www.geneious.com; (Kearse et al. 2012) to investigate genome synteny between the two species in this region. Inversions and translocations were visualised using the RCircos package (Zhang et al. 2013) in R v. 3.2.5 using BLASTn data as input.

2.6. Segregation of nuclear and mitochondrial genetic material

2.6.1. Nuclear genomic material

The segregation ratios of the final set of SNP markers used in the linkage map construction were investigated in the progeny to determine if there was any deviation from the expected 1:1 ratio of a haploid cross. This was based on a Chi-squared test, at α-values of 0.05 and 0.001, performed in JoinMap v. 4.1 (van Ooijen 2006). The position of the SNP markers with significant transmission ratio distortion were investigated to determine their distribution and whether they were clustered in specific regions.

2.6.2. Mitochondrial genomic material

The inheritance of parental mitochondria was investigated in the progeny genomes to determine whether recombination had occurred in the mitochondria of the progeny and to confirm which mating type (MAT1 or MAT2 self-sterile) served as the maternal parent. The mitochondrial contigs in the *C. fimbriata* and *C. manginecans* parental MinION-sequenced genomes were identified through a BLASTn analysis, using the *C. cacaofunesta* mitochondrial genome as reference (NCBI accession number JX185564; (Ambrosio et al. 2013). The gene content of the mitochondria of the two species were compared to identify genes/regions that differed between the species. Genes with SNP or indel differences were used as markers to trace the inheritance of mitochondrial genetic material in all 70 progeny isolates. The sequence data used for the progeny genomes were obtained by converting the bam files (the progeny reads mapped to the parental genome) to consensus sequences in CLC Genomics Workbench v. 6 (CLC bio, www.clcbio.com).

2.7. QTL identification and identification of candidate genes underlying the QTL

QTL mapping was performed using MapQTL v.6 (van Ooijen and Kyazma 2009) in order to identify QTLs associated with each of the phenotypic traits investigated. The mean value for each phenotypic trait and for each isolate was used for analyses. Putative QTLs were identified using interval mapping, and the likelihood for a QTL was calculated using a maximum likelihood mixture model. MQM mapping (multiple-QTL model) was also performed for all phenotypes, to reduce the interval size of the QTL, and was implemented downstream if it improved the mapping. A permutation test with 1,000 replicates was performed with the data of each phenotypic trait to determine the genome wide significance LOD score threshold for each trait at a α-value of 0.05 (significant) and 0.001 (highly significant). Illustrations of the QTL results were constructed using MapChart v. 2.3 (Voorrips 2002). The QTL analysis outputs also indicated what percentage of the variation can be explained by the QTL interval. The overall phenotypic effect of the *C. manginecans* and *C. fimbriata* alleles of the SNP markers spanning each QTL region were calculated based on the average phenotypic score for the

reference allele (*C. manginecans*) versus the alternate allele (*C. fimbriata*) in the progeny, as performed by Mideros et al. (2017).

The SNP markers flanking each QTL interval were correlated back to the *C. manginecans* MinION genome to identify the corresponding genomic interval. The corresponding genome fragment was then aligned to the *C. fimbriata* MinION genome, using LASTZ Alignment v. 1.02 (Harris 2007) in Geneious, to identify the corresponding region in *C. fimbriata*. The alignment was then investigated for regions that were unique to either species or any orthologs that had amino acid differences between the species as these differences might influence the function of the protein.

3. Results

3.1. Reference genomes of the parental isolates

The Illumina sequence data obtained for the *C. manginecans* (CMW46461) genome consisted of 36.2 Gb of data, including 20.6 million mate-pair reads and 102.2 million and 76.8 million paired end reads for the 300 bp and 500 bp libraries, respectively, after quality filtering. This resulted in a genome with a size of 32.05 Mb and a GC content of 47.86%. The genome was distributed among 231 contigs with a 540 times average coverage and an N50 value of 593.7 Kb. In total, 7,260 genes were predicted in the genome. The genome sequence was deposited at GenBank under the accession SGIO00000000, the version described in this paper is SGIO01000000. The BUSCO analyses indicated a completeness score of 98% for both Illumina genomes and 98.3 % and 97.4 % for the *C. manginecans* and *C. fimbriata* MinION genomes, respectively.

3.2. Whole genome re-sequencing of the progeny isolates

The Ion Torrent sequence analysis of the 70 progeny isolates produced an average of 16 Gb data (99 million reads) per Ion Proton chip. Depending on the quality of the DNA and the number of isolates combined per chip, the number of reads per individual, after quality filtering and trimming ranged from 0.48 to 6.5 million reads, with an average of 1.3 million reads per isolate. This resulted in a 19x average genome coverage per isolate. The sequence data are available from the NCBI Short Read Archive (NCBI accession SUB4871173).

3.3. SNP genotyping of parents and progeny isolates

Using the *C. manginecans* Illumina genome as reference, a total of 43,144 SNPs were identified from the combined data of *C. fimbriata* and the 70 progeny isolates, after all quality filtering steps. The final number of SNPs retained for analyses, that were present in >90% of the progeny isolates, were 38,512. The filtered set of 38,512 SNP markers segregated in a ratio of 51.2 (*C. manginecans*) : 48.8 (*C.*

fimbriata) in the 70 progeny isolates, which did not differ significantly from a 1:1 inheritance ratio (P = 0.05).

3.4. Linkage map construction

The 38,512 SNP markers identified in *C. fimbriata* and the 70 progeny isolates were reduced to a smaller subset of 2,300 SNPs, in Rqtl, which was then used for initial linkage map construction in JoinMap v. 4.1. After further manual filtering and removing redundant markers (no recombination between the markers), 467 final SNP markers were used to construct the linkage map. This initial linkage map consisted of eight linkage groups.

A large expansion was present in the middle of LG1 and thus the alignments of the MinION genomes of the two species were investigated in this region. LG1 corresponded to four scaffolds of the *C. manginecans* MinION genome (scaffold 3, 10, 13 and 17) and three scaffolds of the *C. fimbriata* MinION genome (scaffold 29, 31 and 8). LASTZ genome synteny alignments of the scaffolds from each species indicated reciprocal translocations between the ends of scaffold 17 (*C. manginecans*) and 29 (*C. fimbriata*) and between the ends of scaffold 10 (*C. manginecans*) and 31 (*C. fimbriata*) (Fig. 1). This suggested that scaffolds 3 and 10 represent one linkage group and scaffolds 13 and 17 another in *C. manginecans* and, in *C. fimbriata,* scaffolds 29 and 8 formed a separate linkage group from 31. The split of the linkage group into two was confirmed in the *C. manginecans* genome by the presence of eight telomeric repeat sequences $(TTAGGG)_8$ and transposable elements at the end of scaffold 17, suggesting a telomeric end. LG1 was thus split into two smaller linkage groups (LG1 and LG2), more likely representing the number of chromosomes in the species. Scaffolds were joined, *in silico*, to represent the new linkage group 1 and 2 of each species in a final genome synteny map (Fig. 1). The final linkage map consisted of nine linkage groups with the largest linkage group being 386 cM (LG3) and the smallest 46 cM (LG8) resulting in a total length of 1203 cM (Fig. 2). The average spacing between markers was 2.55 cM and the maximum spacing was 11.16 cM (Table 1).

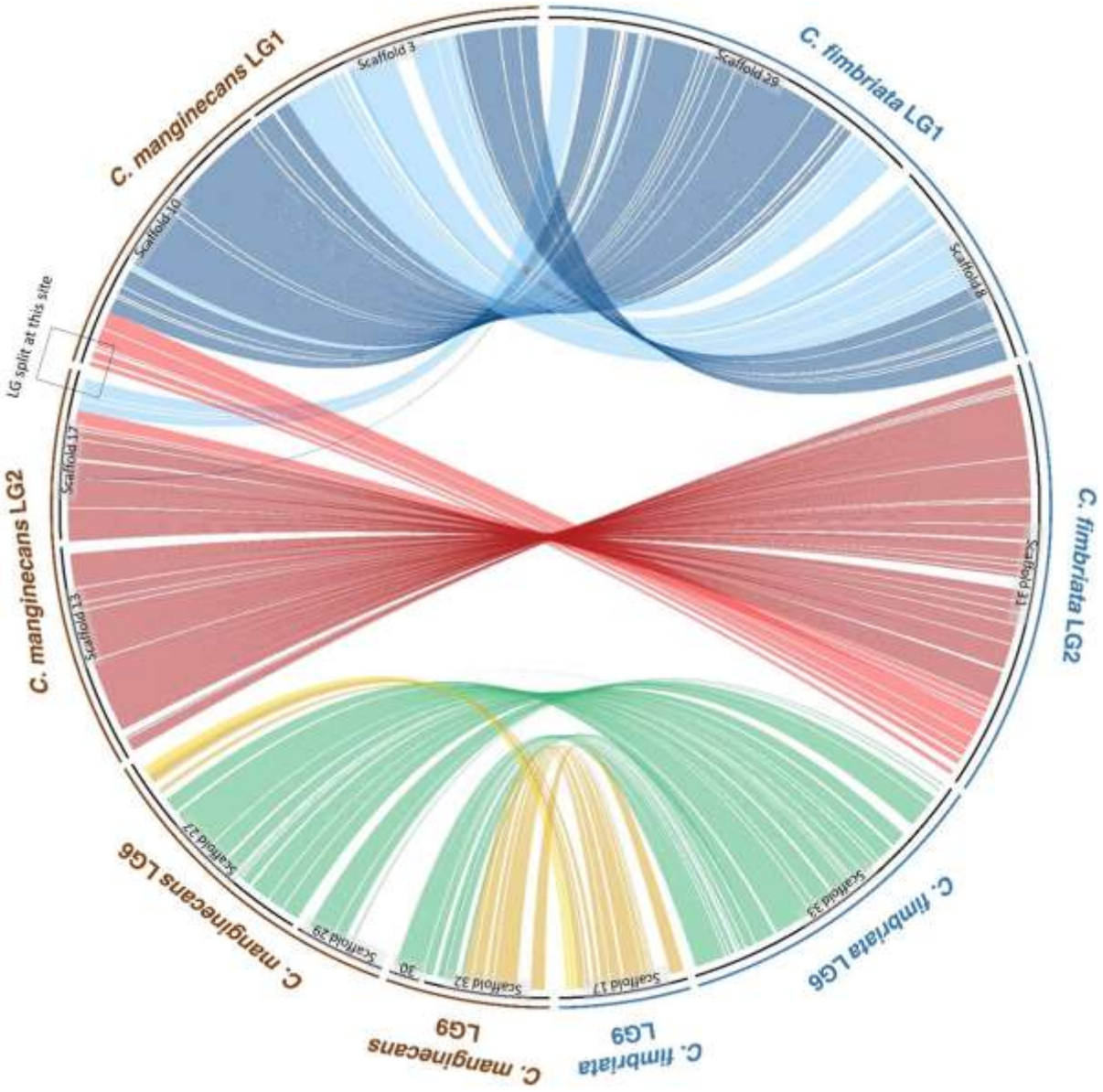

Fig. 1. Genome synteny map of four linkage groups of the *C. fimbriata* (blue lines) and *C. manginecans* (brown lines) genomes displaying translocations. This includes linkage group 1 (blue connector lines), 2 (red connector lines), 6 (green connector lines) and 9 (yellow connector lines). Scaffolds (black lines) were joined, based on evidence from the linkage map, to obtain full linkage groups. The coloured lines indicate genome synteny, darker blue and darker red lines are syntenic in the same orientation and lighter lines are in inverted orientation.

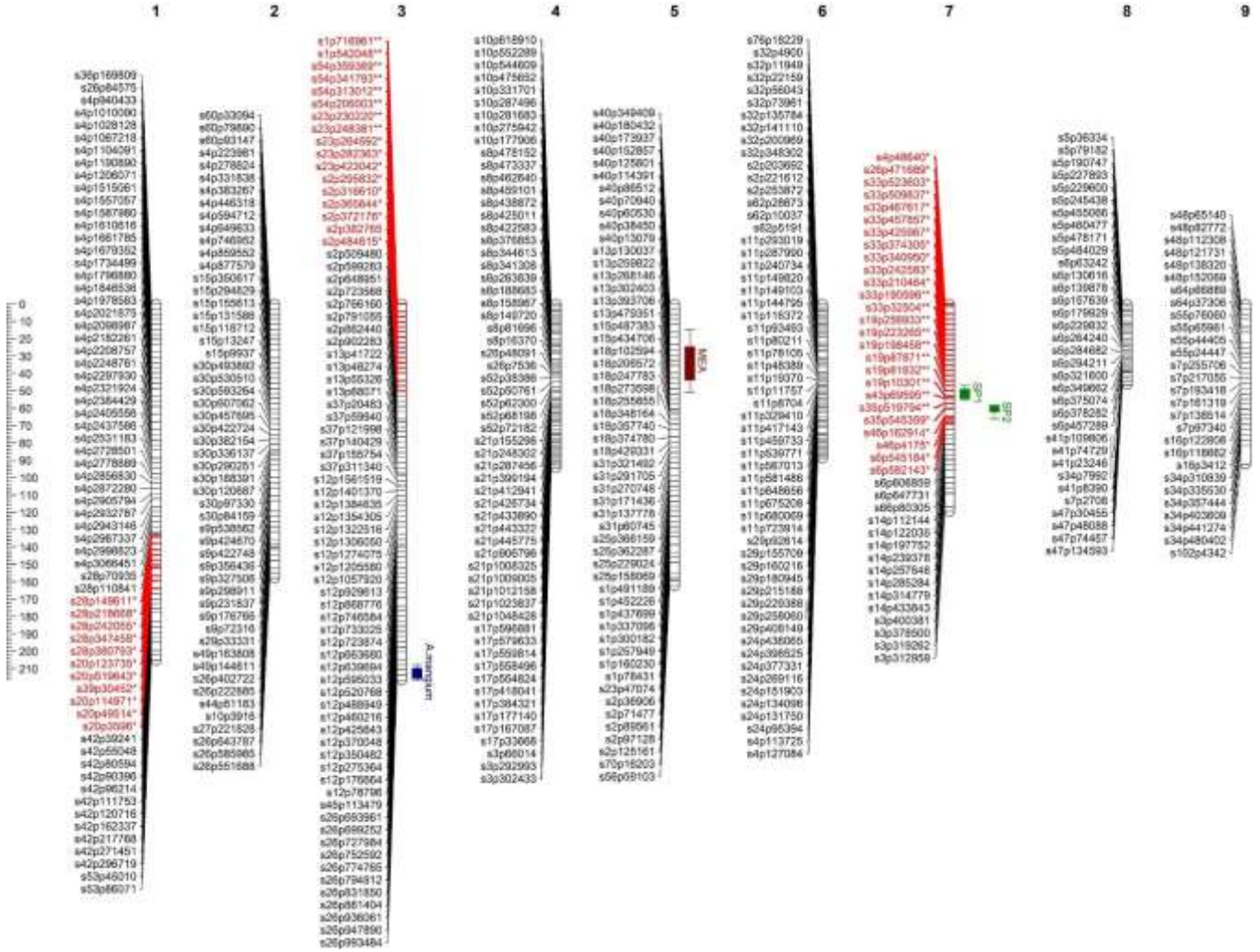

Fig. 2. Linkage map generated from the interspecific cross between *C. fimbriata* and *C. manginecans* showing nine linkage groups. The SNP marker positions are scaled on cM distance. SNP markers coloured in red displayed segregation distortion at significance levels of 0.05 (*) and 0.001 (**). The coloured boxes indicate the regions with significant association with a quantitative trait, including lesion length on *A. mangium* (blue), on *I. batatas* (green) and growth rate on MEA (brown).

Table 1. Summary of the linkage map and cross-over events in the progeny of an interspecific cross between *C. manginecans* (Cm) and *C. fimbriata* (Cf)

Linkage Group	Length (cM)	Estimated actual length (bp)	No. of SNPs	Avg. marker spacing (cM)	Max. marker spacing (cM)	Avg. Kb/ cM[a]	No. of cross-overs [b]								
							Non-rec. (Cm)	Non-rec. (Cf)	1	2	3	4	5	6	Total cross-over events
1	205.3	5,434,539 (Cm), 5,489,302 (Cf)	66	3.16	5.7	25.03 26.74	4	3	21	19	16	5	1	1	138
2	157.9	3,534,763 (Cm), 3,848,799 (Cf)	53	3.04	6.07	24.19 24.37	10	5	22	19	10	3	1	0	107
3	216.6	4,945,956 (Cm), 4,957,660 (Cf)	73	3	4.48	22.83 22.89	3	5	18	22	9	9	2	2	147
4	94.5	3,636,574 (Cm), 3,623,386 (Cf)	60	1.6	2.9	38.48 38.34	9	14	31	14	2	0	0	0	65
5	162.4	3,989,000 (Cf), 3,904,000 (Cm)	54	3.06	6.07	24.56 24.04	7	6	23	20	12	0	1	1	110
6	88.7	2,897,836 (Cm), 2,858,265 (Cf)	58	1.26	2.9	32.67 24.19	15	12	27	14	2	0	0	0	61
7	119.6	3,484,299 (Cf), 3,657,993 (Cm)	41	2.99	7.7	29.13 30.59	14	4	32	12	7	1	0	0	81
8	46.5	2,459,000 (Cm), 2,339,819 (Cf)	34	1.4	2.9	52.88 50.32	24	29	4	11	2	0	0	0	32
9	92.3	1,443,314 (Cm), 1,151,030 (Cf)	28	3.42	11.16	15.64 18.45	14	14	23	18	1	0	0	0	62

[a] Average calculated based on contig size divided by cM size.

[b] The eight columns represent the number of cross-over events observed in the progeny for each linkage group, ranging from non-recombinants of the *C. fimbriata* or *C. manginecans* genotype up to six cross-over events. The values in the columns indicate the number of individuals that displayed a specific number of cross-over events.

By integrating the linkage map and MinION genome assemblies, discrepancies could be identified and genome assemblies could also be improved. The genome synteny alignments indicated a large inversion near the end of LG1 in scaffold 8 of *C. fimbriata* and scaffold 3 of *C. manginecans*, respectively (Fig. 1). The marker order in this region of the linkage map represented that of the *C. fimbriata* genome. Another reciprocal translocation was observed between LG6 (*C. fimbriata* scaffold 33, *C. manginecans* scaffold 27 and 29) and LG9 (*C. fimbriata* scaffold 17, *C. manginecans* scaffold 27, 30, 32) (Fig. 1), where the marker order corresponded to the *C. manginecans* genome. The linkage map was used to improve the *C. manginecans* genome assembly. The linkage map indicated the split of *C. manginecans* scaffold 10 into two fragments that could be assigned to LG1 and LG3, respectively (results not shown). For both species, the linkage map supported the joining of between two to four contigs into a linkage group, reducing the assemblies to nine super-contigs (>1.1 Mb) each. Only 103.7 Kb of smaller contigs in *C. manginecans* and 147.7 Kb in *C. fimbriata* remained unmapped (0.4%) and contained only TE annotation and no protein coding genes.

The number of recombination events per isolate ranged from 5 to 22, with an average of 12.5 per isolate. Among all nine linkage groups, and the 70 progeny isolates, a total of 803 recombination events were observed with the most frequent events (147 crossovers) occurring in LG3 (Table 1). LG8 had the lowest number of recombination events and 53 of the isolates displayed no recombination in this linkage group. The average relationship between physical distance and recombination units varied between the linkage groups and ranged from 15 to 52 Kb/cM in the different groups (Table 1).

3.5. Segregation of nuclear and mitochondrial genetic material

3.5.1. Nuclear genetic material

The inheritance of parental genetic material was investigated in each of the progeny isolates, based on the final set of SNP markers included in the linkage map. Investigation of the progeny genotypes confirmed that all progeny were recombinant and 42 of the isolates inherited almost half (40 to 60%) of their genetic material from each parent (Fig. 3). The inheritance ratio that was most skewed towards *C. fimbriata* in an isolate was 25:74 (*C. manginecans*: *C. fimbriata*) and the ratio most skewed towards *C. manginecans* was 79:21 (*C. manginecans*: *C. fimbriata*). Only 11% of the SNP markers in the linkage map displayed segregation distortion and these were clustered into distinct regions in three of the linkage groups (Fig. 2). A region in LG1 (11 SNP markers) and the one end of LG7 (26 SNP markers) displayed an inheritance skewed towards the *C. manginecans* parent and the one end of LG3 (17 SNP markers) was skewed towards the *C. fimbriata* parent.

The region of segregation distortion in LG1 corresponded with a 1.64 Mb inversion observed between the scaffolds of the *C. fimbriata* (scaffold 8 and 29, LG1) and *C. manginecans* (scaffold 3, LG1) genomes (Fig. 1). This likely explains the limited recombination and skewed inheritance of genetic material in

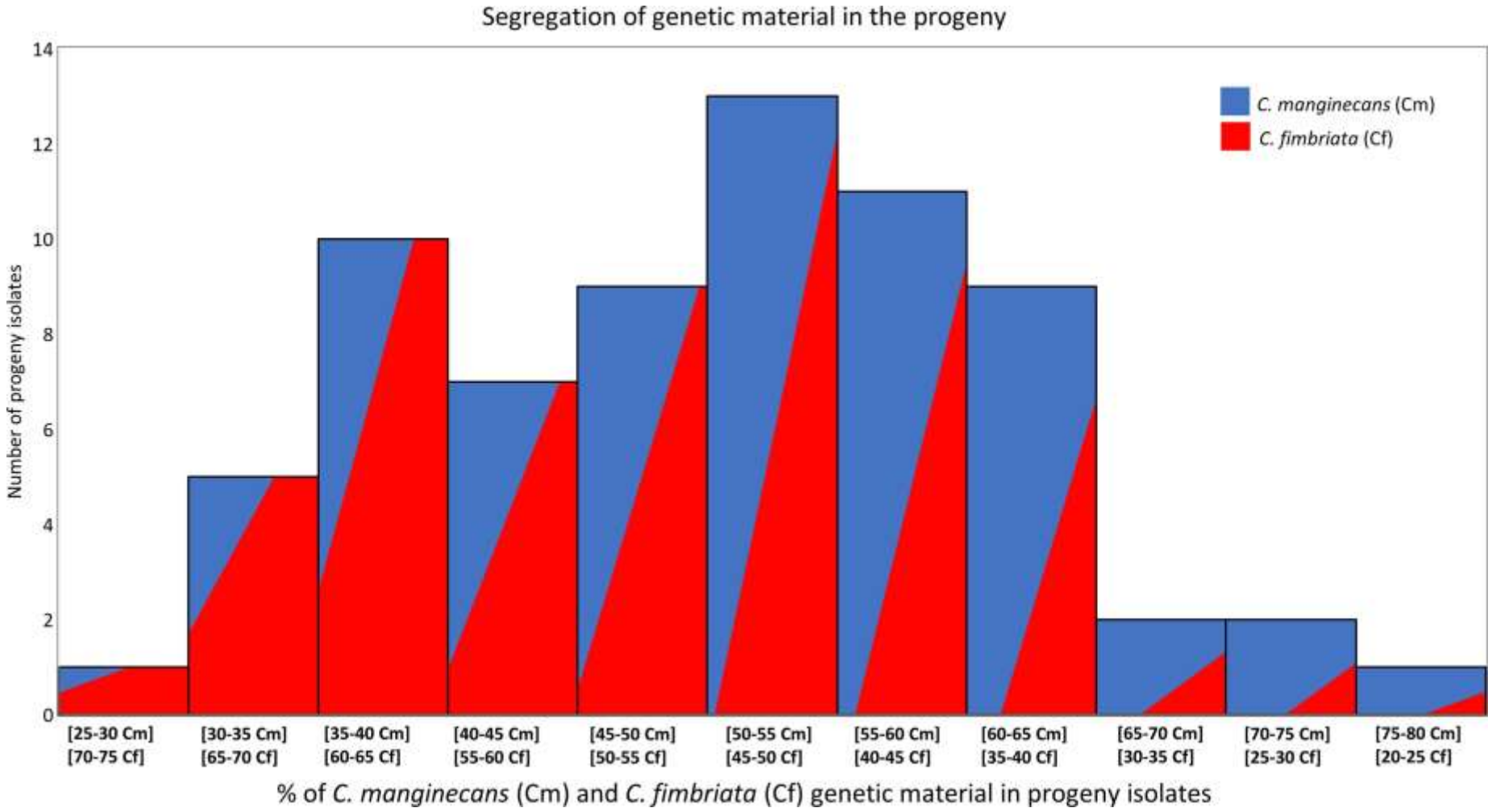

Fig. 3. Segregation of parental genetic material in the progeny isolates. The x-axis represents the % range of parental material inherited (*C. manginecans* in blue and *C. fimbriata* in red) and the y-axis indicates the number of progeny isolates displaying a specific segregation ratio.

this region. LG7 had a smaller 116,000 bp inversion in the region with segregation distortion, alongside but not including the mating-type locus, but this inversion is less likely to have interfered with recombination events due to its size.

3.5.2. Mitochondrial genomic material

From the comparison between the mitochondrial genomes of *C. fimbriata* and *C. manginecans*, five genes with polymorphisms were identified between the two species and these included *rnl, cox2, nad5, cox1* and *rns*. Each of the five genes differed in the presence/absence of certain intron open reading frames (ORF). ORFs missing in *C. manginecans*, in comparison to *C. fimbriata*, included oi1rnl, oi1_2nad5, oi6nad5, oi8cox1 and oi1rns. The progeny that originated from the *C. manginecans* MAT1 x *C. fimbriata* MAT2 cross all inherited the *C. manginecans* mitochondrion (reflecting the same variation in the five mitochondrial genes). Similarly, those from the *C. fimbriata* MAT1 x *C. manginecans* MAT2 cross inherited the *C. fimbriata* mitochondrion. No recombination was observed in the mitochondrial genomes of the progeny isolates.

3.6. QTL identification and investigation of candidate genes

Significant QTL regions could be identified for three of the phenotypic traits investigated. This included one highly significant QTL for growth rate on MEA ($p < 0.001$), two significant QTLs for aggressiveness on *I. batatas* ($p < 0.05$) and one QTL for aggressiveness on *A. mangium* ($p < 0.05$). The QTLs for each phenotype were located on a different linkage group (Fig. 2).

3.6.1. Mycelial growth rate on malt extract agar

A highly significant QTL peak ($p < 0.001$) was present on linkage group 5 (Fig. 4) that could explain 38% of the variation in growth rate on MEA. The 95% confidence interval (LOD > 3.2) was 427 Kb long and 247.6 Kb was above the 0.001 significance threshold (LOD > 5). The region of high significance ($p < 0.001$) contained 60 genes of which only one hypothetical protein was unique to *C. manginecans* (Table S1). The region was very gene dense with few transposable elements.

Of the 60 genes in the highly significant region, 17 genes differed with more than one amino acid between the two species. The polymorphic genes included peptidyl-prolyl cis-trans isomerase, a ras GTPase-activating-like protein (*Ras-GAP*), tRNA modification GTPase, GTPase-activating protein (*GYP7*), CRAL-TRIO domain-containing protein, protein SYG1 (*SYG1*), platelet-activating factor acetylhydrolase, glycoside hydrolase 18, DNA repair protein (*crb2*), microfibrillar-associated like-protein 1, serine/threonine-protein phosphatase 2A (*PP2A*) regulatory subunit delta, 20S-pre-rRNA D-site endonuclease, kynurenine/alpha-aminoadipate aminotransferase, MAU2 chromatid cohesion factor and three hypothetical protein genes. Based on the RNA sequence data used for the *C. fimbriata*

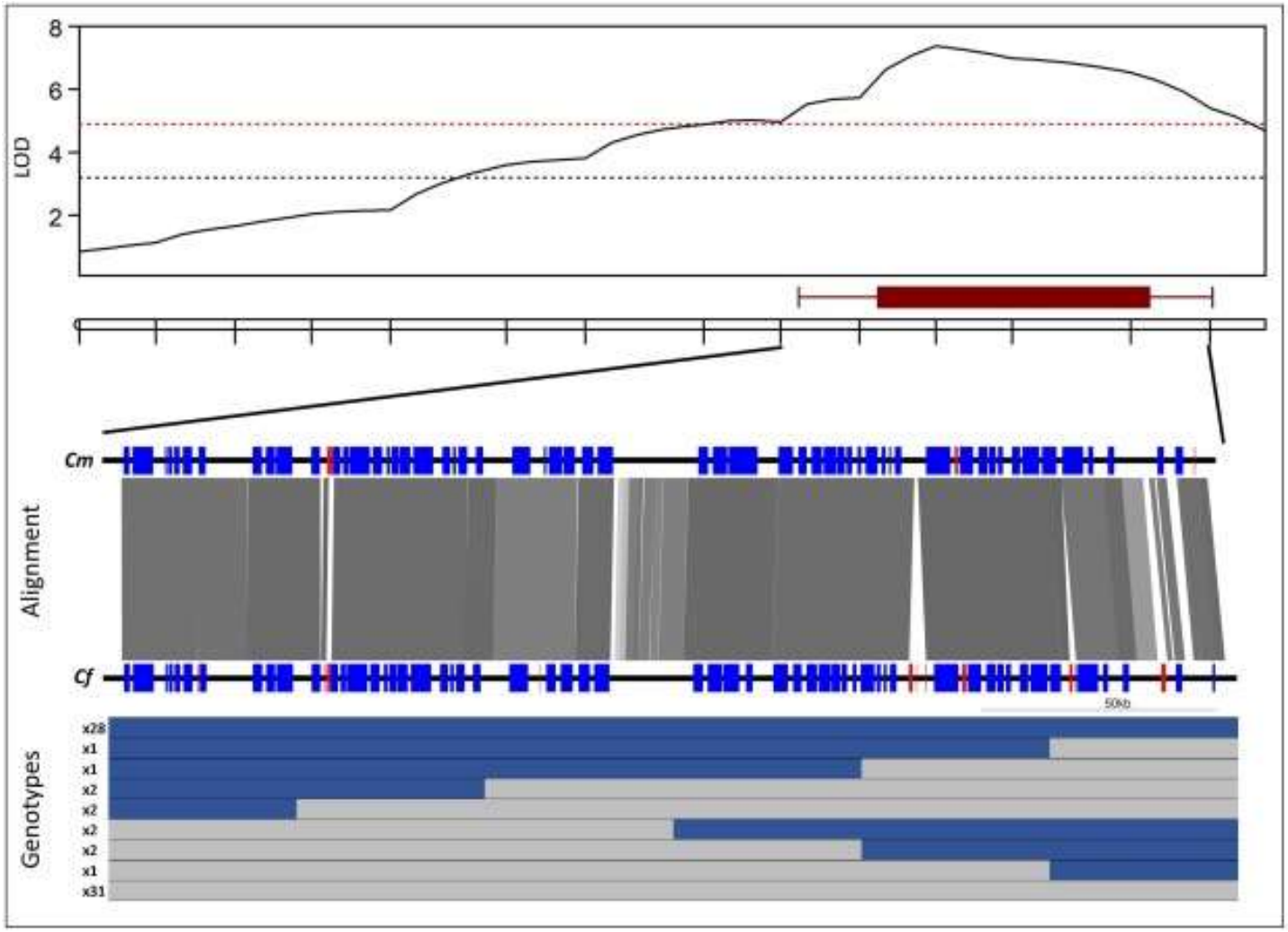

Fig. 4. The QTL region associated with mycelial growth rate on MEA. The top part of the figure indicates the α = 0.05 (black dotted line) and α = 0.001 (red dotted line) LOD thresholds for the QTL interval. Underneath the peak, comparisons of the genic (blue blocks) and TE content (red blocks) between *C. manginecans* (Cm) and *C. fimbriata* (Cf) are presented, with grey lines indicating the sequence synteny between the regions. At the bottom, the inheritance of the parental alleles in the F_1 progeny (*C. fimbriata* alleles in blue and *C. manginecans* in grey bars) are indicated, along with the frequency of the genotypes in the F_1 population.

genome annotation (Fourie et al. 2019), four of the hypothetical proteins had clear expression data and one had only low levels.

In the previously performed phenotypic studies (Fourie et al. 2018), *C. fimbriata* displayed more rapid growth than *C. manginecans* on MEA. The inheritance of the parental allelic variants in this QTL region corresponded with the observed growth phenotypes. The allelic effect of the reference alleles (*C. manginecans*) was negative, with an average growth rate of 8.31 mm/2 days versus an average of 10.23 mm/2 days for the alternate allele (*C. fimbriata*). Thus, a slower growth rate was associated with the reference alleles than with the alternate alleles (Fig. 4). This was corroborated by the fact that the seven fastest growing progeny isolates contained the *C. fimbriata* allelic variants. Eight of the slowest growing isolates contained the *C. manginecans* allelic variants throughout the QTL interval.

3.6.2. Pathogenicity on Ipomoea batatas

Two significant QTL peaks ($p < 0.05$, LOD > 2.9), associated with aggressiveness on *I. batatas,* were mapped alongside each other on LG7 (Fig. 5). The two QTL peaks could explain 19% of the observed variation. This entire region displayed significant segregation distortion, with 70% of the progeny displaying the *C. manginecans* genotype. The first peak (Fig. 5A) had a 95% confidence interval of 370 Kb and contained 46 genes (Table S2). Of these genes, 20 had more than two amino acid differences between *C. manginecans* and *C. fimbriata,* four genes were present only in *C. manginecans* and one gene was unique to *C. fimbriata*. The four unique genes in *C. manginecans* were part of a 30 Kb region, also containing transposable elements, that were mostly absent from *C. fimbriata*. The four unique *C. manginecans* genes included a hypothetical protein, a putative effector protein, alpha-galactosidase (*GLA*) (part of a complex that scavenge waste proteins) and a F-box domain containing protein (modifies glycolipids and glycoproteins). The unique *C. fimbriata* gene was also a putative effector. The other 17 candidate genes with amino acid differences included a cysteine desulphurase, betaine aldehyde dehydrogenase 2 (*BADH*) (oxidative stress response), a putative effector protein and hypothetical proteins (of which 10 had RNA evidence for transcription). The allelic effect of the reference allele (*C. manginecans*) was negative, with an average lesion size of 7.78 mm versus the *C. fimbriata* allele with an average lesion length of 15.16 mm.

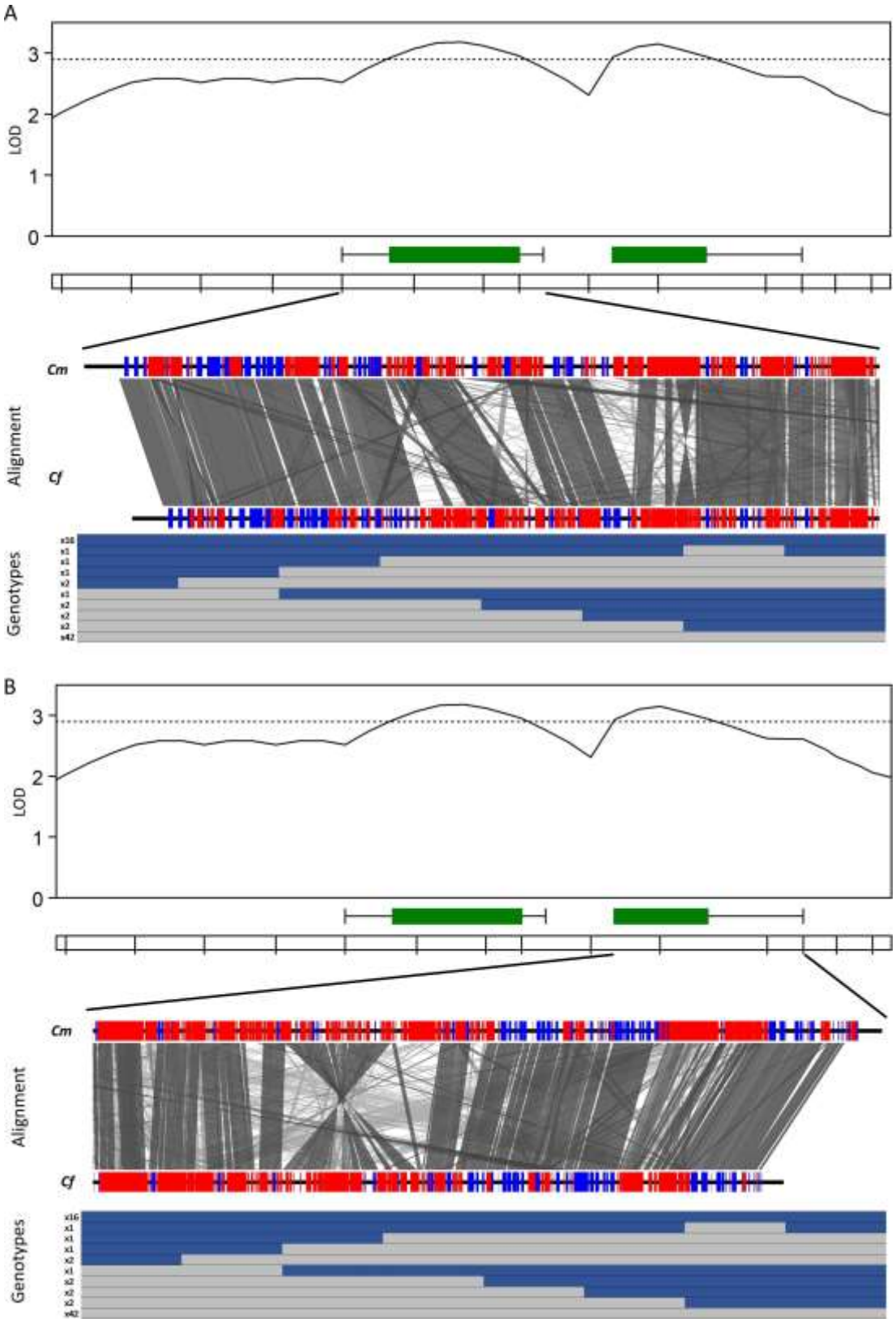

Fig. 5. QTL1 (A) and QTL2 (B) associated with aggressiveness on *I. batatas*. The top part of the figure indicates the α = 0.05 LOD threshold (black dotted line) for the QTL interval. Underneath the peak, comparisons of the gene (blue blocks) and TE content (red blocks) between *C. manginecans* and *C. fimbriata* are presented, with grey lines indicating the sequence synteny between the regions. At the bottom, the inheritance of the parental alleles in the F_1 progeny (*C. fimbriata* alleles in blue and *C. manginecans* in grey bars) are indicated, along with the frequency of the genotypes in the F_1 population.

The region between the two QTL peaks contained 30 genes which, amongst other genes, included the mating type locus. The second QTL peak (Fig. 5B) consisted of a 258 Kb region in the 95% confidence interval and contained 31 genes of which seven had two or more amino acid differences between the two species (Table S2). The genes included a N-acetyltransferase (*eso1*), Proteasome-associated protein and five hypothetical proteins (all of which had RNA evidence for transcription). Three consecutive genes were absent from *C. fimbriata*, a diacylglycerol acyltransferase (*DGAT*), Proteasome-associated protein (*ECM29*) and a hypothetical protein. Further downstream, *C. manginecans* also contained a transposable element with an associated retrovirus Pol polyprotein that aligned very differently to the *C. fimbriata* genome. Both the three unique genes and the TE were inherited together in the majority of the progeny. This QTL region was also associated with a negative allelic effect for the reference allele, where the average lesion length for isolates with the *C. fimbriata* alleles (Fig. 5) was 6.33 mm larger than for those with the *C. manginecans* allele.

3.6.3. Pathogenicity on Acacia mangium

One significant QTL ($p < 0.05$, LOD > 3.1) at the end of LG3 was associated with aggressiveness on *A. mangium* (Fig. 6). This region could explain 19% of the variation observed for the phenotype. This is a putative sub-telomeric region of a chromosome, based on the presence of five repeats of a telomeric sequence (CCCTAA), a high density of TEs and the presence of a telomere capping protein coding gene in the last 120 Kb of the LG. The 95% confidence interval spanned a 281 Kb region which contained 64 genes (Table S3). Sixteen of the genes had more than one amino acid difference between the two species and four genes were unique to *C. manginecans;* of which two unique hypothetical proteins and a putative effector were near the sub-telomeric end.

The genes with amino acid differences included acetoacetyl-CoA synthetase (*AACS*), enhancer of polycomb-like protein (*EPL1*), carboxypeptidase (*CP*), DNA-directed RNA polymerase III subunit (*RPC6*), protein TED1, protein disulphide isomerase (*PDI*), kinesin-like protein 6 (*klp6*), transcription activator protein (*acu-15*), D-isomer specific 2-hydroxyacid dehydrogenase, aryl-alcohol dehydrogenase (*AAD14*), peptide methionine sulfoxide reductase (*Msr*), diatom spindle kinesin (*DSK1*) and two hypothetical protein genes (both with RNA evidence for transcription). The region also contained another putative effector with one amino acid difference between the species, flanked by transposable elements. A pathogenicity-associated gene (PHI-database; (Urban et al. 2015), calcineurin, was also in this region but it was identical in the two species. The reference allele had a negative effect, where isolates with the *C. manginecans* allele (Fig. 6) produced an average lesion length of 1.93 cm and those with the *C. fimbriata* allele, 3.58 cm.

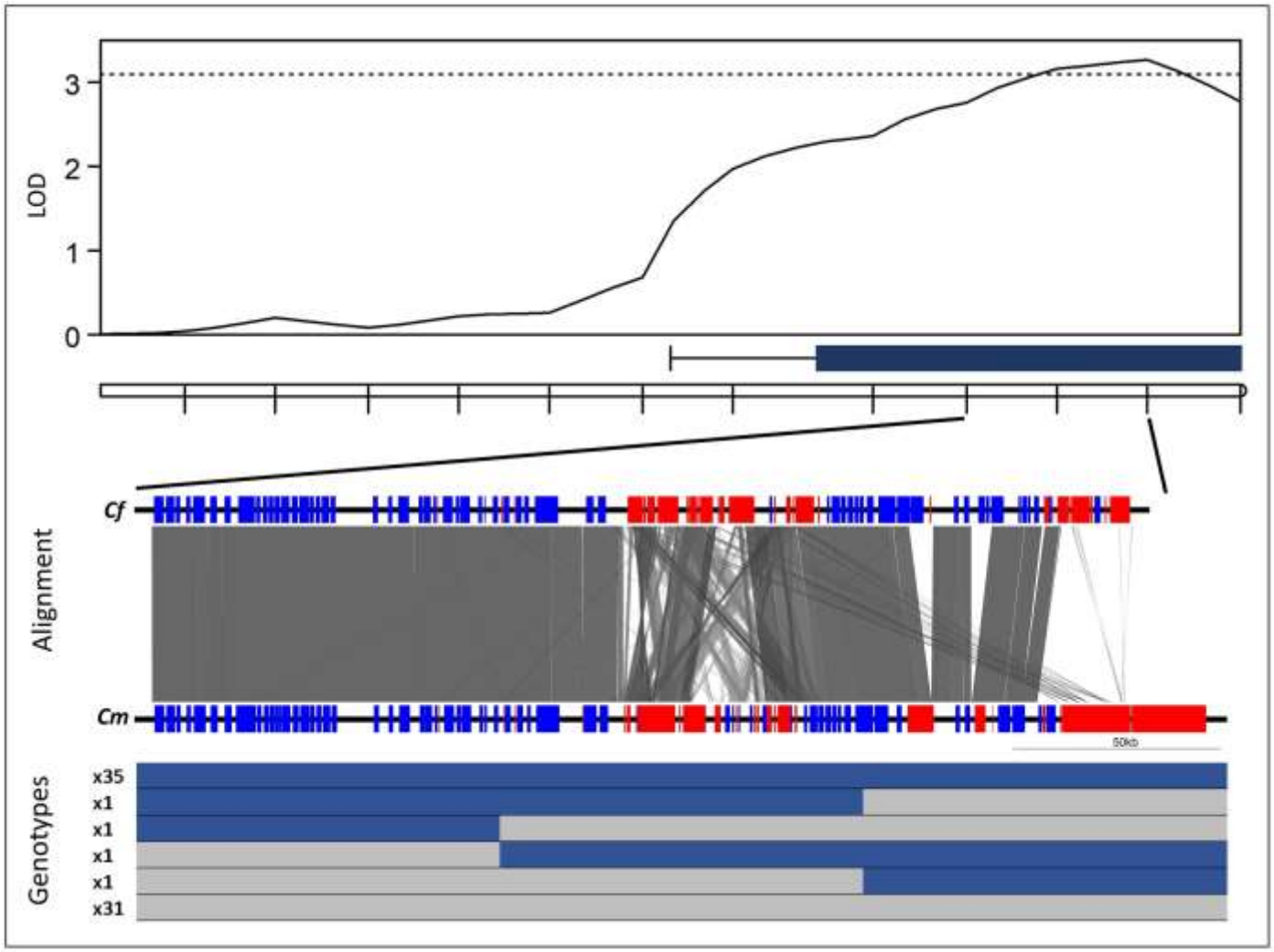

Fig. 6. QTL region associated with aggressiveness on *A. mangium*. The top part of the figure indicates the α = 0.05 LOD threshold (black dotted line) for the QTL interval. Comparisons of the gene (blue blocks) and TE content (red blocks) between *C. manginecans* and *C. fimbriata* are presented underneath with grey lines indicating the sequence synteny between the regions. At the bottom, the inheritance of the parental alleles in the F_1 progeny (*C. fimbriata* alleles in blue and *C. manginecans* in grey bars) are indicated, along with the frequency of the genotypes in the F_1 population.

4. Discussion

An interspecific cross produced between *C. fimbriata sensu stricto* and *C. manginecans* made it possible to construct the first linkage map for any species of *Ceratocystis*. The progeny-based linkage map resulted in nine linkage groups, providing the first indication of the number of chromosomes in the species. The phenotypic inheritance data, in conjunction with the linkage map, facilitated the identification of one genomic region associated with mycelial growth rate and three genomic regions associated with aggressiveness to distinct hosts. The available annotated genomes of both species made it possible to identify candidate genes in the QTL regions and provided a strong foundation for future functional studies on the pathogenicity factors in *Ceratocystis*.

The recombination events observed in the interspecific cross between *C. fimbriata* and *C. manginecans,* and the relatively low levels of segregation distortion, suggest that few pre-zygotic barriers exist during sexual recombination. The number of recombination events observed per isolate (12.5 avg) was similar to that of an interspecific cross of *Fusarium circinatum* x *F. subglutinans* (12 avg) (de Vos et al. 2007). This also did not deviate far from the frequency observed in intraspecific crosses such as *F. graminearum* (15.9 avg) (Laurent et al. 2018) and *Setosphaeria turcica* (15 avg) (Mideros et al. 2017). Some linkage groups in *Ceratocystis*, however, had significantly more recombination events than others. For example, LG9 had more recombination (62 events, 15-18 Kb/cM) than LG 4 and 8, even though it was the smallest linkage group (1.1 – 1.4 Mb). This could be a "hot spot" for recombination to occur, possibly due to transposable elements present in the linkage group (Seidl and Thomma 2017). LG8 had a high number of non-recombinant individuals (75%). This could be due to a lack of recombination, since the genome alignments indicated inverted chromosomal ends between the two species in LG8 (results not shown). Alternatively, recombination in this region might have a lethal effect and thus act as an additional post-zygotic barrier.

The overall inheritance of genomic material did not deviate from a 1:1 ratio in the progeny, but three distinct regions displayed segregation distortion (SD). SD can occur both in intra- and interspecific crosses in fungi but is expected to be more prominent in the latter case (de Vos et al. 2013). In a *F. pseudograminearum* intraspecific cross, five regions showed SD and one explanation was the conservation of essential virulence gene clusters (Gardiner et al. 2018). In the *F. circinatum* x *F. temperatum* interspecific cross, more than 50% of the markers deviated from the 1:1 ratio but they were not clustered in a specific region (de Vos et al. 2013). In that case, the SD was ascribed to chromosomal differences between the species (translocations or inversions) that influenced recombination and the viability of the progeny. New combinations of genes, fixed independently in two species, can also lead to nuclear incompatibility in the progeny (Dobzhansky-Muller hypothesis), contributing to SD (Giraud et al. 2008).

In the *Ceratocystis* cross of the present study, the SD was less extreme (11% skewed markers) than that observed in *Fusarium* (de Vos et al. 2013). In LG1 the SD is likely associated with an inversion and in LG7 the mating-type locus was inside the region of SD. The flanking regions of the mating-type locus of *C. fimbriata* and *C. manginecans* contain regions that are non-syntenic between the two species (Simpson et al. 2018), which could have contributed to the SD in this linkage group. However, other chromosomal differences such as size differences and deletions or unequal cross-over events could also be involved in the SD of the linkage groups (Plissonneau et al. 2016).

Full mitochondrial chromosome sequences were obtained for the progeny produced in this study. Based on DNA fingerprint markers, the inheritance of mitochondrial material in *Ceratocystis* has been associated with the maternal (MAT1) mating type (Engelbrecht and Harrington 2005). The inheritance of the five polymorphic mitochondrial genes considered in the present study confirmed that the MAT1 parent donated the mitochondria to the progeny. Historical recombination events have been observed between the mitochondria in different *Fusarium* species/*formae speciales* both in *F. oxysporum* (Brankovics et al. 2017) and the *F. fujikuroi* (Fourie et al. 2013) species complex. However, results of the present study showed that no recombination had occurred between the mitochondria of the two investigated *Ceratocystis* species during the mating event.

Linkage analysis of the 467 SNP markers resulted in nine linkage groups, which is most likely a close reflection of the number of chromosomes in *Ceratocystis.* In other Sordariomycetes, such as *Fusarium*, the chromosome number can range between four (Gale et al. 2005) and 20 (Waalwijk et al. 2018). In numerous fungal species, the number of linkage groups obtained from linkage analyses closely resembled the actual number of chromosomes and assisted in joining contigs for improved genome assemblies (Lee et al. 2008; Wittenberg et al. 2009; Yuan et al. 2018). It is thus likely that *C. fimbriata* and *C. manginecans* have nine chromosomes but that the sizes of some chromosomes would differ between the two species, based on the current genome alignments. This knowledge will make it possible to detect dispensable chromosomes in the genus and to investigate genome synteny between species.

The MinIon-sequenced genomes corroborated the linkage map constructed in this study and, in addition, the linkage map improved the genome assemblies. The linkage map indicated the split of one *C. manginecans* scaffold (scaffold 10) and could suggest the joining of multiple scaffolds up to the potential chromosomal level in both genome assemblies. The initial linkage map produced an abnormally large LG1 with a size inflation in the middle. Reciprocal translocations have been known to induce pseudo-linkage of genomic markers which results in the terminal joining of two linkage groups (Farré et al. 2011). The MinIon genome alignments of the two *Ceratocystis* species confirmed a reciprocal translocation between the ends of two scaffolds, suggesting that LG1 belongs to two

separate linkage groups. Since large chromosomal rearrangements could influence gene variation and pathogen adaptation and virulence (de Jonge et al. 2013; Wittenberg et al. 2009), the biological significance of the translocations and inversions observed between the two species should be investigated.

Mycelial growth on MEA had a highly significant QTL interval on LG5 which could explain 38% of the variation observed for the phenotype. This QTL explained more variation in mycelial growth rate than what was observed for individual QTLs in other ascomycetes such as *F. circinatum, F. temperatum* (6-32% per QTL; (de Vos et al. 2011) and *Z. tritici* (7-18%; (Lendenmann et al. 2016) but it was similar to that of *Setosphaeria turcica* (6-38%; (Mideros et al. 2017). However, in most of these studies between two and seven QTLs were identified and the combined data could explain more of the variation observed. The results from the present study confirmed that mycelial growth rate is a quantitative trait in *Ceratocystis* and additional QTLs are expected to be involved in this trait.

Of the 60 genes present in the QTL for mycelial growth rate, 17 differed in amino acid sequence between the parental isolates. Two of the candidate genes were of specific interest because they have been associated with mycelial growth in other fungi. For example, the *Ras-GAP* gene has been implicated in hyphal growth, specifically actin cytoskeleton organisation and conidiophore development, in *Aspergillus nidulans* and other fungi (Harispe et al. 2008). The PP2A regulatory sub-unit delta forms part of the protein phosphatase 2A holoenzyme that is involved in regulation of cell division during mitosis as well as cell signalling (Janssens and Goris 2001). This was one of the genes with the highest statistical support, as it was in the centre of the QTL peak. For both of these genes, the *C. fimbriata* SNPs were present in all the fastest growing progeny isolates and the *C. manginecans* SNPs were present in the slowest growing isolates. The results further support the likelihood of this QTL region as an essential regulator of mycelial growth rate in *Ceratocystis*.

Three QTLs could be identified in *Ceratocystis* relating to aggressiveness to distinct hosts; one QTL for aggressiveness on *A. mangium* and two QTLs for aggressiveness on *I. batatas*. However, the QTL for *A. mangium* had a negative correlation to aggressiveness. The QTL regions could explain 19% of the variation observed for aggressiveness on each of the respective hosts. Aggressiveness on a specific host is expected to be a quantitative trait in *Ceratocystis* (Fourie et al. 2018), and this is also true for virulence/aggressiveness in other fungal pathogens (Koladia et al. 2017; Lind et al. 2007; Stewart et al. 2016). This suggests that additional QTLs could be involved in host specificity but were not detected in this study. The inability to detect such QTLs could be due to the low population size and, specifically, the small number of pathogenic isolates compared to the number of non-pathogenic progeny isolates (Fourie et al. 2018).

Two overlapping QTLs were associated with aggressiveness on *I. batatas*. One candidate gene of interest in the first (left) QTL peak was betaine aldehyde dehydrogenase, which had 2 amino acid differences between *C. fimbriata* and *C. manginecans*. It is associated with oxidative stress response that could be involved in mitigating host stress responses (Singh et al. 2013). In addition, five putative effector proteins were identified throughout the two QTL regions. One of these was unique to each pathogenic species and two differed in amino acid sequence, that might influence their functionality. Because effectors have been shown to influence the virulence and host adaptation in numerous fungal pathogens (Dracatos et al. 2018; Sánchez-Vallet et al. 2018; Stewart et al. 2016), it is highly likely that they also play an important role in host adaptation in *Ceratocystis*. The effector that was absent from *C. fimbriata* might also enhance its ability to infect *I. batatas* since it would avoid host detection, as observed for avr genes in *Pyricularia oryzae* (syn. *Magnaporthe oryzae*) (Inoue et al. 2017) and for an effector in *Z. tritici* (Hartmann et al. 2017).

Both QTL regions for aggressiveness on *I. batatas* had a high TE density but the distribution of the TEs differed significantly between the two species of *Ceratocystis*, which could influence the variation and expression of genes. Some effectors were specifically within TE dense regions. This has often been observed in fungal pathogens and it has been suggested that the TEs induce sequence variations and influence the expression of the effectors (Meile et al. 2018; Möller and Stukenbrock 2017). The other protein coding genes located amongst the TEs could also be affected by the TE distribution, specifically if the promoter region is disrupted by the presence of the TE (Krishnan et al. 2018). Hence, although the gene content between the two species was similar, we hypothesise that the activation and regulation of the genes could differ between them and requires further *in planta* gene expression analysis.

The QTL for aggressiveness on *A. mangium* was located at the sub-telomeric end of LG3. The two species of *Ceratocystis* had significant dissimilarities toward the end of the scaffold. This could primarily be attributed to the density and distribution of transposable elements in this region. The content and organisation of genes at the telomeric ends of chromosomes have been implicated in virulence and host adaptation in pathogens such as *Fusarium graminearum* (Gardiner et al. 2012) and *Pyricularia oryzae* (syn. *M. oryzae*) (Orbach et al. 2000). This especially since telomeres are associated with higher sequence variation and rearrangements (Cuomo et al. 2007; Orbach et al. 2000). The TE dense region in this QTL could thus likely serve as a mutational “hotspot” that influences the aggressiveness of *Ceratocystis* pathogens.

Other genes of interest in the *A. mangium* QTL, that differed in two or more amino acids between the two host-specific species of *Ceratocystis*, included those associated with transcriptional regulation (*EPL1*, *RPC6*, *acu-15*), protein processing or modification (*UCHL1*, *CP*, *PDI*) and detoxification (*AAD14*,

Msr) (Yang et al. 2018). These processes could be associated with adaptation to a specific nutritional environment by processing metabolites or regulating the transcription of genes essential in the infection process. This limited number of candidate genes could serve as high priority targets for gene knock-outs in future studies.

The fact that most of the isolates aggressive to *A. mangium* contained the *C. fimbriata* genotype in the *A. mangium* QTL region was unexpected. One explanation could be that these isolates also inherited genetic material from *C. manginecans* in other genomic regions and that protein products from those regions had positive epistatic interactions with the QTL region, providing the strong signal for the QTL. This would relate to the Dobzansky-Muller effect where the rest of the genome content, inherited from the alternate parent, influences the phenotypic effect of an allele (Stukenbrock 2013). For example, a transcription factor or a repressor protein of *C. manginecans,* present in a different region of the progeny genome*,* might upregulate or suppress transcription in the QTL region. A prominent example is the calcineurin-responsive transcription factor *moCRZ1* in *Pyricularia oryzae* which regulates the expression of six other genes that influence the host penetration and pathogenicity of the fungal pathogen (Choi et al. 2009).

Even though only a single QTL was significantly associated with aggressiveness on *A. mangium*, other potential QTLs were detected below the significance threshold. The inclusion of more isolates in future studies can most likely improve the detection of additional QTLs having minor effects (Lendenmann et al. 2016). For example, the one QTL peak was located at the exact same interval on LG7 where the *I. batatas* QTL was located. This suggests that the genomic interval is likely involved in adaptation to both hosts and should be a region of interest for further functional studies.

Identifying specific genes associated with growth rate and aggressiveness on *A. mangium* and *I. batatas*, could enhance resistance breeding of the related hosts. This would also facilitate the development of genetically modified plants (Fan et al. 2015) with high levels of tolerance against *Ceratocystis* infection. This has been prominently illustrated in the wheat - *Stagnospora nodorum* pathosystem where knowledge of the fungal pathogenicity and host resistance factors enabled researchers to assess the aggressiveness of infecting populations and improve screening for resistant wheat cultivars (Oliver et al. 2012).

The inheritance study presented here provided the first linkage map and QTL mapping for a *Ceratocystis* species. This made it possible to identify genomic regions associated with *in vitro* mycelial growth rate and aggressiveness on *A. mangium* and *I. batatas*. A number of candidate genes were identified, either genes with amino acid variation or unique genes, that could be involved in growth rate and aggressiveness on a particular host. To confirm the significance of these genes, the variants

can be screened in natural populations of both species (Hartmann et al. 2017; Stewart et al. 2016) or tested by means of gene knock-out studies (Meile et al. 2018); a system that still needs to be optimised for *Ceratocystis*. In addition, the linkage map and improved genome assembly produced from this study, provides a useful tool for further quantitative studies in *C. fimbriata* and *C. manginecans* and will greatly facilitate the assembly of additional *Ceratocystis* genomes in future.

Acknowledgments

This study was made possible by the financial support of members of the Tree Protection Cooperative Programme, based at the Forestry and Agricultural Biotechnology Institute, and the Genomics Research Institute at the University of Pretoria, as well as the National Research Foundation (Grant UID: 89619 and 95875) and the DST/NRF SARChi chair in Fungal Genomics. The DNA Ion Torrent sequencing facility, University of Pretoria, and Central Analytical Facilities (CAF) at Stellenbosch University are thanked for providing the service for Ion Torrent DNA sequence determination as well as the NRF for funding the Ion Torrent equipment (UID: 88063), based at the CAF of Stellenbosch University. We thank Dr. Tuan Duong for providing access to the MinION sequenced genome assemblies. We also thank Prof. Zander Myburg (University of Pretoria) for advice and assistance with linkage map analyses.

References

Al-Sadi, A.M., Al-Ouweisi, F.A., Al-Shariani, N.K., Al-Adawi, A.O., Kaplan, E.J., Deadman, M.L., 2010. Histological changes in mango seedlings following infection with *Ceratocystis manginecans*, the cause of mango decline. J. Phytopathol. 158**,** 738-743.

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.

Ambrosio, A.B., Do Nascimento, L.C., Oliveira, B.V., Teixeira, P.J.P.L., Tiburcio, R.A., Toledo Thomazella, D.P., Leme, A.F.P., Carazzolle, M.F., Vidal, R.O., Mieczkowski, P., Meinhardt, L.W., Pereira, G.A.G., Cabrera, O.G., 2013. Global analyses of *Ceratocystis cacaofunesta* mitochondria: from genome to proteome. BMC Genomics. 14**,** 1-16.

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.

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.

Barnes, I., Roux, J., Wingfield, M.J., Coetzee, M.P.A., Wingfield, B.D., 2001. Characterisation of *Seiridium* spp. associated with cypress canker based on β-tubulin and histone sequences. Plant Dis. 85**,** 317-321.

Besemer, J., Borodovsky, M., 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 33**,** W451-W454.

Beukes, C.W., Venter, S.N., Law, I.J., Phalane, F.L., Steenkamp, E.T., 2013. South African papilionoid legumes are nodulated by diverse *Burkholderia* with unique nodulation and nitrogen-fixation loci. PLOS ONE. 8**,** e68406.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., Pirovano, W., 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 27**,** 578-579.

Boetzer, M., Pirovano, W., 2012. Toward almost closed genomes with GapFiller. Genome Biol. 13**,** R56.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30**,** 2114-2120.

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.

Brankovics, B., van Dam, P., Rep, M., De Hoog, G.S., van der Lee, T.A.J., Waalwijk, C., Van Diepeningen, A.D., 2017. Mitochondrial genomes reveal recombination in the presumed asexual *Fusarium oxysporum* species complex. BMC Genomics. 18**,** 735.

Broman, K., Wu, H., Sen, Ś., Churchill, G., 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 19**,** 889-890.

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.

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.

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.

Choi, J., Kim, Y., Kim, S., Park, J., Lee, Y.-H., 2009. *MoCRZ1*, a gene encoding a calcineurin-responsive transcription factor, regulates fungal growth and pathogenicity of *Magnaporthe oryzae*. Fungal Genet. Biol. 46**,** 243-254.

Cozijnsen, A.J., Popa, K.M., Purwantara, A., Rolls, B.D., Howlett, B.J., 2000. Genome analysis of the plant pathogenic ascomycete *Leptosphaeria maculans*; mapping mating type and host specificity loci. Mol. Plant Pathol. 1**,** 293-302.

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.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Genomes Project Analysis, G., 2011. The variant call format and VCFtools. Bioinformatics. 27**,** 2156-2158.

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 Long, E.F., 1992. Archaea in coastal marine environments. Proc. Natl. Acad. Sci. 89**,** 5685-5689.
de Vos, L., Myburg, A.A., Wingfield, M.J., Desjardins, A.E., Gordon, T.R., Wingfield, B.D., 2007. Complete genetic linkage maps from an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. Fungal Genet. Biol. 44**,** 701-714.
de Vos, L., van der Merwe, N.A., Wingfield, M.J., Myburg, A.A., Wingfield, B.D., 2013. Transmission ratio distortion in an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. Genes & Genom. 35**,** 177-183.
de Vos, L., van der Nest, M.A., van der Merwe, N.A., Myburg, A.A., Wingfield, M.J., Wingfield, B.D., 2011. Genetic analysis of growth, morphology and pathogenicity in the F1 progeny of an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. Fungal Biol. 115**,** 902-908.
Dean, R., van Kan, J.A., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., Ellis, J., 2012. The top 10 fungal pathogens in molecular plant pathology. Mol. Plant Pathol. 13**,** 414-430.
Deng, C.H., Plummer, K.M., Jones, D.A.B., Mesarich, C.H., Shiller, J., Taranto, A.P., Robinson, A.J., Kastner, P., Hall, N.E., Templeton, M.D., Bowen, J.K., 2017. Comparative analysis of the predicted secretomes of Rosaceae scab pathogens *Venturia inaequalis* and *V. pirina* reveals expanded effector families and putative determinants of host range. BMC Genomics. 18**,** 339.
Divon, H.H., Fluhr, R., 2007. Nutrition acquisition strategies during fungal infection of plants. FEMS Microbiol. Lett. 266**,** 65-74.
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.
Engelbrecht, C.J.B., Harrington, T.C., 2005. Intersterility, morphology and taxonomy of *Ceratocystis fimbriata* on sweet potato, cacao and sycamore. Mycologia. 97**,** 57-69.
Fan, D., Liu, T., Li, C., Jiao, B., Li, S., Hou, Y., Luo, K., 2015. Efficient CRISPR/Cas9-mediated targeted mutagenesis in populus in the first generation. Sci. Rep. 5**,** 12217.
Farré, A., Lacasa Benito, I., Cistué, L., de Jong, J.H., Romagosa, I., Jansen, J., 2011. Linkage map construction involving a reciprocal translocation. Theor. Appl. Genet. 122**,** 1029-1037.
Flutre, T., Duprat, E., Feuillet, C., Quesneville, H., 2011. Considering transposable element diversification in *de novo* annotation approaches. PLOS ONE. 6**,** e16526.
Fourie, A., Simpson, M.C., Duong, T.A., Barnes, I., Coetzee, M.P.A., Van der Nest, M.A., Wingfield, M.J., Wingfield, B.D., 2019. Genome annotation for *Ceratocystis fimbriata*: an aggressive fungal pathogen of sweet potato. IMA Fungus. Submitted.
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., 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.
Fourie, G., van der Merwe, N.A., Wingfield, B.D., Bogale, M., Tudzynski, B., Wingfield, M.J., Steenkamp, E.T., 2013. Evidence for inter-specific recombination among the mitochondrial genomes of *Fusarium* species in the *Gibberella fujikuroi* complex. BMC Genomics. 14**,** 605.
Gale, L.R., Bryant, J.D., Calvo, S., Giese, H., Katan, T., O'Donnell, K., Suga, H., Taga, M., Usgaard, T.R., Ward, T.J., Kistler, H.C., 2005. Chromosome complement of the fungal plant pathogen *Fusarium graminearum* based on genetic and physical mapping and cytological observations. Genetics. 171**,** 985-1001.
Gao, Y., Liu, Z., Faris, J.D., Richards, J., Brueggeman, R.S., Li, X., Oliver, R.P., McDonald, B.A., Friesen, T.L., 2016. Validation of genome-wide association studies as a tool to identify virulence factors in *Parastagonospora nodorum*. Phytopathol. 106**,** 1177-1185.
Gardiner, D.M., Benfield, A.H., Stiller, J., Stephen, S., Aitken, K., Liu, C., Kazan, K., 2018. A high resolution genetic map of the cereal crown rot pathogen *Fusarium pseudograminearum* provides a near complete genome assembly. Mol. Plant Pathol. 19**,** 217–226.

Gardiner, D.M., McDonald, M.C., Covarelli, L., Solomon, P.S., Rusu, A.G., Marshall, M., Kazan, K., Chakraborty, S., McDonald, B.A., Manners, J.M., 2012. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. PLOS Pathog. 8**,** e1002952.

Giraud, T., Refrégier, G., Le Gac, M., de Vienne, D.M., Hood, M.E., 2008. Speciation in fungi. Fungal Genet. Biol. 45**,** 791-802.

Glass, N.L., Donaldson, G.C., 1995. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. Appl. Environ. Microbiol. 61**,** 1323-30.

Goodwin, S.B., Drenth, A., Fry, W.E., 1992. Cloning and genetic analysis of two highly polymorphic, moderately repetitive nuclear DNAs from *Phytophthora infestans*. Curr. Genet. 22**,** 107-115.

Halsted, B.D., 1890. Some fungous disease of the sweet potato. Agric. Coll. Exp. Stn. Bull. 76**,** 1-32.

Harispe, L., Portela, C., Scazzocchio, C., Peñalva, M.A., Gorfinkiel, L., 2008. Ras GTPase-activating protein regulation of actin cytoskeleton and hyphal polarity in *Aspergillus nidulans*. Eukaryot. Cell. 7**,** 141-153.

Harrington, T.C., Huang, Q., Ferreira, M.A., Alfenas, A.C., 2014. 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.

Harris, R.S., 2007. Improved pairwise alignment of genomic DNA. Computer Science and Engineering. The Pennsylvania State University, pp. 1-84.

Harter, L.L., Weimer, J., 1929. A monographic study of sweet-potato diseases and their control. Tech. Bull. 99**,** 1-144.

Hartmann, F.E., Sánchez-Vallet, A., McDonald, B.A., Croll, D., 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 11**,** 1189.

Harwood, C., Nambiar, E., 2014. Productivity of acacia and eucalypt plantations in Southeast Asia. 2. Trends and variations. International Forestry Review. 16**,** 249-260.

Heinzelmann, R., Croll, D., Zoller, S., Sipos, G., Münsterkötter, M., Güldener, U., Rigling, D., 2017. High-density genetic mapping identifies the genetic basis of a natural colony morphology mutant in the root rot pathogen *Armillaria ostoyae*. Fungal Genet. Biol. 108**,** 44-54.

Holland, L.A., Lawrence, D.P., Nouri, M.T., Travadon, R., Harrington, T.C., Trouillas, F.P., 2019. Taxonomic revision and multi-locus phylogeny of the North American clade of *Ceratocystis*. Fungal Syst. Evol. 3**,** 135-156.

Holt, C., Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 12**,** 491.

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**,** 1150-1150.

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.

Janssens, V., Goris, J., 2001. Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. Biochem. J. 353**,** 417-439.

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.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28**,** 1647-1649.

Koladia, V.M., Richards, J.K., Wyatt, N.A., Faris, J.D., Brueggeman, R.S., Friesen, T.L., 2017. Genetic analysis of virulence in the *Pyrenophora teres f. teres* population BB25×FGOH04Ptt-21. Fungal Genet. Biol. 107**,** 12-19.

Korf, I., 2004. Gene finding in novel genomes. BMC Bioinformatics. 5**,** 59.

Krishnan, P., Meile, L., Plissonneau, C., Ma, X., Hartmann, F.E., Croll, D., McDonald, B.A., Sanchez Vallet, A., 2018. Transposable element insertions shape gene regulation and melanin production in a fungal pathogen. BCM Biol. 16**,** 78-96.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. methods. 9**,** 357-359.

Laurent, B., Palaiokostas, C., Spataro, C., Moinard, M., Zehraoui, E., Houston, R.D., Foulongne-Oriol, M., 2018. High-resolution mapping of the recombination landscape of the phytopathogen *Fusarium graminearum* suggests two-speed genome evolution. Mol. Plant Pathol. 19**,** 341-354.

Lee, J., Jurgenson, J.E., Leslie, J.F., Bowden, R.L., 2008. Alignment of genetic and physical maps of *Gibberella zeae*. Appl. Environ. Microbiol. 74**,** 2349-2359.

Lendenmann, M.H., Croll, D., McDonald, B.A., 2015. QTL mapping of fungicide sensitivity reveals novel genes and pleiotropy with melanization in the pathogen *Zymoseptoria tritici*. Fungal Genet. Biol. 80**,** 53-67.

Lendenmann, M.H., Croll, D., Palma-Guerrero, J., Stewart, E.L., McDonald, B.A., 2016. QTL mapping of temperature sensitivity reveals candidate genes for thermal adaptation and growth morphology in the plant pathogenic fungus *Zymoseptoria tritici*. Heredity. 116**,** 384-394.

Lendenmann, M.H., Croll, D., Stewart, E.L., McDonald, B.A., 2014. Quantitative trait locus mapping of melanization in the plant pathogenic fungus *Zymoseptoria tritici*. G3 Genes Genom. Genet. 4**,** 2519-2533.

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.

Lind, M., Dalman, K., Stenlid, J., Karlsson, B., Olson, Å., 2007. Identification of quantitative trait loci affecting virulence in the basidiomycete *Heterobasidion annosum s.l*. Curr. Genet. 52**,** 35-44.

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.

Ma, L.-J., Geiser, D.M., Proctor, R.H., Rooney, A.P., O'Donnell, K., Trail, F., Gardiner, D.M., Manners, J.M., Kazan, K., 2013. *Fusarium* Pathogenomics. Annu. Rev. Microbiol. 67**,** 399-416.

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.

Mideros, S.X., Chung, C.-L., Wiesner-Hanks, T., Poland, J., Wu, D., Fialko, A.A., Turgeon, G., Nelson, R., 2017. Determinants of virulence and *in-vitro* development colocalize on a genetic map of *Setosphaeria turcica*. Phytopathol. 108**,** 254-263.

Möller, M., Stukenbrock, E.H., 2017. Evolution and genome architecture in fungal plant pathogens. Nature Rev. Microbiol. 15**,** 756-771.

Niks, R.E., Marcel, T.C., 2009. Nonhost and basal resistance: how to explain specificity? New Phytol. 182**,** 817-828.

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.

Oliver, R.P., Friesen, T.L., Faris, J.D., Solomon, P.S., 2012. *Stagonospora nodorum*: from pathology to genomics and host resistance. Annu. Rev. Phytopathol. 50**,** 23-43.

Orbach, M.J., Farrall, L., Sweigard, J.A., Chumley, F.G., Valent, B., 2000. A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. The Plant Cell. 12**,** 2019-2032.

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.

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F.E., Croll, D., 2017. Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. Front. Plant Sci. 8**,** 1-15.

Plissonneau, C., Stürchler, A., Croll, D., 2016. The evolution of orphan regions in genomes of a fungal pathogen of wheat. mBio. 7**,** e01231-16.

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/.

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.

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.

Seidl, M.F., Thomma, B.P.H.J., 2017. Transposable elements direct the coevolution between plants and microbes. Trends Genet. 33**,** 842-851.

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.

Simpson, M.C., Coetzee, M.P.A., van der Nest, M.A., Wingfield, M.J., Wingfield, B.D., 2018. *Ceratocystidaceae* exhibit high levels of recombination at the mating-type (MAT) locus. Fungal Biol. 122**,** 1184-1191.

Singh, S., Brocker, C., Koppaka, V., Chen, Y., Jackson, B.C., Matsumoto, A., Thompson, D.C., Vasiliou, V., 2013. Aldehyde dehydrogenases in cellular responses to oxidative/electrophilic stress. Free Radic. Biol. Med. 56**,** 89-101.

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

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.

Stanke, M., Steinkamp, R., Waack, S., Morgenstern, B., 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32**,** W309-W312.

Stewart, E.L., Croll, D., Lendenmann, M.H., Sanchez-Vallet, A., Hartmann, F.E., Palma-Guerrero, J., Ma, X., McDonald, B., 2016. Quantitative trait locus mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. Mol. Plant Pathol. 19**,** 201-216.

Stukenbrock, E.H., 2013. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. New Phytol. 199**,** 895-907.

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.

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 Biology. 123**,** 170-182.

van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From fastQ data to high-confidence variant calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr. Protoc. Bioinformatics. 43**,** 1-33.

van Ooijen, J., 2006. JoinMap 4.0: Software for the calculation of genetic linkage maps in experimental populations. Wageningen, Netherlands: Kyazma BV.

van Ooijen, J., Kyazma, B., 2009. MapQTL® 6, Software for the mapping of quantitative trait in experiment populations of diploid species. Wageningen: Kyazma BV.

van Wyk, M., Al Adawi, A.O., Khan, I.A., Deadman, M.L., Al Jahwari, A.A., Wingfield, B.D., Ploetz, R., Wingfield, M.J., 2007. *Ceratocystis manginecans* sp. nov., causal agent of a destructive mango wilt disease in Oman and Pakistan. Fungal Divers. 27**,** 213-230.

Vleeshouwers, V.G.A.A., Oliver, R.P., 2014. Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. Mol. Plant-Microbe Interact. 27**,** 196-206.

Voorrips, R.E., 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93**,** 77-78.

Waalwijk, C., Taga, M., Zheng, S.-L., Proctor, R.H., Vaughan, M.M., O'Donnell, K., 2018. Karyotype evolution in *Fusarium*. IMA Fungus. 9**,** 13-33.

White, T.J., Bruns, T., Lee, S., Taylor, J., Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols: A sequencing guide to methods and applications. Academic Press, San Diego, USA, 1990, pp. 315-322.

Wingfield, M.J., Brockerhoff, E.G., Wingfield, B.D., Slippers, B., 2015. Planted forest health: The need for a global strategy. Science. 349**,** 832-836.

Wittenberg, A.H.J., van der Lee, T.A.J., Ben M'Barek, S., Ware, S.B., Goodwin, S.B., Kilian, A., Visser, R.G.F., Kema, G.H.J., Schouten, H.J., 2009. Meiosis drives extraordinary genome plasticity in the haploid fungal plant pathogen *Mycosphaerella graminicola*. PLOS ONE. 4**,** e5863.

Yang, D.-D., de Billerbeck, G.M., Zhang, J.-J., Rosenzweig, F., Francois, J.-M., 2018. Deciphering the origin, evolution and physiological function of the subtelomeric aryl-alcohol dehydrogenase gene family in the yeast *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 84**,** e01553-17.

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.

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

Zhang, H., Meltzer, P., Davis, S., 2013. RCircos: an R package for Circos 2D track plots. BMC Bioinformatics. 14**,** 244-249.

Supplementary tables

Table S1. Genes located in the QTL for growth rate on MEA in *C. manginecans* scaffold 5

Table S2A. Genes located in QTL1 for aggressiveness on *I. batatas* in *C. manginecans* scaffold 26

Table S2B. Genes located in QTL2 for aggressiveness on I. batatas in C. manginecans scaffold 26

Table S3. Genes located in the QTL for aggressiveness on *A. mangium* in *C. manginecans* scaffold 12