Abundant secreted hydrolytic enzymes and secondary metabolite gene clusters in genomes of the *Botryosphaeriaceae* reflect their role as important plant pathogens

JH Nagel, MJ Wingfield and B Slippers

Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0001 South Africa

Abstract

The *Botryosphaeriaceae* are important plant pathogens, but unique in their ability to establish asymptomatic infections that persist for extended periods in a latent state. In this study, we used comparative analyses to consider elements that might shed light on the genetic basis of the interactions of these fungi with their plant hosts. For this purpose, we characterised secreted hydrolytic enzymes, secondary metabolite biosynthetic gene clusters and considered general trends in genomic architecture using all available *Botryosphaeriaceae* genomes, and selected Dothideomycetes genomes. The *Botryosphaeriaceae* genomes were rich in carbohydrate-active enzymes (CAZymes), proteases, lipases and secondary metabolic biosynthetic gene clusters (BGCs) compared to other Dothideomycete genomes. The genomes of *Botryosphaeria*, *Macrophomina*, *Lasiodiplodia* and *Neofusicoccum*, in particular, had gene expansions of the major constituents of the secretome, notably CAZymes involved in plant cell wall degradation. The *Botryosphaeriaceae* genomes were shown to have moderate to high GC contents and most had low levels of repetitive DNA. The genomes were not compartmentalized based on gene and repeat densities, but genes of secreted enzymes were slightly more abundant in gene-sparse regions. The abundance of secreted hydrolytic enzymes and secondary metabolite BGCs in the genomes of *Botryosphaeria*, *Macrophomina*, *Lasiodiplodia*, and *Neofusicoccum* were similar to those in necrotrophic plant pathogens, but also endophytes of woody plants. The results provide a foundation for future comparative genomic analyses and hypothesis to explore the mechanisms underlying *Botryosphaeriaceae* host-plant interactions.

Keywords: secretome, CAZyme, secondary metabolism, comparative genomics, endophyte, plant cell wall-degrading enzymes

1. Introduction

Secreted hydrolytic enzymes and fungal toxins play crucial roles in enabling fungal pathogens to establish successful infections on their plant hosts. Among the secreted proteins, carbohydrate-active enzymes (CAZymes), protease and lipases are important for nutrient acquisition, as well as for the breakdown, manipulation or circumvention of host defences (Carlile *et al.* 2000; Christensen and Kolomiets 2011; De Jonge and Thomma 2009; Haridas *et al.* 2020; Jia *et al.* 2000; Ohm *et al.* 2012; Voigt *et al.* 2005; Zhao *et al.* 2013). Fungal toxins are a diverse group of compounds and those most commonly found in fungal pathogens include polyketides, non-ribosomal peptides, terpenes and indole alkaloids (Keller *et al.* 2005). These toxins are secondary metabolites that induce plant cell death and for this reason, necrotrophic plant pathogens usually possess greater numbers of genes involved in secondary metabolite synthesis than biotrophic pathogens (Howlett 2006).

The genomes of many fungal and Oomycetes plant pathogens, especially those rich in repetitive elements, are not homogenous, but rather compartmentalized into repeat-rich, gene sparse regions and repeat poor, gene dense regions (Dong *et al.* 2015; Frantzeskakis *et al.* 2019; Ma *et al.* 2010; Raffaele *et al.* 2010; Raffaele and Kamoun 2012; Rouxel *et al.* 2011). Genes localized to repeat-rich, gene sparse regions also have a higher rate of mutation and are often under stronger selective pressure (Möller and Stukenbrock 2017; Raffaele and Kamoun 2012; Stukenbrock *et al.* 2010). This has given rise to a phenomenon referred to as ‘two-speed’ genomes, due to the stark differences in evolutionary rates between the two different types of genomic regions.

Fungi residing in the *Botryosphaeriaceae* include important plant pathogens. These fungi mostly cause diseases of woody plant species and they can impact negatively on the health of many economically and ecologically significant plant species (Mehl *et al.* 2013; Slippers and Wingfield 2007). The *Botryosphaeriaceae* infect a wide range of plant hosts, most notably grapevine (Urbez-Torres 2011), pome and stone fruits (Slippers *et al.* 2007), plantation forest trees such as *Eucalyptus* spp., *Pinus* spp. and *Acacia mangium* (Alves *et al.* 2013; Mohali *et al.* 2007; Rodas *et al.* 2009), as well as plants in their native habitats (Jami *et al.* 2014; Jami *et al.* 2017; Marincowitz *et al.* 2008;

Pavlic *et al.* 2008). Many of these fungi (e.g. *B. dothidea*, *M. phaseolina*, *Lasiodiplodia theobromae*, *Neofusicoccum parvum*) have wide host ranges, however, some species (e.g. *Diplodia sapinea*) have narrower host ranges or are even very host-specific such as in the case of *Eutiarospora darliae*, *E. pseudodarliae* and *E. tritici-australis*. Many species of *Botryosphaeriaceae* are also known to occur endophytically in asymptomatic plant tissues or to have a latent pathogenic phase, where they inhabit their plant hosts in the absence of symptoms and cause disease only after the onset of stress, such as drought, frost or hail damage (Bihon *et al.* 2011; Slippers and Wingfield 2007).

A few recent studies have investigated secreted proteins and secondary metabolites in species of the *Botryosphaeriaceae*. Proteomic studies analyzing the secreted proteins of *Diplodia seriata* (Cobos *et al.* 2010) and *D. corticola* (Fernandes *et al.* 2014) identified secreted proteins involved in pathogenesis. Studies of grapevine pathogens also predicted secreted CAZymes and genes involved in the production of secondary metabolites of *D. seriata* and *Neofusicoccum parvum* (Massonnet *et al.* 2016; Morales-Cruz *et al.* 2015). Secondary metabolite biosynthetic gene clusters (BGCs) have also been shown to play a role in host range determination, e.g. in *E. darliae* and *E. pseudodarliae* causing white grain disorder, where the presence of a secondary metabolite biosynthetic gene cluster allows woody hosts to be infected (Thynne *et al.* 2019). Despite the many publicly available genomes of species of *Botryosphaeriaceae* (Blanco-Ulate *et al.* 2013; Islam *et al.* 2012; Liu *et al.* 2016; Marsberg *et al.* 2016; Massonnet *et al.* 2016; Morales-Cruz *et al.* 2015; Robert-Siegwald *et al.* 2017; van der Nest *et al.* 2014), no comprehensive comparative studies have been undertaken using these genomes; neither have analyses been conducted to characterise secreted proteins and secondary metabolites in most these fungi. Such studies are also hampered by the lack of publically available genome annotations.

The manner in which plants interact with beneficial microorganisms, while at the same time restricting the negative effects of pathogens, is an important and intriguing question in plant biology (Southworth 2012). One proposed model referred to as the ‘balanced antagonism model’ (Schulz and Boyle 2005) holds that endophytism is a result of both the host plant and the fungus employing

antagonistic measures against each other, in such a way that neither overwhelms the other. Disruption of this balance either results in the pathogen causing disease or in the host plant successfully killing the fungus. The model thus predicts that known endophytic species should have similar genetic repertoires to their closely related plant pathogenic relatives. This appears to be the case when considering recent comparative genomics studies conducted on endophytic fungi (Hacquard *et al.* 2016; Schlegel *et al.* 2016; Xu *et al.* 2014; Yang *et al.* 2019), although some endophytic species, e.g. *Xylonia heveae* had fewer CAZymes than expected and were more similar to mutualistic species (Gazis *et al.* 2016). Indeed, the above-mentioned endophytes (other than *X. heveae*) commonly had high numbers of plant cell wall degrading enzymes and secondary metabolite genes.

Despite their ubiquity as endophytes and their importance as latent pathogens, very little is known regarding how *Botryosphaeriaceae* species interact with their plant hosts at a molecular level. Key questions in this regard relate to the secreted hydrolytic enzymes and secondary metabolic biosynthesis genes present in their genomes. Based on the results of previous studies on Ascomycetes that are endophytes of woody plants, we have hypothesised that these genes and gene clusters in the *Botryosphaeriaceae* will resemble those of closely related plant pathogens. To test this hypothesis, we compared the secreted hydrolytic enzyme and secondary metabolite genes of *Botryosphaeriaceae* species with those of other Dothideomycetes. We also characterised the genome architecture of the *Botryosphaeriaceae* in terms of gene density, repeat content and prevalence of repeat-induced point mutations (RIP), and considered how these associate with secreted hydrolytic enzymes and secondary metabolite BGCs.

2. Materials and Methods

2.1. Genomic data

All available, published *Botryosphaeriaceae* genomes were retrieved from public databases (NCBI and JGI). Additionally, we sequenced and assembled 12 genomes representing three *Lasiodiplodia* spp. and five *Neofusicoccum* spp. (Table 1). To standardize protein annotations for downstream application, all of the above *Botryosphaeriaceae* genomes were annotated using the same pipeline

described below. Additionally, the genomes and protein annotations of 41 Dothideomycetes and *Aspergillus nidulans* (Eurotiales), which had both genomic sequences and annotated protein sequences available on NCBI, were retrieved (Table 1). These genomes were used for comparative purposes in the phylogenomic analyses, hydrolytic enzyme and secondary metabolite BGC analyses and in the statistical clustering analyses, described below.

2.1.1. DNA extraction, genome sequencing and assembly

Cultures of three *Lasiodiplodia* and five *Neofusicoccum* species (Table 1) were inoculated onto cellophane covered 2 % malt extract agar (MEA; Biolab, Merck) and incubated at 22 °C. After five days, mycelium was harvested from the surface of the cellophane using a sterile scalpel and DNA was extracted using a modified phenol/chloroform protocol. Mycelium was ground to a fine powder using liquid Nitrogen and approximately 500 mg of ground mycelium was used for DNA extraction. To the ground mycelia, 18 ml of a 200 mM Tris-HCl (pH 8.0), 150 mM NaCl, 25 mM EDTA (Ethylenediaminetetraacetic acid, pH 8.0) and 0.5% SDS (Sodium dodecyl sulfate) solution and 125 µl of 20 mg/ml Proteinase K was added and incubated at 60 °C for two hours. This was followed by addition of 6 ml of 5M potassium acetate and 30 minutes incubation at 0 °C. Samples were then centrifuged at 5000 g for 20 minutes. The aqueous phase was kept and 24 ml of a 1:1 phenol:chloroform solution was added; samples were then centrifuged as above. Two chloroform washes were performed on the aqueous phase, followed by addition of 100 µl of 10 mg/ml Rnase A and incubated for two hours. DNA was precipitated using one volume isopropanol and centrifuged for 30 minutes. The pellet was cleaned through two 70% ethanol washes and resuspended in 1 x Tris-EDTA buffer.

The extracted DNA was used for paired-end sequencing (average fragment size of 500 bp). All samples were sequenced on an Illumina HiSeq 2500 platform, except for *N. parvum* (isolate CMW9080) that was sequenced on a Miseq platform. The quality of the resulting reads was assessed using FastQC 0.10.1 (Andrews 2010) and low quality and short reads were trimmed or discarded using Trimmomatic 0.30 (Bolger *et al.* 2014). *De novo* genome assembly was performed Velvet

1.2.10 (Zerbino and Birney 2008) and Velvetoptimiser 2.2.5 (Gladman and Seemann 2012). Paired-end reads were used to scaffold the assembly and insert size statistics were determined by Velvet for each genome assembly. Genome assembly summary statistics were calculated with the AssemblyStatsWrapper tool of BBtools 38.00 (Bushnell *et al.* 2017).

2.1.2. Genome annotation

The twelve sequenced genomes described above, as well as the fourteen *Botryosphaeriaceae* genomes retrieved from public databases, were annotated as follows: Custom repeat libraries were constructed for each genome assembly using RepeatModeler 1.0.10 (Smit and Hubley 2017). BRAKER 1.10 (Hoff *et al.* 2015) was used to create trained GeneMark-ET 4.29 (Lomsadze *et al.* 2014) and AUGUSTUS 3.2.3 (Stanke *et al.* 2006) profiles using previously published *N. parvum* transcriptome data (SRR3992643 and SRR3992649) (Massonnet *et al.* 2016). Genomes were annotated through the MAKER2 2.31.8 (Holt and Yandell 2011) pipeline using the custom repeat libraries and the BRAKER trained GeneMark-ET 4.29 and AUGUSTUS profiles. Genomes and genome annotations were assessed for completeness with BUSCO 4.0.5 (Waterhouse *et al.* 2017) using the Ascomycota ortholog library (Creation date 2020-09-10, 1706 core orthologous genes). The annotations for six *Botryosphaeriaceae* genomes available on public databases prior to this study were also assessed using BUSCO and compared to those of the annotations generated in this study.

2.2. Phylogenomic analyses

To illustrate the relationships between species and genera of the *Botryosphaeriaceae*, as well as the relationship of this family to the rest of the Dothideomycetes, a robust phylogeny was created from the genome data. Single copy core orthologous genes were identified from individual genomes (Table 1) using BUSCO (as described above). The BUSCO genes present in all taxa (207 genes) were selected for further analysis. Each orthogroup was aligned using MAFFT 7.407 (Kato and Standley 2013) before being concatenated into a single matrix. RAxML 8.2.4 (Stamatakis 2014) was used to perform maximum likelihood phylogenetic inference using the PROTGAMMAAUTO option and a

thousand bootstrap replicates were performed. Trees were rooted using sequences from *Aspergillus nidulans*.

2.3. Functional annotation

The genome annotation data for the *Botryosphaeriaceae*, the other Dothideomycetes and the outgroup *A. nidulans* were used to perform functional annotation for hydrolytic enzymes and secondary metabolite BGCs. CAZymes were predicted by searching the total predicted proteins of each genome against the CAZy database (Lombard *et al.* 2013) using dbCAN2 (HMMdb release v9.0) (Zhang *et al.* 2018). Only those CAZyme predictions supported by two or more tools (HMMR, DIAMOND, Hotpep) were retained. Proteases and protease inhibitors were predicted by subjecting the predicted protein sequences to a BLASTP (Camacho *et al.* 2009) search against the MEROPS protease database 12.0 (Rawlings *et al.* 2017) using a cut off E-value of 1E-04. Lipases and cutinases were predicted by searching protein sequences against lipase and cutinase hidden Markov model (HMM) profiles retrieved from the Lipase Engineering Database v 3.0 (Fischer and Pleiss 2003) using HMMER 3.1b2 (Eddy 2011). Secondary metabolite BGCs were identified from all annotated genomes. using AntiSMASH v5.1.1 (Blin *et al.* 2019).

The total predicted proteins were also analyzed for the presence of signal peptides, involved in protein secretion. Phobius 1.01 (Käll *et al.* 2007) and SignalP 4.1 (Petersen *et al.* 2011) were used to assess the presence of signal peptides and TMHMM 2.0 (Krogh *et al.* 2001) was used to determine if any transmembrane regions occurred within these proteins. Only proteins with a signal peptide predicted by both Phobius and SignalP, as well as no transmembrane domains outside of the signal peptide predicted by both Phobius and TMHMM were regarded here as predicted secreted proteins.

2.4. Analysis of gene family evolution

CAFE (Computational Analysis of gene Family Evolution) v4.2 (Han *et al.* 2013) was used to study the evolution of gene family size in the hydrolytic enzymes and secondary metabolite BGCs. To this end, the phylogeny described above was converted to an ultrametric tree using r8s v1.81 (Sanderson

2003). This was then calibrated by fixing the age of the *Botryosphaeriaceae* to 61 million years (Phillips *et al.* 2019) and constraining the age of the Dothideomycetes to 303-357 million years (Hyde *et al.* 2017). Gene family sizes for the total predicted CAZymes, proteases, lipases and secondary metabolite BGCs (Supplementary File 1) were used as input for the CAFE analyses. CAFE was run using separate lambda (birth) and mu (death) rate parameters. Additionally, two separate rate classes were allowed in that the rate parameters were calculated independently for the *Botryosphaeriaceae* and the remaining Dothideomycete taxa. CAFE was run using a P-value cutoff of 0.01 and Viterbi P-values were calculated to significant expansions/contractions across branches.

2.5. Hierarchical clustering and principal component analysis

Hierarchical clustering was done to determine the similarity between taxa based on the secreted hydrolytic enzyme classes and secondary metabolite BGC types. The heatmap.2 function of the gplots (Warnes *et al.* 2016) R package was used to perform the analysis. Additionally, principal component analysis (PCA) was performed using the same functional annotations as used in the hierarchical clustering, with the exception that for CAZymes and proteases the number of genes associated with each family (e.g AA1, A01) was used instead of those of each class (e.g AA, Aspartic). The FactoMineR (Lê *et al.* 2008) R package was used to perform the analysis and Factoshiny (Vaissie *et al.* 2015) was used to generate PCA plots.

2.6. Genome architecture

Gene densities were analyzed for each *Botryosphaeriaceae* genome to assess the level of genome compartmentalization. Intergenic distances were used as a measure of gene density, by considering the 5' and 3' flanking intergenic regions (FIRs) of each gene. FIR lengths were used for two-dimensional data binning to construct gene density heat maps (Saunders *et al.* 2014) using R (Team 2013). Possible differences in the gene density distributions between the total and secreted CAZymes, proteases and lipases were also considered. We compared the relative amounts (i.e. the percentage) of these genes located in gene-sparse regions. In this case, gene sparse regions were defined as genes with both 5' and 3' FIRs larger than 1500 bp, as previously used by Raffaele and Kamoun (2012)

The repeat contents of the *Botryosphaeriaceae* genomes were determined using RepeatMasker (Smit *et al.* 2017) with custom repeat libraries created by RepeatModeler. The presence of TA rich regions and the genes present therein were determined using OcculterCut (Testa *et al.* 2016). Specifically, the numbers of the total secreted genes, the secreted CAZymes, proteases and lipases, as well as the number of secondary metabolite BGCs associated with TA rich regions were determined. The occurrence of RIP, including large regions affected by RIP (LRARs), was determined in each genome using TheRIPper (Van Wyk *et al.* 2019)

3. Results

3.1. Genome sequencing, assembly and annotation

Nine genomes of *Neofusicoccum* and three genomes of *Lasiodiplodia* species were sequenced using Illumina sequencing. These included two isolates each of *N. cordaticola*, *N. kwambonambiense*, *N. parvum* and *N. ribis* were sequenced. A single isolate was sequenced for *L. gonubiensis*, *L. pseudotheobromae*, *L. theobromae* and *N. umdonicola*.

De novo genome assembly resulted in genome lengths of approximately 43 MB for both *Lasiodiplodia* spp. and *Neofusicoccum* spp. (Table 2). The number of scaffolds/contigs was very variable between the sequenced genomes, but the three *Lasiodiplodia* genomes had a lower number of scaffolds than the *Neofusicoccum* genomes. The *N. parvum* CMW9080 genome that was sequenced on the Miseq platform had a higher degree of fragmentation, as seen from the high total number of scaffolds and a large number of short contigs and scaffolds. The percentage of repetitive elements of each genome was higher on average in the *Neofusicoccum* genomes (6.84%) than in the *Lasiodiplodia* genomes (3.33%).

Twenty-six *Botryosphaeriaceae* genomes were annotated by predicting protein-coding genes with MAKER using BRAKER trained profiles. BUSCO analysis (Table 3) indicated that all *Botryosphaeriaceae* genomes had a high degree of completeness (average of 98.%, minimum of

95.1%). The genome annotations that were generated also had a high BUSCO completeness score (average of 97.8%, minimum of 94.3%). When comparing these BUSCO results to those of species with existing genome annotations on NCBI/JGI, it was clear that in five out of the six cases the genome annotations from the present study were more complete than others.

3.2. Phylogenomic analyses

We identified 207 core orthologous genes from the collection of 26 *Botryosphaeriaceae*, 39 other Dothideomycetes and the outgroup (*Aspergillus nidulans*) genomes. Only orthologous genes that were represented by a single gene per species were retained. The results of the phylogenomic analyses corresponded well to previous phylogenies for the Dothideomycetes (Schoch *et al.* 2009; Schoch *et al.* 2006; Zhang *et al.* 2011). The phylogeny indicated the early divergence of the Dothideomycetidae (Dothideales, Capnodiales and Myrangiiales) from the lineage containing the Pleosporomycetidae (Pleosporales, Hysteriales and Mytilinidiales) and other Dothideomycetes without current subclass designation (Figure 1). This phylogeny further supported the early divergence of the Botryosphaeriales from the ancestral Dothideomycetes lineage after the divergence of the Dothideomycetidae and Venturiales. The phylogenetic relationships between the *Botryosphaeriaceae* were well defined and the phylogenetic placement of species and genera corresponded with that found in previous studies (Slippers *et al.* 2013; Yang *et al.* 2017).

3.3. Functional annotation

The genomes of the *Botryosphaeriaceae* differed noticeably in the number of predicted secreted proteins, hydrolytic enzymes and secondary metabolite BGCs they encode. However, distinct trends became apparent when considering the genus to which these genomes belong. Invariably, *Eutiarospora* spp. had the fewest numbers of each of these functional annotation categories that we considered, followed by *Diplodia* spp. and *Botryosphaeria*, *Lasioidiplodia*, *Macrophomina*, and *Neofusicoccum* species had the most of these genes (Figure 1). This trend was also evident when considering the different classes of CAZymes/proteases/lipases and the different types of secondary metabolite BGCs (Figure 2).

The *Botryosphaeriaceae*, on average, had more of each CAZyme class than the Dothideomycetidae and a greater or roughly equal amounts than the Pleosporomycetidae, including the three *incertae sedis* taxa (Figure 2). The exception was for the glycosyltransferases (GT), where the *Botryosphaeriaceae* had fewer genes than the Dothideomycetidae and Pleosporomycetidae. The *Botryosphaeriaceae* were furthermore particularly rich in polysaccharide lyase (PL) genes, both the total number encoded in the genome and within the secretome. The most abundant secreted CAZyme families in the *Botryosphaeriaceae* were CBM1, AA3, GH3, GH43, GH5, AA9, CBM18, AA1, GH28 and CBM13 (Supplementary File 1).

The *Botryosphaeriaceae* had above-average numbers of aspartic- (A), metallo- (M) and serine- (S) proteases, especially in the species of *Botryosphaeria*, *Lasiodiplodia*, *Macrophomina*, and *Neofusicoccum* (Figure 2). These three protease classes were also the dominant proteases in the secretome. Furthermore the *Botryosphaeriaceae* had fewer than average number of cysteine (C) proteases and protease inhibitors (I) than the other Dothideomycetes. Notably, the *Botryosphaeriaceae* possessed a single secreted protease inhibitor family, namely I51.001 (serine carboxypeptidase Y inhibitor), whereas many other Dothideomycetes secreted protease inhibitors were of this family, as well as I09.002 (peptidase A inhibitor 1) or I09.003 (peptidase B inhibitor). *Diplodia sapinea* and *D. scrobiculata* had no secreted protease inhibitors. The most abundant secreted protease families among the *Botryosphaeriaceae* were S09, A01, S10, S08, M28, S53, S33, M43, M35, S12 (Supplementary File 1).

In the *Botryosphaeriaceae* and the Dothideomycetes, the most abundant lipases/lipase-like families were abH04 (Moraxella lipase 2 like), abH03 (Candida rugosa lipase-like), abH36, (cutinase) and abH23 (Filamentous fungi lipases) (Figure 2). The abH03, abH36 and abH23 lipase families were the main constituents of the predicted secretomes among the Dothideomycetes and the *Botryosphaeriaceae* had above-average numbers for these families.

The *Botryosphaeriaceae* genomes were rich in gene clusters involved in the synthesis of secondary metabolites. Type 1 polyketide synthases (t1PKS) were the most abundant type of gene cluster, followed by non-ribosomal peptide synthetases (NRPS) and NRPS-like, terpene synthases (TS) and t1PKS-NRPS hybrid clusters. The genomes of *Botryosphaeria*, *Lasiodiplodia*, *Macrophomina*, and *Neofusicoccum* contained higher than average numbers of t1PKS, NRPS-like, NRPS, and TS BGCs (Figure 3). Additionally, the *Botryosphaericeae* genomes contained betalactone BGCs (Supplementary File 1). The genomes of *Diplodia* and *Eutiarosporella* contained fewer than average of these gene clusters. Certain Dothideomycetes genomes, predominantly those of the Pleosporomycetidae also contained indole and type 3 PKS BGCs, however, these were not present in any of the *Botryosphaeriaceae* genomes.

In the *Botryosphaeriaceae*, both genome size and the total gene number were strongly correlated with the number of secreted proteins, CAZymes, proteases, lipases and secondary metabolite gene clusters (Supplementary File 1). Within the Dothideomycetes, however, the numbers of secreted proteins, CAZymes, proteases, lipases, and secondary metabolite BGCs present within a genome were correlated to one another (i.e. species that contained large numbers of secreted proteins also contained large numbers of CAZymes, proteases, lipases and secondary metabolite BGCs), but only weakly correlated to genome size and the total number of genes (Supplementary File 1).

3.4. Gene family evolution

These results of the CAFE analyses (Supplementary file 2) indicated that expansions and contractions of CAZyme gene families occurred at roughly similar levels. This was after the divergence of the Botryosphaeriales ancestor from the Pleosporomycetidae until the formation of the *Botryosphaeriaceae* crown group (61 MYA). During this time, protease gene families apparently experienced more contractions than expansions, lipase gene families had slightly more expansions than contractions and secondary metabolite BGCs experienced a large amount of gene family contractions. Several CAZyme gene families (AA1, AA3, AA7, AA8, AA9, CBM1, CBM18, CE4, CE5, GH3, GH10, GH28, GH43, GH78, GT1, GT2, GT25, PL1 and PL3) experienced rapid

expansion (i.e. greater than expected under the birth/death model of gene family evolution) prior to the divergence of the *Botryosphaeriaceae* crown group.

After the divergence of the *Botryosphaeriaceae*, the genera *Botryosphaeria*, *Lasiodiplodia*, *Macrophomina*, and *Neofusicoccum* experienced more gene family expansions than contractions, whereas the opposite was observed for the *Diplodia* and *Eutiarospora*. Among the *Neofusicoccum* spp., the AA3, AA7, GH3 and GT2 gene families were rapidly expanding. Similarly, among the *Lasiodiplodia* spp., the AA7 and GH106 gene families were rapidly expanding. Conversely, among the *Diplodia* spp. the AA7 gene family was rapidly contracting, as were the AA1, AA3, AA7, GH28, PL1 and PL3 gene families among the *Eutiarospora* spp.

3.5. Principal component analysis and hierarchical clustering

Hierarchical clustering separated the taxa into four groups (Supplementary File 3). The taxa in the Pleosporomycetidae and Dothideomycetidae generally clustered separately, however there was no overall clustering based on taxonomic placement. *Botryosphaeriaceae* species were present in three of the four dominant clusters.

A first cluster included *B. dothidea*, *B. kuwatsukai*, *M. phaseolina*, *L. theobromae*, *L. pseudotheobromae*, and *Neofusicoccum* spp., as well as several Pleosporales (*Alternaria alternata*, *Clohesyomyces aquaticus*, *Corynespora cassicola*, *Paraphaeosphaeria sporulosa* and *Periconia macrospinosus*). A second cluster mostly contained taxa from the Pleosporales (*Aschochyta rabiei*, *Bipolaris* spp., *Epicoccum nigrum* and *Stemphylium lycopersici*), but also contained *D. seriata*, *D. corticola* and *L. gonubiensis*. A third cluster contained taxa from both the Dothideomycetidae and Pleosporomycetidae, as well as *D. sapinea*, *D. scrobiculata* and *Eutiarospora* spp. A fourth cluster was dominated by taxa from the Dothideomycetidae with the exception of *L. palustris*, *C. geophilum*, *C. apollinis* and *V. gallopava*.

PCA of the functional annotation categories clustered the data along 65 dimensions/principal components. The first two dimensions (Figure 3) accounted for 29.9% of the variance among the taxa. The first dimension accounted for 17.4% and the second dimension for 12.5% of the variance. The first dimension was most strongly influenced by a number of secreted CAZyme (AA3, CBM1, GH131, PL3, CBM13, CBM18, CE8, CE12, AA7, GH43, PL4, PL1, CE5 and CBM63), cutinase (abH36) and lipase (abH03 and abH23) and protease (S09 and A01) families, as well as terpene BGCs. The second dimension was most strongly influenced by CAZyme (GH145, CBM3, GH6, GH11, PL26, CBM60, GH7, AA12, AA9, CBM2, GH16, CBM6, CBM87 and CE18), protease (S01, M14 and M36) families, as well as the indole and T3PKS BGC types.

The *Botryosphaeriaceae* were distributed mainly along the first dimension of the PCA and clustered into three groups. *Eutiarospora* spp. clustered at the lower ranges of the first dimension (x-axis) followed by clusters accommodating the *Diplodia* species towards the middle ranges and the other genera of *Botryosphaeriaceae* clustered at the high ranges of the x-axis. The *Botryosphaeriaceae* clustered along a relatively narrow range along the second dimension (y-axis) compared to the other Dothideomycetes. The clustering of the other Dothideomycetes along the x-axis was correlated to their clustering on the y-axis: taxa towards the higher end of the x-axis also occurred towards the higher end of the y-axis. The clustering of taxa did not correspond to their nutritional lifestyle, but their taxonomic placement was reflected in their clustering.

3.6. Genome Architecture

Two-dimensional heatmaps of the 5' and 3' FIRs of the 26 *Botryosphaeriaceae* genomes indicated no genome compartmentalization (Figure 4 and Supplementary File 4). This was evident from the unimodal gene density distributions of these genomes. Genomes of *Botryosphaeria*, *Lasiodiplodia* and *Macrophomina* had a higher proportion of genes in gene sparse regions than the other species in this family. *Eutiarospora* spp., *D. sapinea*, *D. scrobiculata* and *Neofusicoccum* spp. had fewer genes in gene sparse regions.

The predicted secreted proteins of the *Botryosphaeriaceae* contained a greater number of genes in gene sparse regions than the total predicted genes (Supplementary File 4). The total CAZymes contained a higher proportion of genes in gene sparse regions than the total predicted genes. The secreted CAZyme gene density distribution was very similar to that of the total CAZymes. The secreted lipases and cutinases, however, occurred more frequently in gene sparse regions than the total lipases and cutinases (Supplementary File 4). This trend was also observed for the secreted proteases, although not as strongly. Genes associated with secondary metabolite BGCs were less prevalent in gene sparse regions.

The levels of repetitive sequences for most *Botryosphaeriaceae* genomes were between 3 and 8 percent of the total genome size (Table 4). The two *Botryosphaeria* spp. differed considerably in their repeat content (3.48 vs. 11.88%). The genome of *M. phaseolina* also had a higher than average repeat content (16.37%). Among the *Neofusicoccum* species the genomes of *N. parvum* and *N. undoncola* had less repetitive sequences than the other genomes of this genus. The genomes of *D. sapinea* and *D. scrobiculata* also contained less repetitive sequences than the other two *Diplodia* species.

On average, the *Botryosphaeriaceae* genomes had less than 10% of their genomes composed of TA rich regions (GC<50%) (Table 4). The genomes of *B. kuwatsukai* LW030101 and *M. phaseolina* had 18.9% and 19.8% of their genomes as TA rich regions, respectively. The genomes of *Diplodia* and *Eutiarospora* had fewer TA rich regions (approximately 5% of the genome) compared to the other genera. Less than 5% of genes were present in TA rich regions and secreted genes were not found to be over-represented among these genes (Supplementary file 4). The genomes of *Diplodia* and *Eutiarospora* had considerably fewer genes associated with TA rich regions, than the other taxa of this family.

Analysis of the prevalence of RIP in the genomes of *Botryosphaeriaceae* indicated that this has occurred to varying degrees in these genomes (Table 4). *Neofusicoccum* spp. had between 1.36 and 4.71 % of their genome affected by RIP. The level of RIP was similar between different genomes of

the same *Neofusicoccum* species. *Neofusicoccum parvum* and *N. umdonicola* had lower proportions of RIP affected sequences than the other species of the genus. There was a large (>10-fold) difference in the level of RIP between *B. dothidea* and *B. kuwatsukai*. The genome of *M. phaseolina* had the highest (12.92 %) amount of RIP of all the *Botryosphaeriaceae*. *Lasiodiplodia* spp. had RIP levels between 0.72 and 1.18 %. The level of RIP in the genomes of *D. sapinea* and *D. scrobiculata* was lower than in *D. seriata* and *D. corticola*. The genome of *E. tritici-australis* had more than double the level of RIP affected regions than the other two species of the genus.

4. Discussion

This study represents the first large-scale comparative genomics-level consideration of all available genomes of *Botryosphaeriaceae*. The results showed that the included *Botryosphaeriaceae* genomes, especially those of *Botryosphaeria*, *Macrophomina*, *Lasiodiplodia*, and *Neofusicoccum*, encode high numbers of secreted hydrolytic enzymes and secondary metabolite BGCs. This apparently emerges due to these fungi having large numbers of genes belonging to enzyme classes and secondary metabolite types involved in interactions with plants. The results also indicate that the *Botryosphaeriaceae* are most similar to species of the Pleosporamycetidae based on secreted enzyme and secondary metabolite profiles and that nutritional lifestyle could not be deduced from the data. *Botryosphaeriaceae* genomes were furthermore determined not be be compartmentalized based on gene density or GC-content.

There was a strong correlation between the number of hydrolytic enzymes and secondary metabolite BGCs, and the genome size and gene number of the *Botryosphaeriaceae* considered in this study. This correlation between genome size and gene number has generally not been seen in other fungi (Mohanta and Bae 2015; Ohm *et al.* 2012; Raffaele and Kamoun 2012) because transposable elements and repetitive DNA vary significantly among species (Raffaele and Kamoun 2012). A recent comparison of Dothideomycetes genomes (Haridas *et al.* 2020) also showed that genome size and gene number were not correlated to the abundance of functional annotation classes; neither to the lifestyle or phylogenetic placement of a species.

The genomes of *Botryosphaeria*, *Macrophomina*, *Lasiodiplodia*, and *Neofusicoccum* had abundant secreted hydrolytic enzyme and secondary metabolite BGCs. This is a pattern that is most similar to prominent necrotrophic plant pathogens (*A. alternata*, *C. casicola*), saprobes (*C. aquaticus*, *P. sporulosa*) and the endophyte/latent pathogen *P. macrospinosa* in the Pleosporales. The pattern consistent with reports that necrotrophic pathogens tend to have higher numbers of hydrolytic enzymes and secondary metabolite toxins than biotrophs and symbiotic fungi (Lyu *et al.* 2015; Zhao *et al.* 2014). An abundance of secreted hydrolytic enzymes and secondary metabolite BGCs found in the *Botryosphaeriaceae* is also similar to that of other species of woody endophytes. Studies on such endophytic species have shown that they have similar or higher amounts of various secreted enzymes (notably plant cell wall degrading enzymes) or secondary metabolites than closely related plant pathogenic species (Hacquard *et al.* 2016; Schlegel *et al.* 2016; Xu *et al.* 2014; Yang *et al.* 2019). The specific gene families that are enriched, however, differ among endophytic lineages, due to the evolutionary independent origins of endophytism (Hacquard *et al.* 2016; Schlegel *et al.* 2016; Xu *et al.* 2014; Yang *et al.* 2019). It has furthermore been noted that fungi with dual lifestyles (e.g. fungi with endophytic and pathogenic phase) have large numbers of CAZymes (Queiroz and Santana 2020; Wang *et al.* 2015). These observations also emerging from the present study are consistent with the known lifestyle of *Botryosphaeriaceae* as latent pathogens.

The *Botryosphaeriaceae* genomes were rich in CAZymes, especially those involved in plant cell wall degradation (PCWD), although at lower numbers in the genomes of *Diplodia* and *Eutiarospora* species. CAZymes involved in the degradation of cellulose, hemicellulose and pectin were present in all *Botryosphaeriaceae*. The genomes of *Botryosphaeria*, *Macrophomina*, *Lasiodiplodia* and *Neofusicoccum* were particularly rich in CAZyme families involved in cell wall degradation. Specifically, CAZymes involved in plant, general and fungal cell wall degradation (Levasseur *et al.* 2013; Ohm *et al.* 2012; Zhao *et al.* 2013) were abundant in the genomes of the above-mentioned genera.

The *Botryosphaeriaceae* secretomes were rich in CAZyme families involved in the recognition of cellulose (CBM1) and chitin (CBM18). Although carbohydrate-binding domains have no catalytic activity of their own they play important roles in substrate recognition and binding of other CAZymes (Boraston *et al.* 2004; Christiansen *et al.* 2009), they are also involved in the protection of fungal cell walls from degradation by host enzymes and prevention of host detection (De Jonge and Thomma 2009; Marshall *et al.* 2011). High numbers of secreted CAZymes involved with PCWD have also been found in previous studies of *N. parvum* and *D. seriata* (Massonnet *et al.* 2016; Morales-Cruz *et al.* 2015) and in other, especially necrotrophic, Dothideomycetes (Lopez *et al.* 2018; Manning *et al.* 2013; Verma *et al.* 2016; Zeiner *et al.* 2016; Zhao *et al.* 2013). The abundance of these CAZyme families in some genera of *Botryosphaeriaceae* suggests that cell wall degradation plays an important role in the biology of these fungi.

Several important CAZyme families that are common among Dothideomycetes were absent from all the *Botryosphaeriaceae* genomes, i.e. Acetyl xylan esterase (CE3) (Biely *et al.* 1985), Pyrroloquinoline quinone-dependent oxidoreductase (AA12) (Takeda *et al.* 2015), endo- α -1,4-polygalactosaminidase (GH114) (Bamford *et al.* 2019) and α -L-arabinofuranosidase/ β -xylosidase (GH54) (Miyanaaga *et al.* 2006). The absence of these CAZyme families in the *Botryosphaeriaceae*, does not necessarily indicate a gap in the metabolic repertoire of these fungi because a large degree of functional redundancy is commonly seen in fungal CAZyme repertoires (Couger *et al.* 2015; Couturier *et al.* 2016; Goulet and Saville 2017). Interestingly, some of the CAZyme families that can functionally compensate for the absence of the above-mentioned CAZyme families are those that were found to be among the most abundant secreted CAZyme families of the *Botryosphaeriaceae* (e.g. CE16, AA3, AA7, GH15 and GH3).

The *Botryosphaeriaceae* were rich in secreted serine-, metallo- and aspartic-proteases. Protease families (A01, S08, S09, S10) that were previously identified as the most common secreted proteases among Dothideomycetes (Ohm *et al.* 2012) were also the most abundant in the *Botryosphaeriaceae*. Secreted proteases play important roles in nutrient acquisition, signalling and degradation of plant

defences (Carlile *et al.* 2000; Olivieri *et al.* 2002; Plummer *et al.* 2004; Thon *et al.* 2002). Although secreted proteases are abundant in several necrotrophic pathogens, e.g. *Corynespora cassicola* (Lopez *et al.* 2018) and several *Colletotrichum* spp. (Baroncelli *et al.* 2016), no patterns between nutritional lifestyle and the abundance of secreted proteases could be distinguished. The precise function of most of these proteases in the *Botryosphaeriaceae* are unknown and their role during infection and disease expression remains to be determined.

A lower abundance and diversity of secreted protease inhibitors of the *Botryosphaeriaceae* suggests a reduced capacity and/or need for extracellular enzyme inhibition. Plant pathogenic fungi secrete protease inhibitors to inhibit plant proteases involved in defence responses (Jashni *et al.* 2015) and several protease inhibitors are known virulence factors, e.g. *avr2* of *Cladosporium fulvum* (Rooney *et al.* 2005; van Esse *et al.* 2008) and *Pit2* of *Ustilago maydis* (Mueller *et al.* 2013). However, the exact role of many fungal protease inhibitors, such as those secreted by the *Botryosphaeriaceae* and Dothideomycetes, remains unknown (Dunaevsky *et al.* 2014).

Botryosphaeriaceae, especially species of *Botryosphaeria*, *Macrophomina*, *Lasiodiplodia* and *Neofusicoccum* possessed high numbers of secreted lipases. Three lipase families were present in high numbers in the secretomes of the Dothideomycetes, i.e. *Candida rugosa* lipase-like (abH03), cutinases (abH36 and CE5) and Filamentous fungi lipases (abH23). The *Botryosphaericeae* genomes were rich in secreted enzymes for these three families. Lipases and cutinases are important for fungal penetration of host tissue (Kolattukudy 1985; Voigt *et al.* 2005), growth and adhesion (Feng *et al.* 2009; Gácsér *et al.* 2007) and manipulation of host defences (Christensen and Kolomiets 2011). The abundance of these secreted lipases and cutinases emphasises the importance of their role during the infection process in the *Botryosphaeriaceae*.

The genomes of *Botryosphaeria*, *Macrophomina*, *Lasiodiplodia* and *Neofusicoccum* contained many BGCs, especially t1PKS, NRPS, NRPS-like and TS type clusters. The products produced by most of these clusters are unknown, however, the products of some clusters could be determined. These

compounds included melanin, phytotoxins (ACT-Toxin II, (-)-Mellein), siderophores (dimethylcoprogen) and antioxidants (pyranonigrin E). The ability to produce secondary metabolite toxins have been associated with pathogenic fungi's lifestyle, host range and virulence (Ito *et al.* 2004; Ohm *et al.* 2012; Stergiopoulos *et al.* 2013). Many plant pathogenic fungi have large numbers of secondary metabolite BGCs, e.g. *Bipolaris* spp. (Zaccaron and Bluhm 2017), *Corynespora cassicola* (Lopez *et al.* 2018), *Colletotrichum* spp (Baroncelli *et al.* 2016) and *Pyrenophora teres* (Ohm *et al.* 2012), but so also do fungi with other lifestyles such as the saprobic *Annulohypoxylon stygium* (Wingfield *et al.* 2018), *Hysterium pulicare*, and *Rhytidhysterium rufulum* (Ohm *et al.* 2012). Despite the observation that the total abundance of secondary metabolite BGCs does not predict lifestyle, several fungal toxins are able to modulate a fungal species' host range or virulence, e.g. the AF-toxin of *A. alternata* (Ito *et al.* 2004), the *Hybrid-1,2* and *3* genes of *Eutiarosporella darliae* and *E. pseudodarliae* (Thynne *et al.* 2019) and the HC-toxins of *Bipolaris zeicola* (Wolpert *et al.* 2002).

The *Botryosphaeriaceae* genera were shown to possess non-compartmentalized genomes. These species were characterized as having moderate to high %GC, RIP-affected genomes with low amounts of repetitive DNA, a slight preferential localization of secreted genes to gene sparse regions and no preferential localization of secreted genes to TA rich regions. Most Dothideomycetes do not have compartmentalized genomes, but several important pathogenic species (e.g. *L. maculans* and *Pseudocercospora* spp.), have genomes with high levels of repetitive DNA (often transposable elements) (Chang *et al.* 2016; Rouxel *et al.* 2011) enriched for fast-evolving genes related to pathogenicity or virulence (Chang *et al.* 2016; Grandaubert *et al.* 2014; Testa *et al.* 2016). However, not all rapidly evolving fungal phytopathogens have this characteristic 'two-speed' genome architecture (Frantzeskakis *et al.* 2019). Where 'two-speed' genomes rely on the action of leaky RIP to generate variation for selection to act on, 'one-speed' genomes of fast-evolving plant pathogens rely on the absence of RIP that allows gene duplication/copy number variation to generate variation (Frantzeskakis *et al.* 2019). The *Botryosphaeriaceae* are not like those species with 'two-speed' genomes as they don't have compartmentalized genomes, but RIP is also not completely absent as seen in species with 'one-speed' fast-evolving genomes.

This study is the first large-scale comparative genomics study to consider all available genomes of *Botryosphaeriaceae*. It has illustrated large variability in the secreted hydrolytic enzyme and secondary metabolite biosynthetic repertoire between genera of this family. Furthermore, we have demonstrated similarities between the *Botryosphaeriaceae* and necrotrophic plant pathogens and endophytes of woody plants, emphasising their role as latent pathogens. This study highlights the importance of these genes in the infection biology of *Botryosphaeriaceae* species and their interaction with plant hosts. This knowledge will be useful in future studies aimed at understanding the mechanisms of endophytic infections and how these transition to a pathogenic state. The results should also help to better understand the genetic factors involved in determining the complex question of host range in the *Botryosphaerianaceae*.

5. Acknowledgements

We are grateful to the University of Pretoria, The Department of Science and Technology (DST)/National Research Foundation (NRF) Centre of Excellence in Tree Health Biotechnology and members of the Tree Protection Cooperative Program for financial support of this study. The authors acknowledge the Centre for Bioinformatics and Computational Biology, University of Pretoria and the Centre for High Performance Computing (CHPC), South Africa for providing computational resources to this research project. Lastly, our sincere thanks to Alisa Postma for her guidance during the genome assembly and annotation process and to Dr Tuan Duong for his critical comments on this manuscript.

6. References

- Alves A, Barradas C, Phillips AJL, Correia A (2013) Diversity of Botryosphaeriaceae species associated with conifers in Portugal. *European Journal of Plant Pathology* **135**, 791-804.
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.
- Bamford NC, Le Mauff F, Subramanian AS, Yip P, Millán C, Zhang Y, Zacharias C, Forman A, Nitz M, Codée JDC (2019) Ega3 from the fungal pathogen *Aspergillus fumigatus* is an endo- α -1, 4-galactosaminidase that disrupts microbial biofilms. *Journal of Biological Chemistry* **294**, 13833-13849.
- Baroncelli R, Amby DB, Zapparata A, Sarrocco S, Vannacci G, Le Floch G, Harrison RJ, Holub E, Sukno SA, Sreenivasaprasad S (2016) Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC genomics* **17**, 555.
- Biely P, Puls J, Schneider H (1985) Acetyl xylan esterases in fungal cellulolytic systems. *FEBS Letters* **186**, 80-84.
- Bihon W, Burgess T, Slippers B, Wingfield MJ, Wingfield BD (2011) Distribution of *Diplodia pinea* and its genotypic diversity within asymptomatic *Pinus patula* trees. *Australasian Plant Pathology* **40**, 540-548.
- Blanco-Ulate B, Rolshausen P, Cantu D (2013) Draft genome sequence of *Neofusicoccum parvum* isolate UCR-NP2, a fungal vascular pathogen associated with grapevine cankers. *Genome announcements* **1**, e00339-00313.
- Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic acids research* **47**, W81-W87.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.
- Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal* **382**, 769-781.

- Bushnell B, Rood J, Singer E (2017) BBMerge—Accurate paired shotgun read merging via overlap. *PLoS ONE* **12**, e0185056.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.
- Carlile AJ, Bindschedler LV, Bailey AM, Bowyer P, Clarkson JM, Cooper RM (2000) Characterization of SNP1, a cell wall-degrading trypsin, produced during infection by *Stagonospora nodorum*. *Molecular Plant-Microbe Interactions* **13**, 538-550.
- Chang T-C, Salvucci A, Crous PW, Stergiopoulos I (2016) Comparative genomics of the Sigatoka disease complex on banana suggests a link between parallel evolutionary changes in *Pseudocercospora fijiensis* and *Pseudocercospora eumusae* and increased virulence on the banana host. *PLoS Genetics* **12**, e1005904.
- Christensen SA, Kolomiets MV (2011) The lipid language of plant–fungal interactions. *Fungal Genetics and Biology* **48**, 4-14.
- Christiansen C, Abou Hachem M, Janeček Š, Viksø-Nielsen A, Blennow A, Svensson B (2009) The carbohydrate-binding module family 20—diversity, structure, and function. *The FEBS journal* **276**, 5006-5029.
- Cobos R, Barreiro C, Mateos RM, Coque J-JR (2010) Cytoplasmic-and extracellular-proteome analysis of *Diplodia seriata*: a phytopathogenic fungus involved in grapevine decline. *Proteome science* **8**, 46.
- Couger MB, Youssef NH, Struchtemeyer CG, Ligginstoffer AS, Elshahed MS (2015) Transcriptomic analysis of lignocellulosic biomass degradation by the anaerobic fungal isolate *Orpinomyces* sp. strain C1A. *Biotechnology for biofuels* **8**, 208.
- Couturier M, Tangthirasun N, Ning X, Brun S, Gautier V, Bennati-Granier C, Silar P, Berrin J-G (2016) Plant biomass degrading ability of the coprophilic ascomycete fungus *Podospora anserina*. *Biotechnology Advances* **34**, 976-983.
- De Jonge R, Thomma BPHJ (2009) Fungal LysM effectors: extinguishers of host immunity? *Trends in Microbiology* **17**, 151-157.

- Dong S, Raffaele S, Kamoun S (2015) The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development* **35**, 57-65.
- Dunaevsky YE, Popova VV, Semenova TA, Beliakova GA, Belozersky MA (2014) Fungal inhibitors of proteolytic enzymes: classification, properties, possible biological roles, and perspectives for practical use. *Biochimie* **101**, 10-20.
- Eddy SR (2011) Accelerated profile HMM searches. *PLoS computational biology* **7**, e1002195.
- Feng J, Wang F, Liu G, Greenshields D, Shen W, Kaminskyj S, Hughes GR, Peng Y, Selvaraj G, Zou J (2009) Analysis of a *Blumeria graminis*-secreted lipase reveals the importance of host epicuticular wax components for fungal adhesion and development. *Molecular Plant-Microbe Interactions* **22**, 1601-1610.
- Fernandes I, Alves A, Correia A, Devreese B, Esteves AC (2014) Secretome analysis identifies potential virulence factors of *Diplodia corticola*, a fungal pathogen involved in cork oak (*Quercus suber*) decline. *Fungal biology* **118**, 516-523.
- Fischer M, Pleiss J (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic acids research* **31**, 319-321.
- Frantzeskakis L, Kusch S, Panstruga R (2019) The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Molecular Plant Pathology* **20**, 3-7.
- Gácsér A, Stehr F, Kröger C, Kredics L, Schäfer W, Nosanchuk JD (2007) Lipase 8 affects the pathogenesis of *Candida albicans*. *Infection and Immunity* **75**, 4710-4718.
- Gazis R, Kuo A, Riley R, LaButti K, Lipzen A, Lin J, Amirebrahimi M, Hesse CN, Spatafora JW, Henrissat B (2016) The genome of *Xylona heveae* provides a window into fungal endophytism. *Fungal biology* **120**, 26-42.
- Gladman S, Seemann T (2012) VelvetOptimiser. *Victorian Bioinformatics Consortium, Clayton, Australia*: <http://bioinformatics.net.au/software/velvetoptimiser.shtml>.
- Goulet KM, Saville BJ (2017) Carbon acquisition and metabolism changes during fungal biotrophic plant pathogenesis: insights from *Ustilago maydis*. *Canadian Journal of Plant Pathology* **39**, 247-266.

- Grandaubert J, Lowe RG, Soyer JL, Schoch CL, Van de Wouw AP, Fudal I, Robbertse B, Lapalu N, Links MG, Ollivier B (2014) Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans*-*Leptosphaeria biglobosa* species complex of fungal pathogens. *BMC genomics* **15**, 891.
- Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, Weimann A, Damm U, Dallery J-F, Hainaut M (2016) Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nature communications* **7**, 1-13.
- Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular biology and evolution* **30**, 1987-1997.
- Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, Andreopoulos B, Baker SE, Barry K, Bills G (2020) 101 Dothideomycetes genomes: a test case for predicting lifestyles and emergence of pathogens. *Studies in mycology* **96**, 141-153.
- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M (2015) BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767-769.
- Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491.
- Howlett BJ (2006) Secondary metabolite toxins and nutrition of plant pathogenic fungi. *Current Opinion in Plant Biology* **9**, 371-375.
- Hyde KD, Maharachchikumbura SSN, Hongsanan S, Samarakoon MC, Lücking R, Pem D, Harishchandra D, Jeewon R, Zhao R-L, Xu J-C (2017) The ranking of fungi: a tribute to David L. Hawksworth on his 70th birthday. *Fungal diversity* **84**, 1-23.
- Islam MS, Haque MS, Islam MM, Emdad EM, Halim A, Hossen QMM, Hossain MZ, Ahmed B, Rahim S, Rahman MS (2012) Tools to kill: genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina*. *BMC genomics* **13**, 493.

- Ito K, Tanaka T, Hatta R, Yamamoto M, Akimitsu K, Tsuge T (2004) Dissection of the host range of the fungal plant pathogen *Alternaria alternata* by modification of secondary metabolism. *Molecular Microbiology* **52**, 399-411.
- Jami F, Slippers B, Wingfield MJ, Gryzenhout M (2014) Botryosphaeriaceae species overlap on four unrelated, native South African hosts. *Fungal biology* **118**, 168-179.
- Jami F, Wingfield MJ, Gryzenhout M, Slippers B (2017) Diversity of tree-infecting Botryosphaeriales on native and non-native trees in South Africa and Namibia. *Australasian Plant Pathology* **46**, 529-545.
- Jashni MK, Mehrabi R, Collemare J, Mesarich CH, De Wit PJGM (2015) The battle in the apoplast: further insights into the roles of proteases and their inhibitors in plant–pathogen interactions. *Frontiers in Plant Science* **6**, 584.
- Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *The EMBO journal* **19**, 4004-4014.
- Käll L, Krogh A, Sonnhammer ELL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic acids research* **35**, W429-W432.
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772-780.
- Keller NP, Turner G, Bennett JW (2005) Fungal secondary metabolism—from biochemistry to genomics. *Nature Reviews Microbiology* **3**, 937-947.
- Kolattukudy PE (1985) Enzymatic penetration of the plant cuticle by fungal pathogens. *Annual Review of Phytopathology* **23**, 223-250.
- Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes1. *Journal of Molecular Biology* **305**, 567-580.
- Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *Journal of statistical software* **25**, 1-18.

- Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for biofuels* **6**, 41.
- Liu Z, Lian S, Li B, Lu H, Dong X, Wang C (2016) Draft genome sequence of *Botryosphaeria dothidea*, the pathogen of apple ring rot. *Genome announcements* **4**, e01142-01116.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2013) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* **42**, D490-D495.
- Lomsadze A, Burns PD, Borodovsky M (2014) Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic acids research* **42**, e119-e119.
- Lopez D, Ribeiro S, Label P, Fumanal B, Venisse J-S, Kohler A, De Oliveira RR, Labutti K, Lipzen A, Lail K (2018) Genome-Wide Analysis of *Corynespora cassicola* Leaf Fall Disease Putative Effectors. *Frontiers in microbiology* **9**, 276.
- Lyu X, Shen C, Fu Y, Xie J, Jiang D, Li G, Cheng J (2015) Comparative genomic and transcriptional analyses of the carbohydrate-active enzymes and secretomes of phytopathogenic fungi reveal their significant roles during infection and development. *Scientific Reports* **5**, 15565.
- Ma L-J, Van Der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367-373.
- Manning VA, Pandelova I, Dhillon B, Wilhelm LJ, Goodwin SB, Berlin AM, Figueroa M, Freitag M, Hane JK, Henrissat B (2013) Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. *G3: Genes, Genomes, Genetics* **3**, 41-63.
- Marincowitz S, Groenewald JZ, Wingfield MJ, Crous PW (2008) Species of Botryosphaeriaceae occurring on Proteaceae. *Persoonia-Molecular Phylogeny and Evolution of Fungi* **21**, 111-118.
- Marsberg A, Kemler M, Jami F, Nagel JH, Postma-Smidt A, Naidoo S, Wingfield MJ, Crous PW, Spatafora JW, Hesse CN, Robbertse B, Slippers B (2016) *Botryosphaeria dothidea*: a latent

- pathogen of global importance to woody plant health. *Molecular Plant Pathology* **18**, 477–488.
- Marshall R, Kombrink A, Motteram J, Loza-Reyes E, Lucas J, Hammond-Kosack K, Thomma B, Rudd J (2011) Analysis of two in planta expressed LysM effector homologues from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant physiology*, pp-111.
- Massonnet M, Morales-Cruz A, Figueroa-Balderas R, Lawrence DP, Baumgartner K, Cantu D (2016) Condition-dependent co-regulation of genomic clusters of virulence factors in the grapevine trunk pathogen *Neofusicoccum parvum*. *Molecular Plant Pathology* DOI: 10.1111/mpp.12491.
- Mehl JWM, Slippers B, Roux J, Wingfield MJ (2013) Cankers and Other Diseases Caused by the Botryosphaeriaceae. In: *Infectious Forest Diseases* (eds. Gonthier P, Nico G), pp. 298-317. CABI.
- Miyanaga A, Koseki T, Miwa Y, Mese Y, Nakamura S, Kuno A, Hirabayashi J, Matsuzawa H, Wakagi T, Shoun H (2006) The family 42 carbohydrate-binding module of family 54 α -L-arabinofuranosidase specifically binds the arabinofuranose side chain of hemicellulose. *Biochemical Journal* **399**, 503-511.
- Mohali S, Slippers B, Wingfield MJ (2007) Identification of Botryosphaeriaceae from *Eucalyptus*, *Acacia* and *Pinus* in Venezuela. *Fungal diversity* **25**, 103-125.
- Mohanta TK, Bae H (2015) The diversity of fungal genome. *Biological Procedures Online* **17**, 8.
- Möller M, Stukenbrock EH (2017) Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology* **15**, 756.
- Morales-Cruz A, Amrine KCH, Blanco-Ulate B, Lawrence DP, Travadon R, Rolshausen PE, Baumgartner K, Cantu D (2015) Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. *BMC genomics* **16**, 469.

- Mueller AN, Ziemann S, Treitschke S, Aßmann D, Doehlemann G (2013) Compatibility in the *Ustilago maydis*–maize interaction requires inhibition of host cysteine proteases by the fungal effector Pit2. *PLoS pathogens* **9**, e1003177.
- Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F (2012) Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathogens* **8**, e1003037.
- Olivieri F, Zanetti ME, Oliva CR, Covarrubias AA, Casalongué CA (2002) Characterization of an extracellular serine protease of *Fusarium eumartii* and its action on pathogenesis related proteins. *European Journal of Plant Pathology* **108**, 63-72.
- Pavlic D, Wingfield MJ, Barber P, Slippers B, Hardy GESJ, Burgess TI (2008) Seven new species of the Botryosphaeriaceae from baobab and other native trees in Western Australia. *Mycologia* **100**, 851-866.
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786.
- Phillips AJL, Hyde KD, Alves A, Liu J-KJ (2019) Families in Botryosphaeriales: a phylogenetic, morphological and evolutionary perspective. *Fungal diversity* **94**, 1-22.
- Plummer KM, Clark SJ, Ellis LM, Loganathan A, Al-Samarrai TH, Rikkerink EHA, Sullivan PA, Templeton MD, Farley PC (2004) Analysis of a secreted aspartic peptidase disruption mutant of *Glomerella cingulata*. *European Journal of Plant Pathology* **110**, 265-274.
- Queiroz CBd, Santana MF (2020) Prediction of the secretomes of endophytic and nonendophytic fungi reveals similarities in host plant infection and colonization strategies. *Mycologia*, 1-13.
- Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RHY, Zody MC, Kunjeti SG, Donofrio NM (2010) Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* **330**, 1540.
- Raffaele S, Kamoun S (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* **10**, 417-430.

- Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD (2017) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic acids research* **46**, D624-D632.
- Robert-Siegwald G, Vallet J, Abou-Mansour E, Xu J, Rey P, Bertsch C, Rego C, Larignon P, Fontaine F, Lebrun M-H (2017) Draft Genome Sequence of *Diplodia seriata* F98.1, a Fungal Species Involved in Grapevine Trunk Diseases. *Genome announcements* **5**, e00061-00017.
- Rodas CA, Slippers B, Gryzenhout M, Wingfield MJ (2009) Botryosphaeriaceae associated with *Eucalyptus* canker diseases in Colombia. *Forest Pathology* **39**, 110-123.
- Rooney HCE, van't Klooster JW, van der Hoorn RAL, Joosten MHAI, Jones JDG, de Wit PJGM (2005) Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. *Science* **308**, 1783-1786.
- Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature communications* **2**, 202.
- Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301-302.
- Saunders DGO, Win J, Kamoun S, Raffaele S (2014) Two-dimensional data binning for the analysis of genome architecture in filamentous plant pathogens and other eukaryotes. In: *Plant-Pathogen Interactions*, pp. 29-51. Springer.
- Schlegel M, Münsterkötter M, Güldener U, Bruggmann R, Duò A, Hainaut M, Henrissat B, Sieber CMK, Hoffmeister D, Grünig CR (2016) Globally distributed root endophyte *Phialocephala subalpina* links pathogenic and saprophytic lifestyles. *BMC genomics* **17**, 1015.
- Schoch CL, Crous PW, Groenewald JZ, Boehm EWA, Burgess TI, De Gruyter J, De Hoog GS, Dixon LJ, Grube M, Gueidan C (2009) A class-wide phylogenetic assessment of Dothideomycetes. *Studies in mycology* **64**, 1-15.
- Schoch CL, Shoemaker RA, Seifert KA, Hambleton S, Spatafora JW, Crous PW (2006) A multigene phylogeny of the Dothideomycetes using four nuclear loci. *Mycologia* **98**, 1041-1052.

- Schulz B, Boyle C (2005) The endophytic continuum. *Mycological research* **109**, 661-686.
- Slippers B, Boissin E, Phillips AJL, Groenewald JZ, Lombard L, Wingfield MJ, Postma A, Burgess T, Crous PW (2013) Phylogenetic lineages in the Botryosphaeriales: a systematic and evolutionary framework. *Studies in mycology* **76**, 31-49.
- Slippers B, Smit WA, Crous PW, Coutinho TA, Wingfield BD, Wingfield MJ (2007) Taxonomy, phylogeny and identification of Botryosphaeriaceae associated with pome and stone fruit trees in South Africa and other regions of the world. *Plant Pathology* **56**, 128-139.
- Slippers B, Wingfield MJ (2007) Botryosphaeriaceae as endophytes and latent pathogens of woody plants: diversity, ecology and impact. *Fungal Biology Reviews* **21**, 90-106.
- Smit A, Hubley R (2017) RepeatModeler-1.0.10.
- Smit AFA, Hubley R, Green P (2017) RepeatMasker Open-4.0.5.
- Southworth D (2012) *Biocomplexity of plant-fungal interactions* John Wiley & Sons.
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313.
- Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62.
- Stergiopoulos I, Collemare J, Mehrabi R, De Wit PJGM (2013) Phytotoxic secondary metabolites and peptides produced by plant pathogenic Dothideomycete fungi. *FEMS Microbiology Reviews* **37**, 67-93.
- Stukenbrock EH, Jorgensen FG, Zala M, Hansen TT, McDonald BA, Schierup MH (2010) Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola*. *PLoS Genet* **6**, e1001189.
- Takeda K, Matsumura H, Ishida T, Samejima M, Ohno H, Yoshida M, Igarashi K, Nakamura N (2015) Characterization of a novel PQQ-dependent quinoxaline pyranose dehydrogenase from *Coprinopsis cinerea* classified into auxiliary activities family 12 in carbohydrate-active enzymes. *PLoS ONE* **10**, e0115722.
- Team RC (2013) R: A language and environment for statistical computing.

- Testa AC, Oliver RP, Hane JK (2016) OcculterCut: a comprehensive survey of AT-rich regions in fungal genomes. *Genome biology and evolution* **8**, 2044-2064.
- Thon MR, Nuckles EM, Takach JE, Vaillancourt LJ (2002) CPR1: a gene encoding a putative signal peptidase that functions in pathogenicity of *Colletotrichum graminicola* to maize. *Molecular Plant-Microbe Interactions* **15**, 120-128.
- Thynne E, Mead OL, Chooi Y-H, McDonald MC, Solomon PS (2019) Acquisition and loss of secondary metabolites shaped the evolutionary path of three emerging phytopathogens of wheat. *Genome Biology and Evolution*, evz037.
- Urbez-Torres JR (2011) The status of Botryosphaeriaceae species infecting grapevines. *Phytopathologia Mediterranea* **50**, 5-45.
- Vaissie P, Monge A, Husson F (2015) Factoshiny: perform factorial analysis from FactoMineR with a shiny application. *R package version* **1**.
- van der Nest MA, Bihon W, De Vos L, Naidoo K, Roodt D, Rubagotti E, Slippers B, Steenkamp ET, Wilken PM, Wilson A (2014) Draft genome sequences of *Diplodia sapinea*, *Ceratocystis manginecans*, and *Ceratocystis moniliformis*. *IMA fungus* **5**, 135-140.
- van Esse HP, van't Klooster JW, Bolton MD, Yadeta KA, Van Baarlen P, Boeren S, Vervoort J, de Wit PJGM, Thomma BPHJ (2008) The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense. *The Plant Cell* **20**, 1948-1963.
- Van Wyk S, Harrison CH, Wingfield BD, De Vos L, van Der Merwe NA, Steenkamp ET (2019) The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. *PeerJ* **7**, e7447.
- Verma S, Gazara RK, Nizam S, Parween S, Chattopadhyay D, Verma PK (2016) Draft genome sequencing and secretome analysis of fungal phytopathogen *Ascochyta rabiei* provides insight into the necrotrophic effector repertoire. *Scientific Reports* **6**, 24638.
- Voigt CA, Schäfer W, Salomon S (2005) A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals. *The Plant Journal* **42**, 364-375.

- Wang X, Zhang X, Liu L, Xiang M, Wang W, Sun X, Che Y, Guo L, Liu G, Guo L (2015) Genomic and transcriptomic analysis of the endophytic fungus *Pestalotiopsis fici* reveals its lifestyle and high potential for synthesis of natural products. *BMC genomics* **16**, 28.
- Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M (2016) gplots: various R programming tools for plotting data. R package version 3.0. 1. *The Comprehensive R Archive Network*.
- Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM (2017) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution*.
- Wingfield BD, Bills GF, Dong Y, Huang W, Nel WJ, Swalarsk-Parry BS, Vaghefi N, Wilken PM, An Z, de Beer ZW (2018) Draft genome sequence of *Annulohypoxyton stygium*, *Aspergillus mulundensis*, *Berkeleyomyces basicola* (syn. *Thielaviopsis basicola*), *Ceratocystis smalleyi*, two *Cercospora beticola* strains, *Coleophoma cylindrospora*, *Fusarium fructicaudum*, *Phialophora cf. hyalina*, and *Morchella septimelata*. *IMA fungus* **9**, 199.
- Wolpert TJ, Dunkle LD, Ciuffetti LM (2002) Host-selective toxins and avirulence determinants: what's in a name? *Annual Review of Phytopathology* **40**, 251-285.
- Xu X-H, Su Z-Z, Wang C, Kubicek CP, Feng X-X, Mao L-J, Wang J-Y, Chen C, Lin F-C, Zhang C-L (2014) The rice endophyte *Harpophora oryzae* genome reveals evolution from a pathogen to a mutualistic endophyte. *Scientific Reports* **4**, 5783.
- Yang T, Groenewald JZ, Cheewangkoon R, Jami F, Abdollahzadeh J, Lombard L, Crous PW (2017) Families, genera, and species of Botryosphaeriales. *Fungal biology* **121**, 322-346.
- Yang Y, Liu X, Cai J, Chen Y, Li B, Guo Z, Huang G (2019) Genomic characteristics and comparative genomics analysis of the endophytic fungus *Sarocladium brachiariae*. *BMC genomics* **20**, 1-20.
- Zaccaron AZ, Bluhm BH (2017) The genome sequence of *Bipolaris cookei* reveals mechanisms of pathogenesis underlying target leaf spot of sorghum. *Scientific Reports* **7**, 17217.

Zeiner CA, Purvine SO, Zink EM, Paša-Tolić L, Chaput DL, Haridas S, Wu S, LaButti K, Grigoriev IV, Henrissat B (2016) Comparative analysis of secretome profiles of manganese (II)-oxidizing Ascomycete fungi. *PLoS ONE* **11**, e0157844.

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-829.

Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic acids research*.

Zhang Y, Crous PW, Schoch CL, Bahkali AH, Guo LD, Hyde KD (2011) A molecular, morphological and ecological re-appraisal of Venturiales—a new order of Dothideomycetes. *Fungal diversity* **51**, 249-277.

Zhao Z, Liu H, Wang C, Xu J-R (2013) Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC genomics* **14**, 274.

Zhao Z, Liu H, Wang C, Xu J-R (2014) Correction: Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC genomics* **15**, 6.

Tables and Figures

Table 1. List of genome sequences used in this study

Table 2: Genome statistics of new draft Botryosphaeriaceae genomes

Table 3. Genome and genome annotation completeness assesment

Table 4. Summary of genomic architecture features of Botryosphaeriaceae genomes

Figure 1. Phylogenomic tree and functional annotation summary of 26 *Botryosphaeriaceae*, 39 Dothideomycetes and one outgroup (*Aspergillus nidulans*). The supermatrix maximum likelihood phylogeny was determined using the sequence data of 207 single-copy core orthologous genes. Branches with 100% bootstrap support are indicated in black, those with less than 100% support are indicated in grey. The number of genes (total and secreted) annotated with CAZyme, proteases or lipase activity, as well as the number of gene cluster types involved in secondary metabolite biosynthesis are indicated using bar graphs.

Figure 2. Box-and-whisker plots of the number of prominent CAZyme, protease and lipase families (total and secreted) and secondary metabolite BGC types present within the genomes of the considered *Botryosphaeriaceae* and other Dothideomycetes taxa. Taxa are placed into three categories: Dothideomycetidae, Pleosporomycetidae (plus related taxa without subclass designation) and *Botryosphaeriaceae*. The upper and lower bounds of the box represent the 1st and 3rd quartiles and the bar inside the median. The error lines (whiskers) represents 1.5 times the interquartile range (IQR) and outliers are indicated as dots. Additionally, the genera of the *Botryosphaeriaceae* are indicated by means of a scatterplot overlain on the *Botryosphaeriaceae* box-and-whisker plot.

Figure 3. Principal component analysis of functional annotation categories of secreted proteins and secondary metabolite BGCs from *Botryosphaeriaceae* and other Dothideomycetes. Taxa are indicated using abbreviated names (Supplementary File 1) and colours indicate their Order/Family. The percentage variation accounted for by each principal component is indicated at each axis.

Figure 4. Gene density landscape of the genera of *Botryosphaeriaceae*. The two-dimensional heatmap shows the distribution of genes to gene dense or sparse regions of the genome based on their 5' and 3' flanking intergenic regions (FIRs). Two-dimensional heatmaps of each genus is the average across bins of all genomes in the group. The heatmaps labelled as “*Botryosphaeria*-clade” include the genomes of *B. dothidea*, *B. kuwatsukai* and *M. phaseolina*.

The values on the axes are distances in base pairs and signify the upper limit of each bin. The colours of the heatmap represents the number of genes present within each two-dimensional bin.

Supplementary File 1: Excel file of the functional annotation data

Supplementary File 2: Phylogenetic tree indicating the gene family contractions/expansions of CAZymes, proteases, lipases and secondary metabolite BGCs among the *Botryosphaeriaceae* and other Dothideomycetes

Supplementary File 3: Hierarchical clustering and heatmap of *Botryosphaeriaceae* and other representative Dothideomycetes species based on the number of functional annotation categories of secreted and secondary metabolite BGCs. Overrepresented (red and dark red) and underrepresented (blue and dark blue) values are scaled relative to the column mean.

Supplementary File 4: Excel file containing data of gene association with gene-sparse/TA-rich regions

Table 1. List of genome sequences used in this study*

Species	Reference collection/Isolate number	Assembly Size (Mbp)	Number of gene models	Genome accession number	Reference
Dothideomycetes					
Botryosphaerales					
<i>Botryosphaeria kuwatsukai</i>	LW030101	47.39	11 278	MDSR01000000	Lui et al. 2016
<i>Botryosphaeria dothidea</i>	CMW8000	43.50	11 368	http://genome.jgi.doe.gov/Botdo1_1/Botdo1_1.home.html	Marsberg et al. 2017
<i>Diplodia corticola</i>	CBS112549	34.99	9 376	MNUE01000001	Fernandes et al. 2014
<i>Diplodia sapinea</i>	CBS117911	36.05	9 589	AXCF00000000	Bihon et al. 2014
	CBS138184	35.24	9 386	JHUM00000000	Bihon et al. 2014
<i>Diplodia scrobiculata</i>	CBS139796	34.93	9 204	LAEG00000000	Wingfield et al. 2015
<i>Diplodia seriata</i>	UCDDS831	37.12	9 759	MSZU00000000	Morales-Cruz et al. 2015
	F98.1	37.27	9 832	LAQI00000000	Robert-Siegwald et al. 2017
<i>Eutiarosporella darliae</i>	2G6	27.27	7 904	GFXH01000000	Thynne et al. 2019
<i>Eutiarosporella pseudodarliae</i>	V4B6	26.74	7 846	GFXI01000000	Thynne et al. 2019
<i>Eutiarosporella tritici-australis</i>	153	26.59	7 783	GFXG01000000	Thynne et al. 2019
<i>Lasiodiplodia gonubiensis</i>	CBS115812	41.14	10 649	RHKG00000000	Present study
<i>Lasiodiplodia pseudotheobromae</i>	CBS116459	43.01	10 964	RHKG00000000	Present study
<i>Lasiodiplodia theobromae</i>	CBS164.96	42.97	10 961	RHKG00000000	Present study
	CSS01	43.28	11 017	RHKB00000000	Yan et al. 2017
<i>Macrophomina phaseolina</i>	MS6	48.88	10799	AHHD00000000	Islam et al. 2012
<i>Neofusicoccum cordaticola</i>	CBS123634	45.71	12822	RHKG00000000	Present study
	CBS123638	43.56	12630	RHKG00000000	Present study
<i>Neofusicoccum kwambonambiense</i>	CBS123639	44.17	12839	RHKE00000000	Present study
	CBS123642	44.21	12904	RKSS00000000	Present study
<i>Neofusicoccum parvum</i>	CMW9080	41.41	12870	RHJX00000000	Present study
	CBS123649	42.16	12453	RHJY00000000	Present study
	UCRNP2	42.52	12691	AORE00000000	Blanco-Ulate et al. 2013
<i>Neofusicoccum ribis</i>	CBS115475	43.18	12708	RHJZ00000000	Present study
	CBS121.26	43.12	12733	RHKA00000000	Present study
<i>Neofusicoccum umdonicola</i>	CBS123644	42.29	12816	RHKB00000000	Present study
Capnodiales					

<i>Acidomyces richmondensis</i>	meta	26.82	10338	JOOL000000000	Mosier et al. 2016
<i>Baudoinia panamericana</i>	UAMH 10762	21.88	10508	AEIF000000000	Ohm et al. 2012
<i>Cercospora berteroae</i>	CBS538.71	33.89	11903	PNEN000000000	De Jonge et al. 2018
<i>Cercospora beticola</i>	09-40	37.06	12463	LKMD000000000	De Jonge et al. 2018
<i>Cercospora zeina</i>	CMW25467	40.76	10193	MVDW000000000	Wingfield et al. 2017
<i>Dothistroma septosporum</i>	NZE10	30.21	12415	AIEN000000000	Ohm et al. 2012
<i>Pseudocercospora eumusae</i>	CBS 114824	47.12	12632	LFZN010000000	Chang et al. 2016
<i>Pseudocercospora fijiensis</i>	CIRAD86	29.98	13066	AIHZ000000000	Ohm et al. 2012
<i>Pseudocercospora musae</i>	CBS 116634	60.44	13129	LFZO000000000	Chang et al. 2016
<i>Ramularia collo-cygni</i>	URUG2	32.25	11612	FJUY000000000	McGrann et al. 2016
<i>Sphaerulina musiva</i>	SO2202	29.35	10233	AEFD000000000	Ohm et al. 2012
<i>Zymoseptoria tritici</i>	IPO323	39.69	10963	ACPE000000000	Goodwin et al. 2011
Dothideales					
<i>Aureobasidium namibiae</i>	CBS 147.97	25.43	10259	AYEM000000000	Gostinčar et al. 2014
<i>Aureobasidium subglaciale</i>	EXF-2481	25.80	10792	AYYB000000000	Gostinčar et al. 2014
Hysteriales					
<i>Hysterium pulicare</i>	CBS 123377	38.43	12352	AJFK000000000	Ohm et al. 2012
<i>Rhytidhysterium rufulum</i>	CBS 306.38	40.18	12117	AJFL000000000	Ohm et al. 2012
Myriangiales					
<i>Elsinoe australis</i>	NL1	23.34	9223	NHZQ000000000	Zhao et al. 2020
Mytilinidiales					
<i>Lepidopterella palustris</i>	CBS 459.81	45.67	13861	LKAR000000000	Peter et al. 2016
Pleosporales					
<i>Alternaria alternata</i>	SRC1lrK2f	32.99	13466	LXPP000000000	Zeiner et al. 2016
<i>Ascochyta rabiei</i>	ArDII	34.66	10596	JYNV000000000	Verma et al. 2016
<i>Bipolaris maydis</i>	ATCC 48331	32.93	12705	AIHU000000000	Ohm et al. 2012
<i>Bipolaris oryzae</i>	ATCC 44560	31.36	12002	AMCO000000000	Condon et al. 2013
<i>Bipolaris sorokiniana</i>	ATCC 44560	34.41	12214	AEIN000000000	Condon et al. 2013
<i>Bipolaris victoriae</i>	FI3	32.83	12882	AMCY000000000	Condon et al. 2013
<i>Bipolaris zeicola</i>	26-R-13	31.27	12853	AMCN000000000	Condon et al. 2013
<i>Clohesyomyces aquaticus</i>	CBS 115471	49.68	15811	MCFA000000000	Mondo et al. 2017
<i>Corynespora cassiicola</i>	CCP	44.85	17158	NSJI000000000	Lopez et al. 2018
<i>Epicoccum nigrum</i>	ICMP 19927	34.74	12025	NCTX000000000	Fokin et al. 2017
<i>Exserohilum turcicum</i>	Et28A	43.01	11698	AIHT000000000	Ohm et al. 2012

<i>Leptosphaeria maculans</i>	JN3	45.12	12469	FP929064:FP929139	Rouxel et al. 2011
<i>Paraphaeosphaeria sporulosa</i>	AP3s5-JAC2a	38.46	14734	LXPO00000000	Zeiner et al. 2016
<i>Parastagonospora nodorum</i>	SN15	37.21	15994	AAGI00000000	Hane et al. 2007
<i>Periconia macrospinos</i>	DSE2036	54.99	18735	PCYO00000000	Knapp et al. 2018
<i>Pyrenophora tritici-repentis</i>	Pt-1C-BFP	38.00	12169	AAXI00000000	Manning et al. 2013
<i>Stemphylium lycopersici</i>	CIDEFI 216	35.17	8997	LGLR00000000	Franco et al. 2015
Venturiales					
<i>Verruconis gallopava</i>	CBS 43764	31.78	11357	JYBX00000000	Teixeira et al. 2017
<i>incertae sedis</i>					
<i>Cenococcum geophilum</i>	1.58	177.56	14709	LKKR00000000	Peter et al. 2016
<i>Coniosporium apollinis</i>	CBS 100218	28.65	9308	AJKL00000000	Teixeira et al. 2017
<i>Glonium stellatum</i>	CBS 207.34	40.52	14277	LKAO00000000	Peter et al. 2016
Eurotiomycetes					
Eurotiales					
<i>Aspergillus nidulans</i>	FGSC A4	30.28	9556	AACD00000000	Galagan et al. 2005

*Entries in boldface represent genomes that were sequenced as part of the present study

Table 2: Genome statistics of new draft *Botryosphaeriaceae* genomes

	# scaffolds	# contigs	Scaffold length (Mb)	Contig length (Mb)	Genome scaffold N50/L50 (#/kb)	Genome contig N50/L50 (#/kb)	Maximum scaffold length (Mb)	Maximum contig length (kb)	% main genome in scaffolds > 50 KB	% GC	% repetitive sequences
<i>Lasiodiplodia gonubiensis</i> CBS115812	376	1578	41.14	40.97	50/234.83	267/45.4	1.06	226.76	0.92	54.94	3.2
<i>Lasiodiplodia pseudotheobromae</i> CBS116459	403	1285	43.01	42.89	48/236.23	218/61.88	1.03	416.59	0.91	54.82	3.2
<i>Lasiodiplodia theobromae</i> CBS164.96	424	1093	42.97	42.88	48/223.84	163/81.56	1.69	591.21	0.91	54.86	3.4
<i>Neofusicoccum cordaticola</i> CBS123634	1912	8698	45.71	45.10	276/47.38	1218/10.76	680.02	143.71	0.48	55.77	7.8
<i>Neofusicoccum cordaticola</i> CBS123638	2393	3252	43.56	43.44	254/51.16	386/33.04	274.15	152.21	0.51	56.15	8.3
<i>Neofusicoccum kwambonambiense</i> CBS123639	1560	2839	44.17	43.99	219/59.44	363/34.21	344.53	186.24	0.58	56.25	8.1
<i>Neofusicoccum kwambonambiense</i> CBS123642	1717	2871	44.21	44.06	210/61.88	352/34.6	387.07	262.94	0.58	56.33	7.6
<i>Neofusicoccum parvum</i> CBS123649	2185	6686	42.16	41.76	312/39.13	995/12.5	439.22	92.09	0.40	56.65	4.7
<i>Neofusicoccum parvum</i> CMW9080	5188	5739	41.41	41.39	830/14.81	897/13.55	99.22	80.57	0.04	56.65	4.8
<i>Neofusicoccum ribis</i> CBS115475	1994	3261	43.18	43.03	301/42.08	474/26.80	231.82	186.29	0.43	55.96	7.5
<i>Neofusicoccum ribis</i> CBS121.26	2417	3145	43.12	43.03	245/51.65	371/33.92	280.49	167.59	0.51	56.03	7.3
<i>Neofusicoccum umdonicola</i> CBS123644	1343	2424	42.29	42.15	165/73.88	345/36.48	422.57	210.10	0.67	56.75	5.2

Table 3. Genome and genome annotation completeness assesment*

Genome	Genome BUSCO %					Current annotation BUSCO %					Prior annotation BUSCO %				
	C	S	D	F	M	C	S	D	F	M	C	S	D	F	M
<i>Botryosphaeria kuwatsukai</i> LW030101	99.2	98.9	0.3	0.1	0.7	98.4	98.1	0.3	0.6	1.0					
<i>Botryosphaeria dothidea</i> CMW8000	99.0	98.6	0.4	0.3	0.7	98.5	98.1	0.4	0.5	1.0	98.1	97.7	0.4	0.3	1.6
<i>Diplodia corticola</i> CBS112549	98.9	98.7	0.2	0.1	1.0	98.7	98.6	0.1	0.5	0.8	99.4	99.2	0.2	0.3	0.3
<i>Diplodia sapinea</i> CBS117911	98.0	97.8	0.2	0.5	1.5	97.6	97.5	0.1	0.9	1.5					
<i>Diplodia sapinea</i> CBS138184	96.0	95.8	0.2	0.8	3.2	95.6	95.4	0.2	1.2	3.2					
<i>Diplodia scrobiculata</i> CBS139796	95.4	95.2	0.2	2.3	2.3	95.0	94.8	0.2	1.8	3.2					
<i>Diplodia seriata</i> UCDD831	99.1	98.9	0.2	0.2	0.7	98.8	98.7	0.1	0.5	0.7	87	86.9	0.1	5.9	7.1
<i>Diplodia seriata</i> F98.1	99.3	98.9	0.4	0.1	0.6	98.9	98.5	0.4	0.5	0.6	90.1	89.9	0.2	2.6	7.3
<i>Eutiarospora darliae</i> 2G6	98.1	97.8	0.3	0.6	1.3	98.5	97.8	0.7	0.5	1.0					
<i>Eutiarospora pseudodarliae</i> V4B6	98.2	98.0	0.2	0.6	1.2	98.1	97.8	0.3	0.7	1.2					
<i>Eutiarospora tritici-australis</i> 153	98.3	98.0	0.3	0.4	1.3	98.6	98.4	0.2	0.4	1.0					
<i>Lasiodiplodia gonubiensis</i> CBS115812	97.2	97.1	0.1	0.2	2.6	98.3	98.2	0.1	0.8	0.9					
<i>Lasiodiplodia pseudotheobromae</i> CBS116459	99.1	98.9	0.2	0.2	0.7	98.3	98.0	0.3	0.8	0.9					
<i>Lasiodiplodia theobromae</i> CBS164.96	97.2	97.0	0.2	0.2	2.6	98.6	98.2	0.4	0.5	0.9					
<i>Lasiodiplodia theobromae</i> CSS01	99.1	98.7	0.4	0.0	0.9	99.1	98.7	0.4	0.4	0.5					
<i>Macrophomina phaseolina</i> MS6	98.9	98.7	0.2	0.0	1.1	98.1	97.9	0.2	0.8	1.1	91.2	91	0.2	4.5	4.3
<i>Neofusicoccum cordaticola</i> CBS123634	97.0	94.9	2.1	0.9	2.1	96.1	94.2	1.9	1.6	2.3					
<i>Neofusicoccum cordaticola</i> CBS123638	98.5	98.0	0.5	0.5	1.0	98.3	97.9	0.4	0.6	1.1					
<i>Neofusicoccum kwambonambiense</i> CBS123639	99.0	98.6	0.4	0.3	0.7	98.6	98.2	0.4	0.7	0.7					
<i>Neofusicoccum kwambonambiense</i> CBS123642	98.9	98.4	0.5	0.4	0.7	98.8	98.3	0.5	0.5	0.7					
<i>Neofusicoccum parvum</i> CMW9080	95.4	95.0	0.4	2.0	2.6	94.3	93.9	0.4	2.0	3.7					
<i>Neofusicoccum parvum</i> CBS123649	95.1	94.5	0.6	0.9	4.0	96.4	95.9	0.5	1.3	2.3					
<i>Neofusicoccum parvum</i> UCRNP2	98.1	97.9	0.2	0.5	1.4	97.8	97.6	0.2	0.9	1.3	84.9	84.8	0.1	6.6	8.5
<i>Neofusicoccum ribis</i> CBS115475	98.3	97.9	0.4	0.9	0.8	97.6	97.2	0.4	1.3	1.1					
<i>Neofusicoccum ribis</i> CBS121.26	98.7	98.3	0.4	0.6	0.7	98.2	97.7	0.5	1.2	0.6					
<i>Neofusicoccum umdonicola</i> CBS123644	98.7	98.2	0.5	0.5	0.8	98.1	97.5	0.6	1.1	0.8					

*BUSCO percentages for Complete (C), single (S), duplicated (D), fragmented (F) and missing (M) genes

Table 4. Summary of genomic architecture features of *Botryosphaeriaceae* genomes

	Genome size (Mb)	Repetitive sequences (bp)	% of repetitive sequences	Genome GC %	% genome that is GC rich (>50%)	Genes in TA rich regions	Secreted genes in TA rich regions	% genome affected by RIP	Number of LRARs	Total size (bp) of all LRARs
<i>Eutiarospora darliae</i> 2G6	27.27	1 415 283	5.19	61	98.3	3	0	0.91	4	32 276
<i>E. pseudodarliae</i> V4B6	26.74	1 290 421	4.83	61.53	98.6	7	0	0.4	0	0
<i>E. tritici-australis</i> 153	26.59	1 544 103	5.81	61.66	97.3	0	0	2.54	14	95 508
<i>Diplodia sapinea</i> CBS117911	36.05	1 344 948	3.73	56.84	95	73	4	1.12	2	9 500
<i>D. sapinea</i> CBS138184	35.24	1 305 464	3.7	56.73	94.7	68	6	0.96	2	12 500
<i>D. scrobiculata</i> CBS139796	34.93	1 110 716	3.18	57.01	95.6	63	6	0.82	1	13 000
<i>D. seriata</i> UCDDS831	37.27	1 711 911	4.61	56.6	96.8	29	2	2.63	21	125 537
<i>D. seriata</i> F98.1	37.12	1 593 160	4.27	48.47	94.5	51	1	1.87	34	554 000
<i>D. corticola</i> CBS112549	34.99	2 102 085	6.01	47.92	94.5	31	3	3.39	42	690 014
<i>Lasiodiplodia theobromae</i> CBS164.96	42.97	1 498 597	3.49	54.74	89.9	354	32	0.72	3	59 500
<i>L. theobromae</i> CSS01	43.28	1 543 231	3.57	48.15	89.8	306	33	1.04	18	375 000
<i>L. pseudotheobromae</i> CBS116459	43.01	1 402 977	3.26	54.66	89.2	377	36	0.92	12	375 000
<i>L. gonubiensis</i> CBS115812	41.14	1 330 240	3.23	54.72	92.1	223	24	1.18	12	345 500
<i>Botryosphaeria dothidea</i> CMW8000	43.5	1 515 197	3.48	54.3	89.8	385	53	0.88	1	5 000
<i>B. kuwatsukai</i> LW030101	47.39	5 628 118	11.88	53.09	81.1	406	41	9.65	170	1 387 657
<i>Macrophomina phaseolina</i> MS6	48.88	8 000 688	16.37	52.33	80.2	228	23	12.92	172	2 658 966
<i>Neofusicoccum cordaticola</i> CBS123634	45.71	3 594 350	7.86	54.9	86.5	566	54	3.7	105	1 071 347
<i>N. cordaticola</i> CBS123638	43.56	3 614 552	8.3	55.92	88.7	393	40	3.99	47	324 966
<i>N. kwambonambiense</i> CBS123639	44.17	3 369 216	7.62	55.92	89.1	408	40	2.75	64	693 395
<i>N. kwambonambiense</i> CBS123642	44.21	3 592 960	8.13	56.04	89.2	459	49	2.35	56	554 628
<i>N. parvum</i> CMW9080	41.41	1 944 617	4.7	56.54	91.3	406	40	1.36	11	63 557
<i>N. parvum</i> CBS123649	42.16	2 032 314	4.82	56.06	91.8	360	36	1.65	14	88 222
<i>N. parvum</i> UCRNP2	42.52	2 167 908	5.1	56.76	91.4	407	35	1.59	11	72 874
<i>N. ribis</i> CBS115475	43.18	3 241 261	7.51	55.71	89.6	379	38	4.52	79	911 957
<i>N. ribis</i> CBS121.26	43.12	3 171 232	7.35	55.88	89.1	423	51	4.71	25	153 638
<i>N. umdonicola</i> CBS123644	42.29	2 237 372	5.29	56.51	91.6	415	48	1.38	11	76 232