Mycosphere X(X): X–X (XXXX) www.mycosphere.org ISSN 2077 7019

Article

Doi 10.5943/mycosphere/si/1f/5

Comparative genomics provides new insights into the evolution of *Colletotrichum*

Chen YP[1], Wu T[1], Tian WH[1], Ilyukhin E[2], Hyde KD[3,4] and Maharachchikumbura SSN[1,*]

[1]*School of Life Science and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China*
[2]*Swift Current Research and Development Centre, Agriculture and Agri-Food Canada, Swift Current, Saskatchewan S9H 3X2, Canada*
[3]*Center of Excellence in Fungal Research, Mae Fah Luang University, Chaing Rai 57100, Thailand*
[4]*Innovative Institute for Plant Health, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China*

Chen YP, Wu T, Tian WH, Ilyukhin E, Hyde KD, Maharachchikumbura SSN XXXX – Comparative genomics provides new insights into the evolution of *Colletotrichum*. Mycosphere X(X), X–X, Doi Doi 10.5943/mycosphere/si/1f/5

Abstract

Colletotrichum is an economically important pathogen with a worldwide distribution associated with a wide range of plant hosts. Characterization of the genomic profiles of *Colletotrichum* can generate new insights to understand its speciation, diversity, and pathogenicity potential. Although some genomic studies on this genus have been reported, a systematic comparative genomic analysis of the genus has so far been lacking. In this study, we first compared genome completeness generated by second- and third-generation sequencing platforms and confirmed the effectiveness of second-generation sequencing techniques for genomic studies. We then integrated the taxonomic concept of *Colletotrichum* species complexes into the comparative genomic analysis and depicted the genomic features of 102 *Colletotrichum* genomes from 39 species across ten species complexes and two singletons. Genome sizes of *Colletotrichum* species vary tremendously (44.15–109.66 Mb), and the average genome sizes of species complexes are significantly different. Repetitive sequences are the key drivers for genome size and GC content variation. The number of predicted genes of each genome (10,354–17,809) is positively correlated to the size of non-repeat genome sequences. Phylogenetically close strains consist of similar composition of secretomes, however, the number of major components is significantly different between different species complexes. The *C. gloeosporioides* species complex possessed the most abundant CAZymes, which may be responsible for its diverse distribution and higher pathogenicity. A total of 7,971 putative SMGCs covering 20 types were detected in these genomes, suggesting their immense biosynthetic potential to produce natural products. We further constructed a genome-scale TimeTree of *Colletotrichum* and investigated the evolution of gene families. The result suggested that gene gain and loss are both important for environmental adaption but less useful in delineating species complexes. Our study characterized the genomes of *Colletotrichum* species from different perspectives, providing insights into the genomic evolution of *Colletotrichum* species.

Key words – CAZymes – Divergence time – Gene family – Pathogenic fungi – Secondary metabolic gene cluster – Secretome – Sordariomycetes – Species complex

---

Submitted 14 July 2022, Accepted 27 September 2022, Published online 14 November 2022, Published (to be added)
Corresponding Author: Maharachchikumbura S.S.N. – e-mail – sajeewa83@yahoo.com

Introduction

Colletotrichum is the sole genus of the family Glomerellaceae (Glomerellales, Sordariomycetes), which was introduced by Corda (1831) with *C. lineola* as the type species. The genus is listed in the top 10 genera of fungi of plant pathology (Dean et al. 2012), species of which are capable of infecting many fruit crops including apples, strawberries, citrus, and stone fruits (Chang et al. 2018, Oo et al. 2018, Zhang et al. 2020, Wang et al. 2021), as well as major cereals such as maize and sorghum (Lima et al. 2013, Tao et al. 2013, Schena et al. 2014). Several *Colletotrichum* species, viz. *C. coccodes, C. crassipes, C. dematium, C. gloeosporioides, and C. graminicola*, cause human infections (Cano et al. 2004, Shivaprakash et al. 2011). The lifestyles of *Colletotrichum* are generally categorized as necrotrophic, hemibiotrophic, and endophytic, the most common of which is hemibiotrophic (De Silva et al. 2017, Jayawardena et al. 2021).

Colletotrichum comprises over 280 species, and 1,000 epithets in Index Fungorum (http://indexfungorum.org/; Accessed 1st June 2022). Accurate species identification is vital for understanding its taxonomic positions, controlling plant diseases, and developing quarantine measures. However, limited diagnostic characters mainly composed of spore shape and size, coupled with morphological plasticity, present challenges in solely using morphology to identify *Colletotrichum* species (Cannon et al. 2012, Hyde et al. 2014, Maharachchikumbura et al. 2021). The internal transcribed spacer (ITS) known as the universal DNA barcode for fungi, is capable of identifying species to the genus level but features poor resolution for distinguishing *Colletotrichum* spp. at the species level (Crous et al. 2015, Jayawardena et al. 2020). The ITS and multi-locus phylogenies indicated that the genus could be subclassified into many clades, and these clades were assigned a formal nomenclature system termed a “complex” (also called “aggregate”) by Cannon et al. (2012). Recently, Liu et al. (2022) used a comprehensive multi-locus dataset to assign *Colletotrichum* species into 16 species complexes including 1 new species complex and 15 singleton species.

Protein secretion is an important feature of fungi that is vital for decaying substrates and interacting with neighboring microbes, plants, or animals and human beings (Martinez et al. 2009, Kloppholz et al. 2011, Shah et al. 2013, Liang et al. 2018a, Prasanth et al. 2019). The fungal secretome (the complete set of secreted proteins) consists of various functional groups, including proteases, lipases, Carbohydrate-Active enZymes (CAZymes), secreted proteins of unknown function, and small-secreted proteins (SSPs) (Alfaro et al. 2014). Proteases from pathogenic species can digest lung parenchyma as well as cuticles and skin, facilitating the infection of humans, animals, and plants (Yike 2011). Extracellular lipases contribute to the virulence of pathogenic fungi, interfering with the production of secreted lipases can impede the growth of pathogenic yeast in the absence of de novo fatty acid synthesis (Nguyen et al. 2011b, Park et al. 2013). CAZymes are enzymes involved in the metabolism and synthesis of carbohydrates, especially in the breakdown of the plant cell wall and storage compounds (Drula et al. 2021). SSPs secreted by fungi contain a signal peptide and fewer than 300 amino acids in size (Kim et al. 2016). Biotrophic and symbiotic fungi usually have a higher proportion of species-specific SSPs than hemibiotrophic and necrotrophic (Feldman et al. 2020). Ashwin et al. (2017) revealed the relationship between light and expression of effectors in *Colletotrichum falcatum* via comparative secretome analysis. To date, we have limited knowledge about the secretome of *Colletotrichum* spp., particularly whether the components of the secretome differ between different species and species complexes.

As next-generation sequencing (NGS) technologies improve and become more cost-effective, the number of available genomes has greatly increased in the public domains (https://www.ncbi.nlm.nih.gov/genbank/statistics/). New sequencing techniques, such as RNA sequencing (RNA-seq), single molecule real time (SMRT) sequencing and single-cell RNA sequencing (scRNA-seq) (Wang et al. 2009, Goodwin et al. 2016, Longo et al. 2021) have yielded new insights into fungal studies. Gan et al. (2013) conducted comparative genomic and transcriptomic analyses of *C. orbiculare* 104-T (MAFF 240422) and *C. gloeosporioides* Nara gc5, revealing the mechanism of hemibiotrophic stage shift in *Colletotrichum* species. Comparing the genome and gene variations among *Colletotrichum* and non-*Colletotrichum* genomes, Liang et al.

(2018b) found fungus-to-fungus horizontal gene transfer and many duplication/loss events at key phylogenetic nodes. Lineage-specific losses of CAZymes and proteases-encoding genes in *Colletotrichum* species displayed a narrow host range as well as duplications of these families in the *C. acutatum* species complex (Baroncelli et al. 2016). Genome-wide expression profiling showed that SSPs and secondary metabolite synthesis genes were upregulated during the initial stages of host colonization, whereas the necrotrophic stage of growth was related to the upregulation of degradative enzyme-encoding genes (O'Connell et al. 2012, Gan et al. 2013). Tsushima et al. (2019) revealed that the transposable elements (TEs) play important roles in the genomic plasticity of closely related strains by investigating large-scale rearrangements, the existence of strain-specific regions, and candidate effector gene sets. Presently, there are over one hundred *Colletotrichum* genomes available in the NCBI Assembly database, but related studies have mainly focused on functional genomics and paid less attention on basic genomic characterization.

The molecular evolutionary clock, proposed by Zuckerkandl (1962), characterizes the evolutionary rate to time with the simplest clock model assuming that the rate of molecular evolution remains constant across species. The idea allows us to estimate evolutionary and demographic timescales, improving our understanding of biological patterns and processes. Molecular clocks have been used to provide with new insights into fungal taxonomic studies. Divergence times are useful for ranking families, orders, and subclasses (Liu et al. 2017, Hongsanan et al. 2017). Bhunjun et al. (2021) tried to recognize *Colletotrichum* species complexes using time dating, and the estimated interval of divergence time for species complexes spanned from 4.8 to 32.2 million years ago (MYA). In the genomic era, molecular-clock methods become more complex to handle larger datasets, most of which comprise large numbers of taxa and large numbers of genes, representing a growing trend toward genome-scale data sets (Shen et al. 2020, Li et al. 2021). To date, a comprehensive, genome-scale TimeTree of *Colletotrichum* remains unavailable for comparison and verification with the TimeTree generated based on the multi-locus dataset.

Fungi produce a wide variety of molecules known as secondary metabolites (SMs), many of which are beneficial not only to the microbes themselves but also to plants, animals, and human beings (Flemings 1929, Schmitt et al. 2004, Tanaka et al. 2005, Nirlane da Costa Souza et al. 2016, Kanpiengjai et al. 2020). Secondary metabolic gene clusters (SMGCs) are groups of two or more genes contiguously arranged in the genome sequences that form a biosynthetic pathway for the production of the corresponding SMs (Medema et al. 2015). Collemare et al. (2008) revealed that ascomycetes have more SMGCs than basidiomycetes, archeo-ascomycetes, and chytridiomycetes. Some SMs are characterized as pathogenicity factors, such as the host-specific cyclic tetrapeptide, HC-toxin (*Bipolaris carbonum* race 1), or AM-toxin (*Alternaria alternata*) (Ahn & Walton 1996, Akamatsu et al. 1999). At least 109 secondary metabolites have been reported from multiple *Colletotrichum* species to date (Kim & Shim 2019). However, inter-species and -complex differences and diversity remain less clear.

This study aims to conduct a comprehensive phylogenomic analysis of available *Colletotrichum* genomes with the following objectives: 1) to characterize and compare the basic genomic characteristics of *Colletotrichum* species; 2) to study the secretome in different species and complexes; 3) to estimate the divergence time based on genome-scale dataset; 4) to assess the capability of *Colletotrichum* spp. to produce natural products.

Materials & Methods

Genome collection and assessment

In order to collect the most comprehensive genome dataset of *Colletotrichum* as of 11 November 2021, we used the search term "*Colletotrichum*" in NCBI Genome Browser to obtain the basic assembly information. All genomes and related annotation files were downloaded using the NCBI tool datasets version 12.25.0. Genome assemblies were assessed using BUSCO version 5.3.3

with the “ascomycota_odb10” as the benchmarking data set (Simão et al. 2015). Multiple factors, including the related published records, the completeness of genomes, assembly level, and availability of the strains, were taken into consideration for selecting strains. Up to 10 isolates of each species were selected for downstream analysis.

Genome annotation

RepeatMasker version 4.1.2-p1 (Smit et al. 2015) with the curated repeat libraries RepBase Edition-20181026 (Bao et al. 2015) and Dfam release 3.5 (Storer et al. 2021), in conjunction with RepeatModeler version 2.0.2 (Flynn et al. 2020) were used to identify repetitive sequences. Transfer RNA (tRNA) genes were predicted using tRNAscan-SE version 2.0.9 (Chan & Lowe 2019). To increase the gene prediction accuracy, a huge protein database was built for homology-based gene prediction by combining reviewed fungal proteins from the SwissProt Database (The UniProt Consortium 2019) and a custom protein dataset. The latter was constructed based on 269 public RNA-seq datasets of *Colletotrichum* spp. (Supplementary Table 1). De novo transcript assembly of these datasets was conducted using Trinity version 2.1.1 (Grabherr et al. 2011), and protein-coding regions were identified using TransDecoder version 5.5.0 (http://transdecoder.github.io/). The protein sequences from different sources, were filtered using CD-HIT version 4.8.1 (Li & Godzik 2006) with the sequence identity threshold set at 0.8, and only one representative protein of each cluster was saved. Gene annotation was conducted using the eukaryotic genome annotation pipeline BRAKER2 version 2.1.6 (Brůna et al. 2021), which combined ab initio gene prediction from AUGUSTUS version 3.4.0 (Keller et al. 2011) and homology-based gene prediction from GenomeThreader version 1.7.1 (Gremme et al. 2005).

Functional annotation

Annotation of proteins were conducted using InterProScan version 5.54-87.0 (Blum et al. 2020) with the Pfam database version 35.0 (Mistry et al. 2020) to classify proteins into proper families. Fungal secretomes including proteases, lipases, CAZymes, and SSPs were predicted using the published pipeline (Pellegrin et al. 2015). SignalP version 4.1 (Emanuelsson et al. 2007) was used to predict the presence and location of signal peptide cleavage sites. Phobius version 1.01 (Käll et al. 2004) was used to determine the transmembrane topology and signal peptides. TMHMM version 2.0 (Krogh et al. 2001) was used to predict transmembrane helices.

Gene family analysis

The proteins with fewer than 50 amino acids were deleted, and gene families were inferred using OrthoFinder version 2.5.4 (Emms & Kelly 2019) under default parameters. CAFE5 (Mendes et al. 2020) was used to detect gene family evolution, including expansion and contraction. Large gene families containing over 100 gene copies in one or more species were deleted, and a p-value of 0.01 was used as the threshold for determining the significance of family size change.

Maximum-likelihood analysis

Maximum-likelihood (ML) analysis was conducted based on 1,662 single-copy homologous proteins identified by OrthoFinder version 2.5.4 (Emms & Kelly 2019). MAFFT version 7.310 (Katoh et al. 2002) was used to conduct multiple sequence alignment with the option “--auto”. Poorly aligned regions in the alignments were removed using trimAl version v1.4.rev15 (Capella-Gutiérrez et al. 2009) based on the gappyout method. The evolution model of each alignment was determined using ModelFinder (Kalyaanamoorthy et al. 2017). All alignments were concatenated using an in-house python script, and the ML tree was reconstructed using IQ-TREE2 version 2.1.3 (Nguyen et al. 2014) with 1000 ultra-fast bootstrap (UFboot) replicates (Hoang et al. 2017). The species *Verticillium nonalfalfae* VnAa140 (NRRL 66861) was selected as the outgroup taxon (Supplementary Table 3).

Divergence time estimation

The RelTime method implemented in MEGA-X (Kumar et al. 2018) being capable of dealing with genome-wide datasets, was used to estimate divergence times. The concatenation-based ML tree with branch lengths was used as the input tree. Three calibration nodes, which were obtained from the TimeTree database (Kumar et al. 2017) and the previous analysis (Shen et al. 2020), were used for the following molecular dating analysis: the *C. graminicola-sidae* split (42.1 to 85 MYA), the *C. graminicola-coccodes* split (at least 16.6 MYA), and the *C. orchidophilum-nymphaeae* split (at least 9.2 MYA).

Identification of secondary metabolic gene clusters

SMGCs were annotated using the stand-alone antiSMASH fungal version 6.0.1 (Blin et al. 2021) with detection strictness set “relaxed”. Two extra features were used: “KnownClusterBlast” to compare predicted SMGCs to the Minimum Information about a Biosynthetic Gene cluster (MIBIG) database (Medema et al. 2015), with known functions, and “SubClusterBlast” to identify similarly characterized gene clusters.

Statistics and visualization

All statistical analyses and visualizations were performed in R v4.1.3 (R Core Team 2022). Pearson correlation analysis was carried out using the function “cor.test” from the base package stats. The Kruskal-Wallis test with Dunn post hoc tests (FSA package) was used to investigate the differences in genomic features between *Colletotrichum* species complexes. Phylogenetic trees were plotted using ggtree (Yu 2020). The heatmap was plotted using pheatmap package (Kolde 2019).

Results

Genomic characteristics of *Colletotrichum*

A total of 102 *Colletotrichum* genomes including 39 species covering ten species complexes and two singletons, were used in this study. Genomic features of these genomes were depicted (Fig. 1), and details were listed in Supplementary Table 2. Genome sizes of *Colletotrichum* species range from 44.15 Mb (*C. truncatum* KLC.C-4) to 109.66 Mb (*C. trifolii* MAFF 305078). Most genomes (81/102) generated by second-generation sequencing platforms (Illumina or Roche 454), consist of more contigs or scaffolds ($\bar{x}$ = 2,149 > 1355; p = 1.009e-05; Fig. 2a) and smaller N50 values ($\bar{x}$ = 0.57 Mb < 3.13 Mb; p = 7.414e-06; Figure 2b) than the genomes generated by third-generation sequencing platforms (PacBio or Oxford Nanopore), indicating that the former is inferior to the latter in assembly contiguity. Variance analysis showed that the completeness of BUSCO assessments was not determined by sequencing platforms (Fig. 2c). The proportion of repeats in genomes varies greatly from 0.28% (*C. fructicola* Nara-gc5) to 54.67% (*C. trifolii* MAFF 305078). Despite the inevitable influence related to the different sequencing and assembly technologies, it is obvious that repeat content is highly correlated to genome size, with a Pearson correlation coefficient of 0.93 (Fig. 2d), and is significantly negatively correlated to GC content with a Pearson correlation coefficient of -0.95 (Fig. 2e). The number of predicted protein-coding genes for 39 *Colletotrichum* species, ranges from 10,354 (*C. truncatum* KLC.C- 4) to 17,809 (*C. australisinense* CGMCC 3.18886). More than 70% of genes from each genome excluding *C. australisinense* CGMCC 3.18886 were assigned at least one PFAM domain. The number of predicted genes is related to non-repeat genome size with a Pearson correlation coefficient of 0.86 (Fig. 2f).

To make statistically reasonable comparisons between different *Colletotrichum* species complexes, seven species complexes, viz. *C. acutatum*, *C. destructivum*, *C. gloeosporioides*, *C. graminicola-caudatum*, *C. orbiculare*, *C. orchidearum*, and *C. spaethianum* species complexes, with at least 3 species were used in subsequent comparative analysis. Average genome sizes between species complexes are significantly different (p = 5.576e-13). Results of pairwise comparisons suggested that significant differences are present across *C. orbiculare, C.*

gloeosporioides, and *C. acutatum* species complexes, whereas no significant difference was observed in other pairwise comparisons (Fig. 2g). The *Colletotrichum orbiculare* species complex possess the largest genome size ($\bar{x}$ = 95.16 Mb), which is significantly larger than the *C. acutatum* species complex ($\bar{x}$ = 50.37 Mb; p = 3.71e-6), and the latter is notably smaller than the *C. gloeosporioides* species complex ($\bar{x}$ = 58.74 Mb; p = 4.28e-9). Significant differences in the proportion of repetitive DNA content between species complexes were also observed (p = 1.593e-06; Fig. 2h). The *C. orbiculare* species complex features the largest repeat content at 49.74%, which is significantly larger than the *C. gloeosporioides* species complex ($\bar{x}$ = 4.83%; p < 1.22e-3). The *Colletotrichum acutatum* complex features the smallest repeat content of 2.43%, which is significantly smaller than the *C. orbiculare* species complex (p < 4.13e-4), the *C. destructivum* species complex ($\bar{x}$ = 14.14%; p < 3.24e-2) and the *C. graminicola-caudatum* species complex ($\bar{x}$ = 16.63%; p < 1.21e-2). Compared to the *C. gloeosporioides* species complex, the *C. graminicola-caudatum* species complex consists of more repeat sequences (p < 3.52e-2). The *Colletotrichum gloeosporioides* species complex features the largest number of predicted gene (15,367), followed by the *C. orchidearum* (14,260), the *C. acutatum* (13,046), the *C. spaethianum* (12,974), and the *C. orbiculare* species complexes (12,376; Fig. 2i). The *C. graminicola-caudatum* species complex features the smallest average gene number of 11,956.

Secretome of *Colletotrichum*

Secreted proteins play important roles in host-fungi interactions. Here, we identified the secreted proteins with signal peptides and the secretome profiles of *Colletotrichum* species were described by analyzing and comparing the richness of four main categories of proteins (CAZymes, proteases, lipases, and SSPs) across different species and species complexes (Fig. 3, Supplementary Table 4). Closely related strains (belonging to the same complex or species) display similar compositions of secretomes. The size of the secretome of each strain ranges from 854 proteins (*C. graminicola* M5.001) to 1,802 proteins (*C. truncatum* KLC.C5), and the proportion of it from 6.98% (*C. sansevieriae* Sa-1-2) to 10.56% (*C. truncatum* CMES1059), which is positively related to the proteome size (R = 0.85, p < 2.2e-16). One interesting observation is that the species *C. australisinense* (CGMCC 3.18886) features the largest proteome (17,809), while the size of the secretome is smaller than most other *Colletotrichum* strains largely due to the contraction of CAZymes and protease. Pairwise comparisons were conducted to investigate the differences in the size of secretomes among seven species complexes (Fig. 4a). The *Colletotrichum gloeosporioides* species complex has the largest secretome ($\bar{x}$ = 1535), followed by the *C. orchidearum* ($\bar{x}$ = 1372), *C. acutatum* ($\bar{x}$ = 1295), *C. orbiculare* ($\bar{x}$ = 1235), *C. spaethianum* ($\bar{x}$ = 1220), and *C. destructivum* species complexes ($\bar{x}$ = 1073). The *C. graminicola-caudatum* species complex has the smallest average size of secretome ($\bar{x}$ = 967). Generally, SSPs are the most abundantly secreted proteins in all species, followed by CAZymes, unclassified secreted proteins, proteases, and finally lipases.

CAZymes

The CAZymes are divided into six functional groups: carbohydrate-binding modules (CBMs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AAs), glycosyltransferases (GTs), and glycoside hydrolase (GHs) (Drula et al. 2021). The number of CAZymes in *Colletotrichum* species ranges from 246 (*C. falcatum* Cf671) to 512 (*C. fructicola* Cg38 S1). GHs and AAs families contribute 39% ~ 51% and 20% ~ 29% in the total number of CAZymes, respectively. CAZyme families except for GTs are greatly positively related to the total number of CAZymes (Fig. 5). Significant differences in the size of CAZymes were observed across multiple pairwise comparisons (Fig. 4b). The *Colletotrichum gloeosporioides* species complex possesses the largest number of CAZymes (479), which is far larger than the smallest number of CAZymes observed in the *C. graminicola-caudatum* species complex (297).

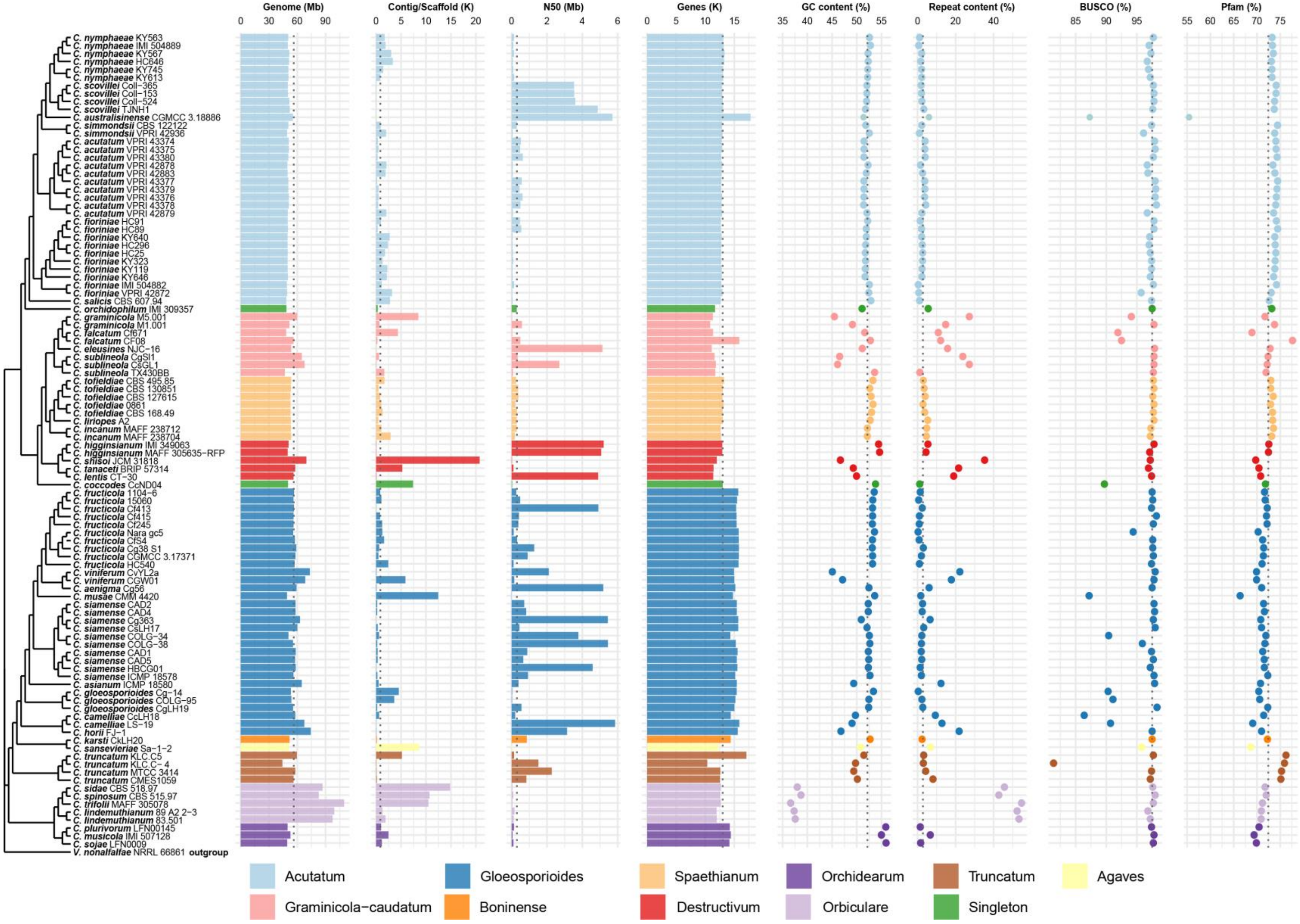

Figure 1 – Genomic properties of the 102 *Colletotrichum* genomes. Strains are ordered according to the genome-scale phylogeny. *Colletotrichum* species complex information is depicted in color. Median values are depicted in the dotted line. Genome (Mb): Genome size. Contig/Scaffold:

The number of scaffolds or contigs. L50 (Mb): N50 length. BUSCO (%): Genome completeness by BUSCO assessment. GC content (%): The percentage of cytosine and guanine bases in genomes. Repeat (%): The proportion of repeats in the genomes. Genes (K): The number of predicted genes. Pfam (%): The proportion of protein-coding genes with at least one PFAM domain.

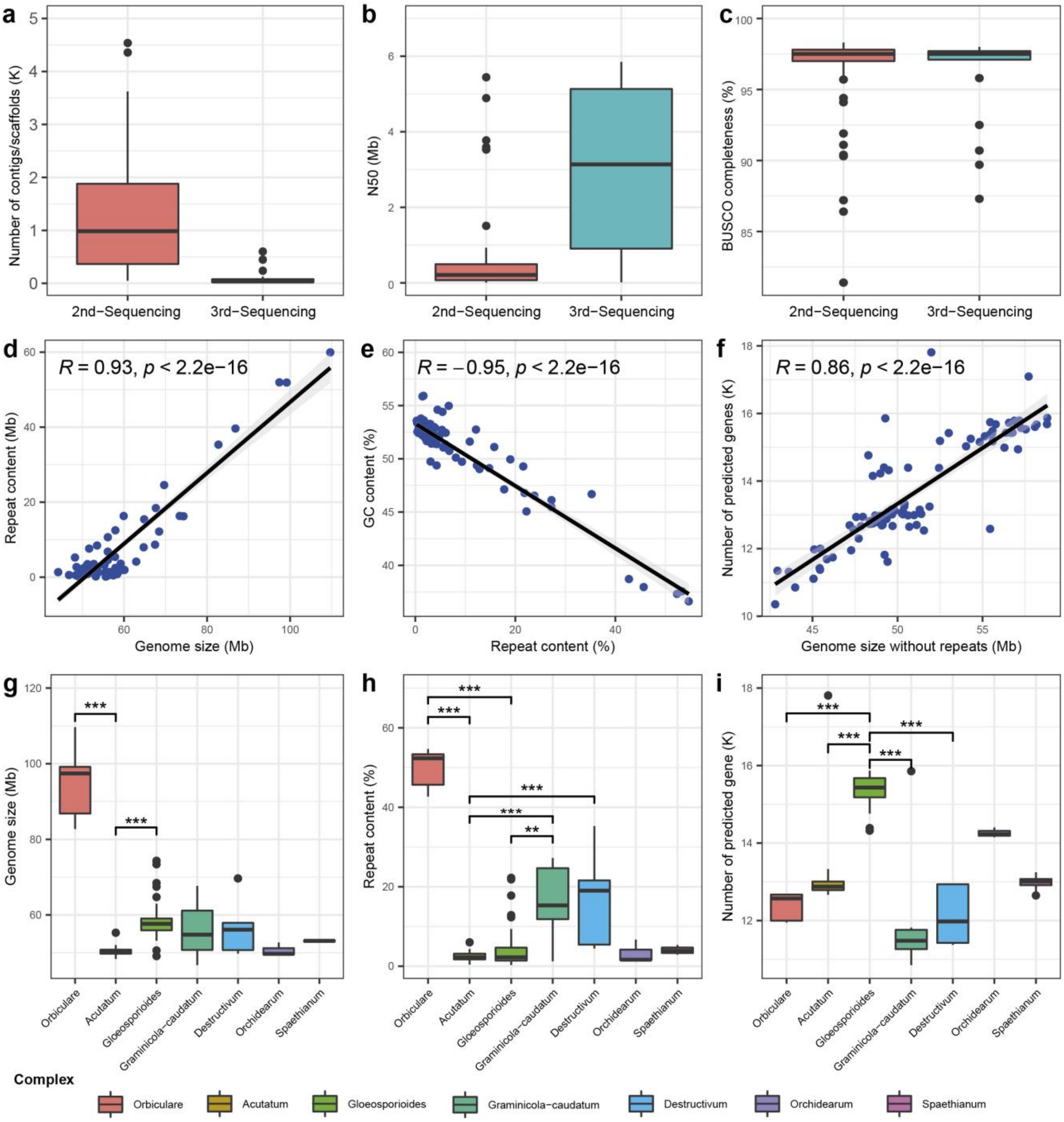

Figure 2 – Comprehensive correlation and comparative analyses of genomic features of *Colletotrichum* genomes. a Number of contigs/scaffolds vs. sequencing platform. b N50 size vs sequencing platform c BUSCO completeness vs. sequencing platform. d Correlation of repeat content and genome size. e Correlation of GC content and genome size. f Correlation of the number of predicted genes and genome size without repeat sequences. g Difference in average genome size among 7 *Colletotrichum* species complexes. h Difference in repeat content among 7 *Colletotrichum* species complexes. i Difference in the number of predicted genes among 7 *Colletotrichum* species complexes. Levels of statistical significance were marked with star (*), ** at 0.05 level, and *** at 0.01 level.

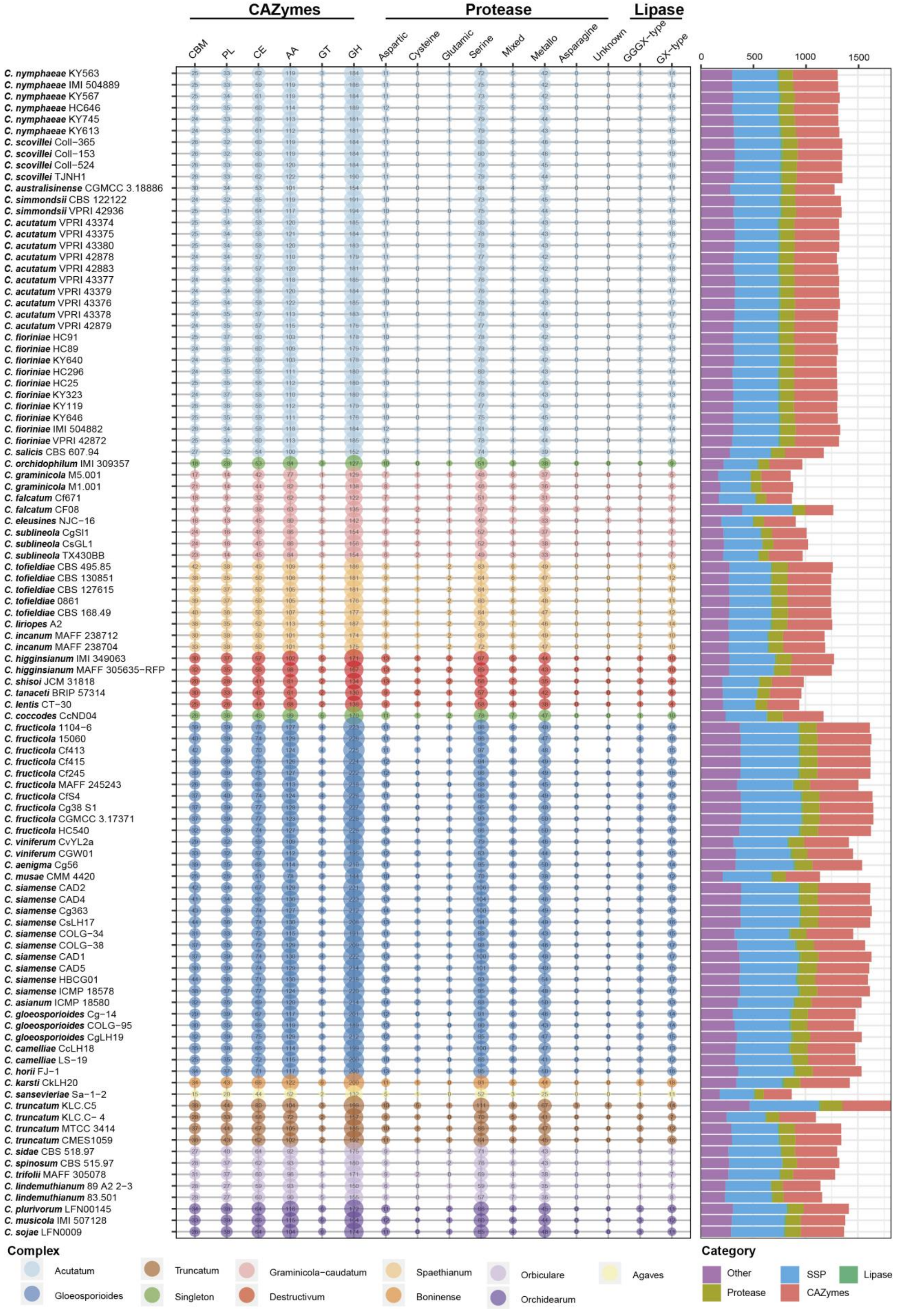

Figure 3 – Composition of 102 *Colletotrichum* secretomes. The strains are ordered according to the genome-scale phylogeny. The number and richness in secreted proteins of each strain are displayed

by bubble size with the exact number. The composition of the secretome of each strain is shown in the stacked bar chart. The *Colletotrichum* species complex information is depicted in color.

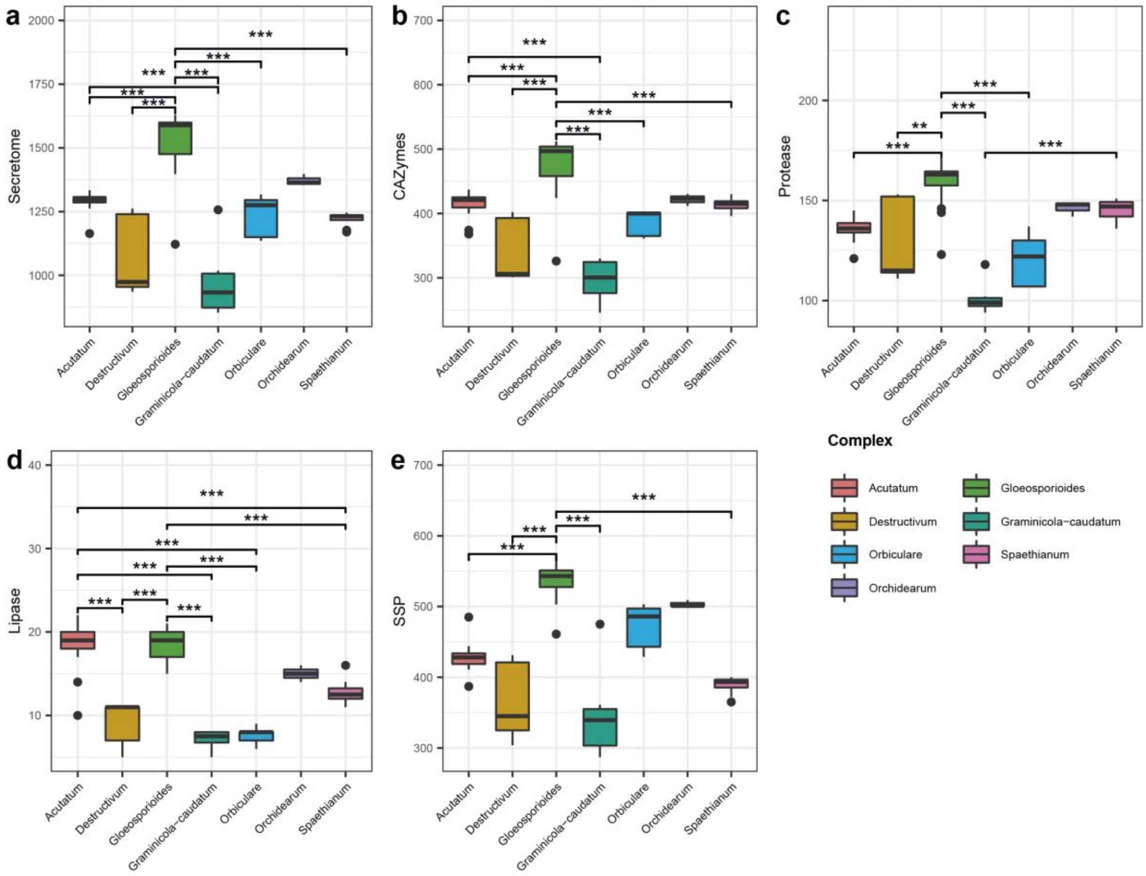

Figure 4 – Difference in components of secretome among 7 *Colletotrichum* species complexes. a Secretome sizes. b Number of CAZymes. c Number of proteases. d Number of lipases. e Small secreted proteins (SSPs). Levels of statistical significance were marked with star (*), ** at 0.05 level and *** at 0.01 level.

Proteases

According to the catalytic type of proteases, proteolytic enzymes are divided into 9 groups: aspartic peptidases, cysteine peptidases, glutamic peptidases, metallo peptidases, asparagine peptidases, mixed peptidases, serine peptidases, unknown catalytic type and threonine peptidases in the *MEROPS* database (Rawlings et al. 2017). The first eight groups of proteases are present in 102 *Colletotrichum* genomes (Fig. 3). The number of total proteases of each strain ranges from 86 (*C. sansevieriae* Sa-1-2) to 210 (*C. truncatum* KLC.C5), and significant differences were observed across multiple pairwise comparisons (Fig. 4c). Serine peptidases and metallo peptidases account for 47%–62% and 26%–38% of total protease, respectively. Correlation analysis revealed that both the number of serine peptidases and metallo peptidases are greatly positively related to the total number of proteases (Fig. 6). MEROPS family S8 peptidases, also known as the subtilase family, are the most abundant in secreted serine proteases of most *Colletotrichum* species.

Lipases

Lipase is a class of enzymes that catalyze the hydrolysis of long-chain triglycerides. In the Lipase Engineering Database (LED), lipases are divided into three types (GX-, GGGX-, or Y-type) by their oxyanion hole signature. This study identified the presentence of GX-type and GGGX-type lipases, and the latter accounts for 69%–100% of total lipases. The number of lipases from each strain ranges from 5 (*C. lentis* CT-30 *and C. graminicola* M5.001) to 24 (*C. karsti* CkLH20).

Significant differences in the number of lipases were observed between species complexes (Fig. 4d). The *Colletotrichum acutatum* species complex possesses the largest number of secreted lipases ($\bar{x}$ = 19), followed by the *C. gloeosporioides* ($\bar{x}$ = 18), *C. orchidearum* ($\bar{x}$ = 15), *C. spaethianum* ($\bar{x}$ = 13), *C. destructivum* ($\bar{x}$ = 9), *C. orbiculare* ($\bar{x}$ = 8), and finally *C. graminicola-caudatum* species complexes ($\bar{x}$ = 7).

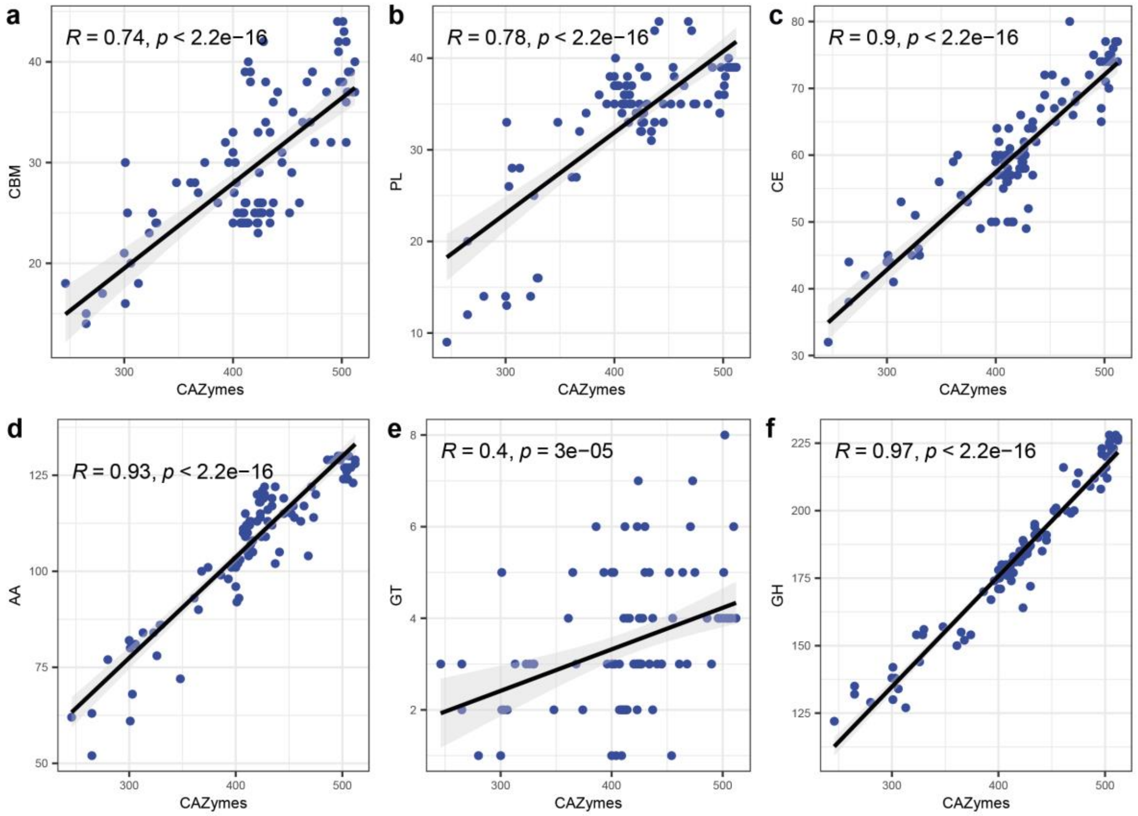

Figure 5 – Correlation analyses of the size of CAZymes and the size of subfamilies. a CBM family. b PL family. c CE family. d AA family. e GT family. f GH family.

SSPs

SSPs are the second largest group of proteins in the secretome of *Colletotrichum*, and the number of SSPs ranges from 287 to 656. A strongly positive correlation between the number of SSPs and the size of the secretome with a Pearson correlation coefficient of 0.94 (p-value < 2.2e-16) was observed. Pairwise differences in the number of SSPs across *Colletotrichum* species complexes were investigated (Fig. 4e), and the largest average number of SSPs ($\bar{x}$ = 539) was observed in the *C. gloeosporioides* species complex which is significantly larger than the *acutatum* ($\bar{x}$ = 426), *destructivum* ($\bar{x}$ = 365) and *spaethianum* species complexes ($\bar{x}$ = 389).

Colletotrichum contains diverse metabolic gene clusters

To understand the diversity and composition of SMGCs in *Colletotrichum* genomes, we detected the SMGCs for each genome, and a total of 7,971 putative SMGCs covering 20 types across 102 *Colletotrichum* genomes, were determined (Fig. 7, Supplementary Table 5). The number of SMGCs per genome ranges from the smallest at 50, observed in *C. shisoi* JCM 31818 to the largest at 108 observed in *C. fructicola* CGMCC 3.17371 and the average number of SMGCs per *Colletotrichum* genome is 78. Hierarchical cluster analysis revealed that strains belonging to the same species or species complex shared a similar profile of SMGCs. Furthermore, we also noted the intraspecific differences, which mainly were caused by copy variation. The *Colletotrichum gloeosporioides* species complex contained the largest average number of SMGCs ($\bar{x}$ = 94), significantly larger than the *C. acutatum* ($\bar{x}$ = 68), *C. destructivum* ($\bar{x}$ = 69), *C. graminicola-caudatum* ($\bar{x}$ = 72) and *C. orbiculare* species complexes ($\bar{x}$ = 63) ($p < 0.05$). The most common SMGCs for the production of non-ribosomal peptide synthetase (NRPS; 1,515 in total), NRPS-like (1,016 in total), type 1 polyketide synthase (T1PKS; 2,883 in total), indole (762 in total), terpene (1,454 in total), and T3PKS (154 in total), were present in all genomes. T1PKS was the most widely distributed and abundant SMGC, the products of which are responsible for the synthesis of chemically diverse polyketide natural products (Newman & Cragg 2012). Multiple strain-specific SMGCs including phosphonates, resorcinol, ectoine, linear azoline-containing peptides, and non-alpha poly-amino acids like e-poly lysin (NAPAA) present in *C. falcatum* CF08, trans-AT polyketide synthase (transAT-PKS) presenting in *C. acutatum* VPRI 42879, as well as lanthipeptides class I, lanthipeptides class II, arylpolyene and unspecified ribosomally synthesized and post-translationally modified peptide product (RiPP-like) present in *C. truncatum* KLC.C5 were observed.

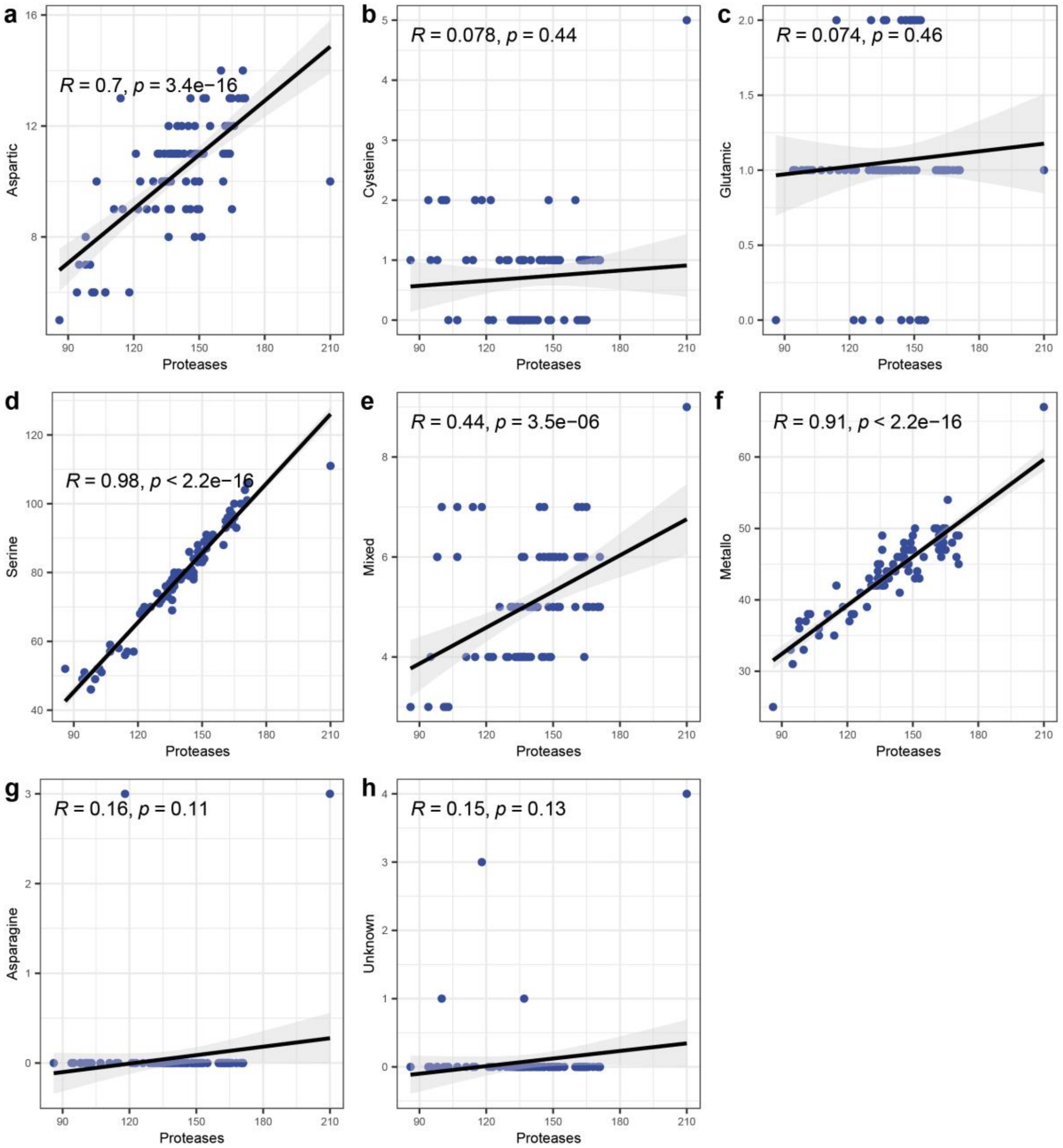

Figure 6 – Correlation analyses of the size of proteases and the size of subfamilies. a Aspartic family. b Cysteine family. c Glutamic family. d Serine family. e Mixed family. f Metallo family. g Asparagine family. h Unknown type.

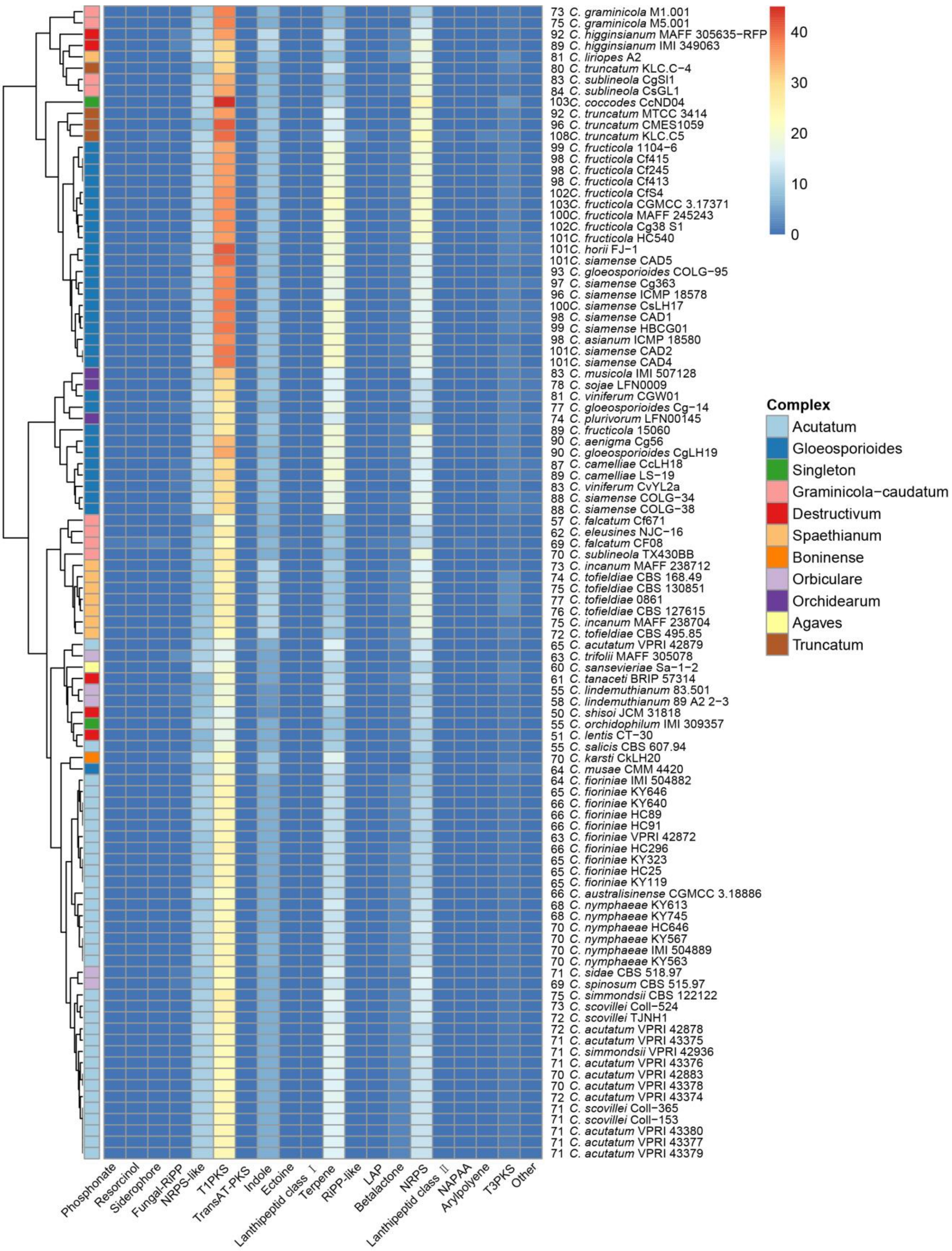

Figure 7 – SMGCs of 102 *Colletotrichum* strains. *Colletotrichum* strains with similar components of SMGCs were clustered. Heatmap represented the number of SMGCs of each fungal species. *Colletotrichum* species complex information is depicted in the first column with different colors. The total number of SMGCs was noted between heatmap and strain labels. RiPP-like, other unspecified ribosomally synthesized and post-translationally modified peptide product. Fungal-RiPP, fungal RiPP with POP or UstH peptidase types and a modification. NRPS,

non-ribosomal peptide synthetase. NRPS-like, non-ribosomal peptide synthetase containing at least two NRPS-specific domains and another domain or one NRPS A domain combined with NAD binding 4 domain or short-chain dehydrogenase. T1PKS, Type 1 polyketide synthase. T1PKS, Type 3 polyketide synthase. transAT-PKS, trans-AT polyketide synthase. LAP, linear azole-containing peptides. NAPAA, non-alpha poly-amino acids like e-poly lysin.

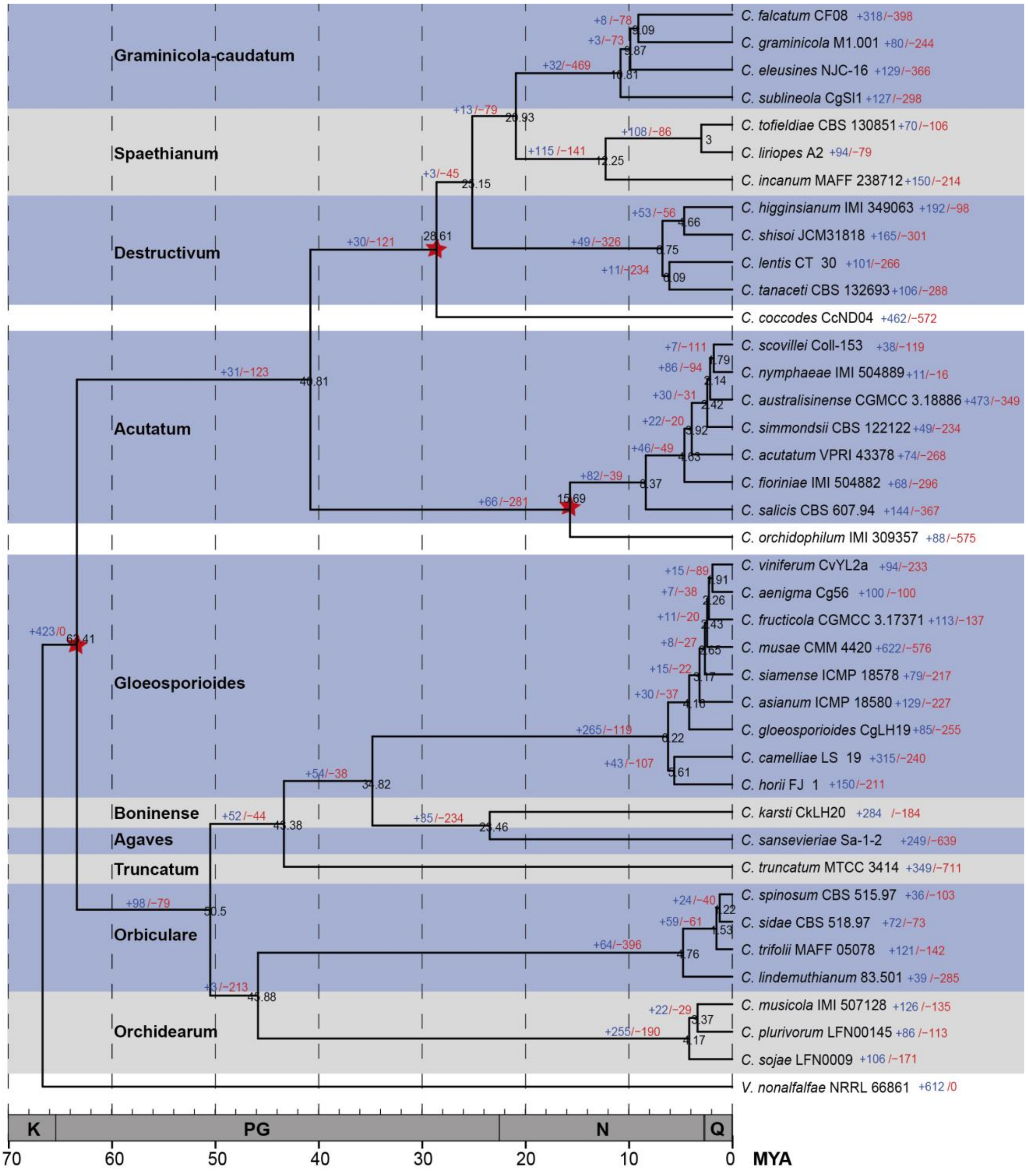

Figure 8 – Gene family expansion and contraction of 39 *Colletotrichum* species. ML tree (-lnL = -9630847.467) was constructed based on a combined dataset by concatenating 1,662 single-copy proteins. Calibration points were noted using red stars. Node ages are given at the nodes, and divergence times are shown in millions of years. The number of significantly expanded (blue) and contracted (P < 0.01) gene families (red) were given on the corresponding branches. MYA, million years ago. K, Cretaceous; PG, Paleogene; N, Neogene; Q, Quaternary.

Gene family expansion and contraction along with the evolutionary history

The concatenation-based ML tree, the RealTime method, and three-time calibration nodes were used to infer the divergence time of *Colletotrichum* species (Fig. 8). Estimated divergence times of species complexes were compared to previous studies. In this study, the youngest species complex *C. acutatum* species complex was dated to 15.69 MYA and the oldest species complexes the *C. orbiculare* species complex and *C. orchidearum* species complex diverged at 45.88 MYA, whereas Bhunjun et al. (2021) recognized the *C. magnum* species complex and *C. orchidearum* species complex as the youngest species complexes diverged at 14.8 MYA. The difference may originate from the missing *C. magnum* species complex in our dataset. Therefore, we could not trace the most recent common ancestor of the *C. magnum* species complex and *C. orchidearum* species complex. The divergence times of most *Colletotrichum* species complexes are in accordance with the estimation of previous studies (Liang et al. 2018b, Shen et al. 2020, Bhunjun et al. 2021). The *Colletotrichum graminicola-caudatum* species complex and *C. spaethianum* species complex diverged at 20.93 MYA. The origin of the *C. destructivum* species complex took place 25.15 MYA. The *Colletotrichum truncatum* species complex was dated to 43.38 MYA, and *C. gloeosporioides* diverged with the *C. boninense* species complex and *C. agaves* species complex 34.82 MYA. The origin of the latter two species complexes took place 23.46 MYA.

A total of 23,139 gene families were identified from 39 *Colletotrichum* species and the outgroup. These gene families were further subjected to detect gene family expansion and contraction at the species level in light of the software failing to detect the relatively small intraspecies changes in family size. A total of 449 gene families changed significantly (expanded or contracted; $p < 0.01$) along the phylogeny. Gene Ontology (GO) annotation results (Supplementary Table 6) showed most of these gene families are related to molecular function, suggesting that changes in gene families related to molecular function play more important roles in speciation. Pfam annotation (Supplementary Table 7) revealed that the top five most abundant Pfam domains are major facilitator superfamily (PF07690), methyltransferase domain (PF13489), cytochrome P450 family (PF00067), sugar (and other) transporter (PF00083) and heterokaryon incompatibility protein (PF06985) in descending order. MFS, one of the two largest families of membrane transporters (Reddy et al. 2012), play an important role in the multidrug resistance of fungi (Chen et al. 2017). Cytochrome P450 proteins are involved in many cellular processes, including the conversion of intermediates of primary and secondary metabolic pathways, as well as the clearance of natural and environmental pollutants (Črešnar & Petrič 2011). Compared to the outgroup V*erticillium nonalfalfae* VnAa140 (NRRL 66861), a soilborne vascular wilt pathogen (Kasson et al. 2019), a total of 423 significantly expanded gene families were observed in the *Colletotrichum* lineage. The most abundant GO term of these families was involved in transmembrane transporter activity, followed by oxidoreductase activity (Supplementary Table 8). Concerning the number of changed families of each species, no significant difference between species complex or singleton species was observed, suggesting that gene duplication and loss are less useful in delineating species complexes.

Discussion

The genome size of *Colletotrichum godetiae* was estimated at approximately 36 Mb in length (Liu et al. 2022), which is the smallest genome in *Colletotrichum* to date. However, the strain was excluded in this study due to the low assembly completeness with only 64.50% of BUSCO genes. This difference in assembly contiguity was largely caused by the intrinsic drawbacks and pitfalls of the second-generation sequencing technique (van Dijk et al. 2018), but it does not affect the completeness of assembly that indirectly suggests that second-generation sequencing techniques are capable of obtaining-high-quality genomes for most phylogenomic analyses to some extent. Repeat sequences, especially transposable elements (TEs), play a vital role in fungal genome evolution, especially in plant pathogens (Raffaele & Kamoun 2012, Muszewska et al. 2019). Smaller genomes of *Colletotrichum* species having fewer repeats was in line with previous studies (Haridas et al. 2020, Kidwell 2002), and larger genomes with lower GC content, suggesting that transposable elements as major components of repeats that are key drivers responsible for generating genomic diversity in *Colletotrichum* species (Tsushima et al. 2019), similar to TE functions in other eukaryotic genomes (Kapusta and Suh 2017, Shao et al. 2019, Chen et al. 2020). The number range of predicted genes of *Colletotrichum* species in this study (10,354~17,809) was larger than the data obtained by Liu et al. (2022) (8,424~14,841), but similar to the result (15,744~16,538) of Fu et al. (2020). The number of predicted genes and non-repeat genome size displayed a strong positive correlation, which was also observed by Haridas et al. (2020) in Dothideomycetes genomes.

The largest proteome of *Colletotrichum* does not include the most secreted proteins, and this was also observed in the symbiont *Laccaria bicolor* (Hydnangiaceae, Basidiomycota) and several animal pathogens, such as *Paracoccidioides brasiliensis* (Ajellomycetaceae, Ascomycota), *Histoplasma capsulatum* (Ajellomycetaceae, Ascomycota) and *Coccidioides immitis* (Onygenaceae, Ascomycota) (Lowe & Howlett 2012). The number of CAZymes of *Colletotrichum* species ranges from 246 to 512, moreover this number is generally larger than the human fungal pathogens such as *Candida* spp. (Incertae sedis, Ascomycota) and *Malassezia* spp. (Malasseziaceae, Basidiomycota) (Zhao et al. 2013), and similar to the plant pathogens such as *Diaporthe* spp. (Diaporthaceae, Ascomycota) (Mena et al. 2022). GHs and AAs families are chief components of CAZymes in *Colletotrichum* species. The former can hydrolytically cleave glycosidic bonds in oligo- or polysaccharides, including cellulose and hemicellulose. The latter is responsible for the breakdown of lignocellulose from the plant cell wall (Levasseur et al. 2013). Compared to animal pathogens, increased CAZymes might be useful for increasing the ability to break down plant cell walls and gain entries into the plant tissue. The largest average number of CAZymes was observed in *C. gloeosporioides* species complex, which might be one of the feasible explanations for why the species of *C. gloeosporioides* species complex are capable of infecting 283 host plants at the various developmental stages as pathogens or endophytes (Talhinhas & Baroncelli 2021). The subtilase family of serine peptidases, with optimum activity at neutral-mildly alkali pH, features the second-largest serine proteases, and alkalinization of host tissue has been associated with fungal pathogenicity (Fernandes et al. 2017). Furthermore, members of the subfamily subtilisin are capable of degrading plant cell wall glycoproteins or pathogenesis-related proteins, which may be important for penetration and colonization (Olivieri et al. 2002). Lipases are less commonly secreted proteins in *Colletotrichum* species, but they are still larger than the ectomycorrhizal symbiont *Pisolithus tinctorius* and the yeast *Pichia stipites* that only have two lipases (Pellegrin et al. 2015). Accordingly, we should not ignore potential virulence-related lipases, such as *FGL1* and *FgATG15* secreted by *Fusarium graminearum* (Voigt et al. 2005, Nguyen et al. 2011a). Some SSPs are related to host-pathogen interactions (Giraldo & Valent 2013), while other SSPs have been recognized as effectors in mutualistic fungi (Kloppholz et al. 2011, Plett et al. 2011). Similar effector proteins of pathogenic fungi have been reported in saprotrophic fungi (Rovenich et al. 2014, Seidl et al. 2015), suggesting that SSPs play versatile roles in fungal development. Pellegrin et al. (2015) found a positive correlation between the secretome size and the number of SSPs in ectomycorrhizal fungi, and this relationship also was observed in this study. Krijger et al. (2014) found that the size of the predicted secretome is related to phylogenetic relationships, while Alfaro et al. (2014) reported that secretome differences are correlated with fungal lifestyles. In the present study, we found that evolutionarily close genomes (same species or same species complex) displayed similar components of the secretome. The relationship between the lifestyle and secretome was not investigated, because of the difficulties in distinguishing necrotrophic, hemibiotrophic, and endophytic lifestyles (Cannon et al. 2012, Liu et al. 2022). Herein, secretomes were fully characterized that would bolster understanding of the role of the secretome in speciation.

Compared to the recent studies, the average number of SMGCs per *Colletotrichum* genome was 78, which was slightly larger than xylarialean genomes ($\bar{x} = 71$, n = 96) (Franco et al. 2022) and *Penicillium* genomes ($\bar{x} = 55$, n = 24) (Nielsen et al. 2017), and about two times larger than Dothideomycetes genomes ($\bar{x} = 33$, n = 101) (Gluck-Thaler et al. 2020). The dramatic decrease in the number of SMGCs in Dothideomycetes might be caused by the use of different prediction tools, as well as niche-specific selection. Among these SMGCs, resorcinol was widely known as an antifungal compound produced by diverse plant species, and not reported in fungi barring *Fusarium culmorum* F1 (Zarnowski et al. 2000) and *Penicillium sclerotiorum* LM 5679 (Oliveira et al. 2021).

Differences in estimated divergence times between this study and the study of Bhunjun et al. (2021) were observed. One possible explanation for this may be that we used genome-scale data and different calibration points. Based on divergence time, Zhao et al. (2016) proposed to divide *Agaricus* into five subgenera and 20 sections. Bhunjun et al. (2021) recognized *Colletotrichum* species complexes based on the divergence time of species complexes ranging from approximately 4.8 MYA to 32.2 MYA. Li et al. (2021) found that the current taxonomic framework is broadly consistent with both genome sequence divergence and divergence time at the higher taxonomic level. Although time-dating has become an alternative in ranking fungi, we have to bear in mind that divergence time estimation intrinsically carries a degree of uncertainty. Accuracy in molecular clock studies depends on multiple factors, such as taxonomic sampling size, substitution rate variation, number of loci, the selection evolution model and calibration choices (van Tuinen & Hadly 2004, Voloch & Schrago 2012, Duchêne et al. 2014, Schenk 2016, Carruthers et al. 2019). van Tuinen and Torres (2015) reported that ages of shallow nodes (<10 MYA) might be overestimated due to bias and low precision, even when appropriate fossil priors and best substitution models are used. The genome-scale multi-locus dataset in this work decreased the impact of substitution rate variation caused by incomplete lineage sorting. However, most studies involving divergence time estimation of *Colletotrichum* spp., including this study, used the secondary calibrations generated from previous studies, which further increased the uncertainty in age estimates. Therefore, in the absence of direct fossil evidence, insufficient taxonomic sampling size and number of loci, we should avoid conducting time dating at low taxonomic levels.

Baroncelli et al. (2016) reported that contracted CAZyme- and protease-coding gene families in *Colletotrichum* species are related to relatively narrow host ranges. However, this conclusion was based on ten *Colletotrichum* species with one genome of each. In this study, we used publicly available genome data of *Colletotrichum* spp., most of which are from non-type materials and without detailed host information. Therefore, we stressed the gene family changes along with the phylogeny instead of investigating the relationship between gene family evolution and the host range. Vela-Corcía et al. (2019) reported that MFS functions as fungitoxic compound efflux and can provide tolerance to phytoanticipins. Cowan et al. (2022) found that cytochrome P450 proteins play important roles in niche adaptation. The abundant genes of MFS and cytochrome P450 protein family in significantly changed families were observed in this study, suggesting that gene expansion and contraction along *Colletotrichum* phylogeny may be associated with species adaptation to the environment.

In conclusion, our study comprehensively characterized the genomic profiles of *Colletotrichum* genomes and unveiled that TEs are key drivers leading to genomic diversity. Pairwise comparisons of genomic properties had revealed inherent genetic difference among *Colletotrichum* complexes. The expansions and contractions of gene families along the phylogeny within the genus could be important determinants in environmental adaptation. We also systematically investigated the SMGCs in *Colletotrichum* genomes, and the result suggested that *Colletotirchum* species are a rich source of natural products. These results provide a solid foundation for understanding the genetic, population genomic, and evolutionary studies of *Colletotrichum* species, and set the stage for exploring these natural products and elucidating their biological roles.

Acknowledgements

We thank Xing-Xing Shen at Zhejiang University for his valuable suggestions in data collecting and divergence time estimation. We thank the computing resources provided by high-performance computing platform at the information technology center, University of Electronic Science and Technology of China. Kevin D Hyde thanks the National Research Council of Thailand (NRCT) grant “Total fungal diversity in a given forest area with implications towards species numbers, chemical diversity and biotechnology” (grant no. N42A650547).

References

Ahn JH, Walton JD. 1996 – Chromosomal organization of TOX2, a complex locus controlling host-selective toxin biosynthesis in *Cochliobolus carbonum*. The Plant Cell 8(5), 887–897. Doi 10.1105/tpc.8.5.887

Akamatsu H, Taga M, Kodama M, Johnson R et al. 1999 – Molecular karyotypes for *Alternaria* plant pathogens known to produce host-specific toxins. Current Genetics 35(6), 647–656. Doi 10.1007/s002940050464

Alfaro M, Oguiza JA, Ramírez L, Pisabarro AG. 2014 – Comparative analysis of secretomes in basidiomycete fungi. Journal of Proteomics 102, 28–43. Doi 10.1016/j.jprot.2014.03.001

Ashwin NMR, Barnabas L, Ramesh Sundar A, Malathi P et al. 2017 – Comparative secretome analysis of *Colletotrichum falcatum* identifies a cerato-platanin protein (EPL1) as a potential pathogen-associated molecular pattern (PAMP) inducing systemic resistance in sugarcane. Journal of Proteomics 169, 2–20. Doi 10.1016/j.jprot.2017.05.020

Bao W, Kojima KK, Kohany O. 2015 – Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 6(1), 11. Doi 10.1159/000084979

Baroncelli R, Amby DB, Zapparata A, Sarrocco S et al. 2016 – Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. BMC Genomics 17(1), 555. Doi 10.1186/s12864-016-2917-6

Bhunjun CS, Phukhamsakda C, Jayawardena RS, Jeewon R et al. 2021 – Investigating species boundaries in *Colletotrichum*. Fungal Diversity 107(1), 107–127. Doi 10.1016/j.funbio.2019.05.002

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z et al. 2021 – antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Research 49(W1), W29–W35. Doi 10.1093/nar/gkab335

Blum M, Chang H-Y, Chuguransky S, Grego T et al. 2020 – The InterPro protein families and domains database: 20 years on. Nucleic Acids Research 49(D1), D344–D354. Doi 10.1093/nar/gkaa977

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021 – BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics and Bioinformatics 3(1), lqaa108. Doi 10.1093/nargab/lqaa108

Cannon PF, Damm U, Johnston PR, Weir BS. 2012 – *Colletotrichum* - current status and future directions. Studies in Mycology 73(1), 181–213. Doi 10.3114/sim0014

Cano J, Guarro J, Gene J. 2004 – Molecular and morphological identification of *Colletotrichum* species of clinical interest. Journal of Clinical Microbiology 42(6), 2450–2454. Doi 10.1128/JCM.42.6.2450-2454.2004

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009 – trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15), 1972–1973. Doi 10.1093/bioinformatics/btp348

Carruthers T, Sanderson MJ, Scotland RW. 2019 – The implications of lineage-specific rates for divergence time estimation. Systematic Biology 69(4), 660–670. Doi 10.1093/sysbio/syz080

Chan PP, Lowe TM. 2019 – tRNAscan-SE: searching for tRNA genes in genomic sequences. In: Kollmar M. (eds) Gene prediction. Springer, 1962, 1–14. Doi 10.1007/978-1-4939-9173-0_1

Chang TH, Hassan O, Lee YS. 2018 – First report of anthracnose of Japanese Plum (*Prunus salicina*) caused by *Colletotrichum nymphaeae* in Korea. Plant Disease 102(7), 1461–1461. Doi 10.1094/PDIS-01-18-0018-PDN

Chen J, Lu L, Robb SMC, Collin M et al. 2020 – Genomic diversity generated by a transposable element burst in a rice recombinant inbred population. Proceedings of the National Academy of Sciences 117(42), 26288–26297. Doi 10.1073/pnas.2015736117

Chen L-H, Tsai H-C, Yu P-L, Chung K-R. 2017 – A major facilitator superfamily transporter-mediated resistance to oxidative stress and fungicides requires *Yap1*, *Skn7*, and *MAP* kinases in the citrus fungal pathogen *Alternaria alternata*. Plos One 12(1), e0169103. Doi 10.1371/journal.pone.0169103

Collemare J, Billard A, Böhnert HU, Lebrun M-H. 2008 – Biosynthesis of secondary metabolites in the rice blast fungus *Magnaporthe grisea*: the role of hybrid PKS-NRPS in pathogenicity. Mycological Research 112(2), 207–215. Doi 10.1016/j.mycres.2007.08.003

Cowan A, Skrede I, Moody SC. 2022 – Cytochrome P450 complement may contribute to niche adaptation in serpula wood-decay fungi. Journal of fungi 8(3), 283. Doi 10.3390/jof8030283

Črešnar B, Petrič Š. 2011 – Cytochrome P450 enzymes in the fungal kingdom. Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics 1814(1), 29–35. Doi 10.1016/j.bbapap.2010.06.020

Crous PW, Hawksworth DL, Wingfield MJ. 2015 – Identifying and naming plant-pathogenic fungi: past, present, and future. Annual Review of Phytopathology 53(1), 247–267. Doi 10.1146/annurev-phyto-080614-120245

De Silva DD, Crous PW, Ades PK, Hyde KD, Taylor PWJ. 2017 – Life styles of *Colletotrichum* species and implications for plant biosecurity. Fungal Biology Reviews 31(3), 155–168. Doi 10.1016/j.fbr.2017.05.001

Dean R, Van Kan JA, Pretorius ZA, Hammond-Kosack KE et al. 2012 – The Top 10 fungal pathogens in molecular plant pathology. Molecular Plant Pathology 13(4), 414–430. Doi 10.1111/j.1364-3703.2011.00783.x

Drula E, Garron M-L, Dogan S, Lombard V et al. 2021 – The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Research 50(D1), D571–D577. Doi 10.1093/nar/gkab1045

Duchêne S, Lanfear R, Ho SYW. 2014 – The impact of calibration and clock-model choice on molecular estimates of divergence times. Molecular Phylogenetics and Evolution 78, 277–289. Doi 10.1016/j.ympev.2014.05.032

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007 – Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols 2(4), 953–971. Doi 10.1038/nprot.2007.131

Emms DM, Kelly S. 2019 – OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology 20(1), 238. Doi 10.1186/s13059-019-1832-y

Feldman D, Yarden O, Hadar Y. 2020 – Seeking the roles for fungal small-Secreted proteins in affecting saprophytic lifestyles. Frontiers in Microbiology 11, 455. Doi 10.3389/fmicb.2020.00455

Fernandes TR, Segorbe D, Prusky D, Di Pietro A. 2017 – How alkalinization drives fungal pathogenicity. Plos Pathogens 13(11), e1006621. Doi 10.1371/journal.ppat.1006621

Flemings A. 1929 – On the antibacterial action of cultures of a penicillium with special reference to their use in the isolation of *B. influenza*. British Journal of Experimental Pathology 10, 226–236.

Flynn JM, Hubley R, Goubert C, Rosen J et al. 2020 – RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences 117(17), 9451–9457. Doi 10.1073/pnas.1921046117

Franco MEE, Wisecaver JH, Arnold AE, Ju Y-M et al. 2022 – Ecological generalism drives hyperdiversity of secondary metabolite gene clusters in xylarialean endophytes. New Phytologist 233(3), 1317–1330. Doi 10.1111/nph.17873

Fu F-F, Hao Z, Wang P, Lu Y et al. 2020 – Genome sequence and xomparative analysis of *Colletotrichum gloeosporioides* isolated from *Liriodendron* leaves. Phytopathology 110(7), 1260–1269. Doi 10.1094/PHYTO-12-19-0452-R

Gan P, Ikeda K, Irieda H, Narusaka M et al. 2013 – Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytologist 197(4), 1236–1249. Doi 10.1111/nph.12085

Giraldo MC, Valent B. 2013 – Filamentous plant pathogen effectors in action. Nature Reviews Microbiology 11(11), 800–814. Doi 10.1038/nrmicro3119

Gluck-Thaler E, Haridas S, Binder M, Grigoriev IV et al. 2020 – The architecture of metabolism maximizes biosynthetic diversity in the largest class of fungi. Molecular Biology and Evolution 37(10), 2838–2856. Doi 10.1093/molbev/msaa122

Goodwin S, McPherson JD, McCombie WR. 2016 – Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17(6), 333–351. Doi 10.1038/nrg.2016.49

Grabherr MG, Haas BJ, Yassour M, Levin JZ et al. 2011 – Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29(7), 644–652. Doi 10.1038/nbt.1883

Gremme G, Brendel V, Sparks ME, Kurtz S. 2005 – Engineering a software tool for gene structure prediction in higher organisms. Information and Software Technology 47(15), 965–978. Doi 10.1016/j.infsof.2005.09.005

Haridas S, Albert R, Binder M, Bloem J et al. 2020 – 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. Studies in Mycology 96, 141–153. Doi 10.1016/j.simyco.2020.01.003

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2017 – UFBoot2: Improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35(2), 518–522. Doi 10.1093/molbev/msx281

Hongsanan S, Maharachchikumbura SSN, Hyde KD, Samarakoon MC et al. 2017 – An updated phylogeny of Sordariomycetes based on phylogenetic and molecular clock evidence. Fungal Diversity 84(1), 25–41. Doi 10.1007/s13225-017-0384-2

Hyde KD, Nilsson RH, Alias SA, Ariyawansa HA et al. 2014 – One stop shop: backbones trees for important phytopathogenic genera: I (2014). Fungal Diversity 67(1), 21–125. Doi 10.1007/s13225-014-0298-1

Jayawardena RS, Bhunjun CS, Hyde KD, Gentekaki E, Itthayakorn P. 2021 – *Colletotrichum*: lifestyles, biology, morpho-species, species complexes and accepted species. Mycosphere 12(1), 519–669. Doi 10.5943/mycosphere/12/1/7

Jayawardena RS, Hyde KD, Chen YJ, Papp V et al. 2020 – One stop shop IV: taxonomic update with molecular phylogeny for important phytopathogenic genera: 76–100 (2020). Fungal Diversity 103(1), 87–218. Doi 10.1007/s13225-020-00460-8

Käll L, Krogh A, Sonnhammer ELL. 2004 – A combined transmembrane topology and signal peptide prediction method. Journal of Molecular Biology 338(5), 1027–1036. Doi 10.1016/j.jmb.2004.03.016

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017 – ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14(6), 587–589. Doi 10.1038/nmeth.4285

Kanpiengjai A, Khanongnuch C, Lumyong S, Kummasook A, Kittibunchakul S. 2020 – Characterization of *Sporidiobolus ruineniae* A45.2 cultivated in tannin dubstrate for use as a potential multifunctional probiotic Yeast in aquaculture. Journal of Fungi 6(4), 378. Doi 10.3390/jof6040378

Kapusta A, Suh A. 2017 – Evolution of bird genomes – a transposon's-eye view. Annals of the New York Academy of Sciences 1389(1), 164–185. Doi 10.1111/nyas.13295

Kasson MT, Kasson LR, Wickert KL, Davis DD et al. 2019 – Genome Sequence of a Lethal Vascular Wilt Fungus, *Verticillium nonalfalfae*, a biological control used against the invasive *Ailanthus altissima*. Microbiology Resource Announcements 8(4), e01619–01618. Doi 10.1111/nyas.13295

Katoh K, Misawa K, Kuma Ki, Miyata T. 2002 – MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30(14), 3059–3066. Doi 10.1093/nar/gkf436

Keller O, Kollmar M, Stanke M, Waack S. 2011 – A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics 27(6), 757–763.

Doi 10.1093/bioinformatics/btr010

Kidwell MG. 2002 – Transposable elements and the evolution of genome size in eukaryotes. Genetica 115(1), 49–63. Doi 10.1023/a:1016072014259

Kim JW, Shim SH. 2019 – The fungus *Colletotrichum* as a source for bioactive secondary metabolites. Archives of Pharmacal Research 42(9), 735–753. Doi 10.1007/s12272-019-01142-z

Kim K-T, Jeon J, Choi J, Cheong K et al. 2016 – Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. Frontiers in Plant Science 7, 1886. Doi 10.3389/fpls.2016.00186

Kloppholz S, Kuhn H, Requena N. 2011 – A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. Current Biology 21(14), 1204–1209. Doi 10.1016/j.cub.2011.06.044

Kolde R. 2019 – pheatmap: Pretty Heatmaps.

Krijger J-J, Thon MR, Deising HB, Wirsel SGR. 2014 – Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. BMC Genomics 15(1), 722. Doi 10.1186/1471-2164-15-722

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001 – Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Journal of Molecular Biology 305(3), 567–580. Doi 10.1006/jmbi.2000.4315

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018 – MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35(6), 1547–1549. Doi 10.1093/molbev/msy096

Kumar S, Stecher G, Suleski M, Hedges SB. 2017 – TimeTree: A resource for timelines, timetrees, and divergence times. Molecular Biology and Evolution 34(7), 1812–1819. Doi 10.1093/molbev/ms

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(1), 41. Doi 10.1186/1754-6834-6-41

Li W, Godzik A. 2006 – Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13), 1658–1659. Doi 10.1093/bioinformatics/btl158

Li Y, Steenwyk JL, Chang Y, Wang Y et al. 2021 – A genome-scale phylogeny of the kingdom Fungi. Current Biology 31(8), 1653–1665. e1655. Doi 10.1016/j.cub.2021.01.074

Liang X, Shang S, Dong Q, Wang B et al. 2018a – Transcriptomic analysis reveals candidate genes regulating development and host interactions of *Colletotrichum fructicola*. BMC Genomics 19(1), 557. Doi 10.1186/s12864-018-4934-0

Liang X, Wang B, Dong Q, Li L et al. 2018b – Pathogenic adaptations of *Colletotrichum* fungi revealed by genome wide gene family evolutionary analyses. Plos One 13(4), e0196303. Doi 10.1371/journal.pone.0196303

Lima NB, Batista MVdA, De Morais MA, Barbosa MA et al. 2013 – Five *Colletotrichum* species are responsible for mango anthracnose in northeastern Brazil. Fungal Diversity 61(1), 75–88. Doi 10.1007/s13225-013-0237-6

Liu F, Ma ZY, Hou LW, Diao YZ et al. 2022 – Updating species diversity of *Colletotrichum,* with a phylogenomic overview. Studies in Mycology, 101, 1–56. Doi 10.3114/sim.2022.101.01

Liu J-K, Hyde KD, Jeewon R, Phillips AJL et al. 2017 – Ranking higher taxa using divergence times: a case study in Dothideomycetes. Fungal Diversity 84(1), 75–99. Doi 10.1007/s13225-017-0385-1

Longo SK, Guo MG, Ji AL, Khavari PA. 2021 – Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. Nature Reviews Genetics 22(10), 627–644. Doi 10.1038/s41576-021-00370-8

Lowe RGT, Howlett BJ. 2012 – Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi. Plos Pathogens 8(3), e1002515. Doi 10.1371/journal.ppat.1002515

Maharachchikumbura SSN, Chen Y, Ariyawansa HA, Hyde KD et al. 2021 – Integrative approaches for species delimitation in Ascomycota. Fungal Diversity 109(1), 155–179. Doi 10.1007/s13225-021-00486-6

Martinez D, Challacombe J, Morgenstern I, Hibbett D et al. 2009 – Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. Proceedings of the National Academy of Sciences 106(6), 1954–1959. Doi 10.1073/pnas.0809575106

Medema MH, Kottmann R, Yilmaz P, Cummings M et al. 2015 – Minimum information about a biosynthetic gene cluster. Nature Chemical Biology 11(9), 625–631. Doi 10.1038/nchembio.1890

Mena E, Garaycochea S, Stewart S, Montesano M, Ponce De León I. 2022 – Comparative genomics of plant pathogenic *Diaporthe* species and transcriptomics of *Diaporthe caulivora* during host infection reveal insights into pathogenic strategies of the genus. BMC Genomics 23(1), 175. Doi 10.1186/s12864-022-08413-y

Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2020 – CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics 36(22–23), 5516–5518. Doi 10.1093/bioinformatics/btaa1022

Mistry J, Chuguransky S, Williams L, Qureshi M et al. 2020 – Pfam: The protein families database in 2021. Nucleic Acids Research 49(D1), D412–D419. Doi 10.1093/nar/gkaa913

Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K. 2019 – Transposable elements contribute to fungal genes and impact fungal lifestyle. Scientific Reports 9(1), 4307. Doi 10.1038/s41598-019-40965-0

Newman DJ, Cragg GM. 2012 – Natural products as sources of new drugs over the 30 years from 1981 to 2010. Journal of Natural Products 75(3), 311–335. Doi 10.1021/np200906s

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2014 – IQ-TREE: a fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. Molecular Biology and Evolution 32(1), 268–274. Doi 10.1093/molbev/msu300

Nguyen LN, Bormann J, Le GTT, Stärkel C et al. 2011a – Autophagy-related lipase *FgATG15* of *Fusarium graminearum* is important for lipid turnover and plant infection. Fungal Genetics and Biology 48(3), 217–224. Doi 10.1016/j.fgb.2010.11.004

Nguyen LN, Gacser A, Nosanchuk JD. 2011b – Secreted lipases supply fatty acids for yeast growth in the absence of de novo fatty acid synthesis. Virulence 2(6), 538–541. Doi 10.4161/viru.2.6.18244

Nielsen JC, Grijseels S, Prigent S, Ji B et al. 2017 – Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species. Nature Microbiology 2(6), 17044. Doi 10.1038/nmicrobiol.2017.44

Nirlane da Costa Souza P, Luiza Bim Grigoletto T, Alberto Beraldo de Moraes L, Abreu LM et al. 2016 – Production and chemical characterization of pigments in filamentous fungi. Microbiology 162(1), 12–22. Doi 10.1099/mic.0.000168

O'Connell RJ, Thon MR, Hacquard S, Amyotte SG et al. 2012 – Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nature Genetics 44(9), 1060–1065. Doi 10.1038/ng.2372

Oliveira L, Macedo M, Rodrigues J, Lima E et al. 2022 – Plant metabolite 5-pentadecyl resorcinol is produced by the Amazonian fungus *Penicillium sclerotiorum* LM 5679. Brazilian Journal of Biology 82, e241863. Doi 10.1590/1519-6984.241863

Olivieri F, Eugenia Zanetti M, 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(1), 63–72. Doi 10.1023/A:1013920929965

Oo MM, Yoon HY, Jang HA, Oh SK. 2018 – Identification and characterization of *Colletotrichum* species associated with bitter rot disease of apple in South Korea. The Plant Pathology Journal 34(6), 480–489. Doi 10.5423/PPJ.FT.10.2018.0201

Park M, Do E, Jung WH. 2013 – Lipolytic enzymes involved in the virulence of human pathogenic fungi. Mycobiology 41(2), 67–72. Doi 10.5941/MYCO.2013.41.2.67

Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. 2015 – Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. Frontiers in Microbiology 6, 1278. Doi 10.3389/fmicb.2015.01278

Plett Jonathan M, Kemppainen M, Kale Shiv D, Kohler A et al. 2011 – A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. Current Biology 21(14), 1197–1203. Doi 10.1016/j.cub.2011.05.033

Prasanth CN, Viswanathan R, Malathi P, Sundar AR. 2019 – Comparative transcriptome analysis of candidate secretory effector proteins from *Colletotrichum falcatum* infecting sugarcane. Agri Gene 13, 100089. Doi 10.1016/j.aggene.2019.100089

R Core Team. 2022 – R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raffaele S, Kamoun S. 2012 – Genome evolution in filamentous plant pathogens: why bigger can be better. Nature Reviews Microbiology 10 (6), 417–430. Doi 10.1038/nrmicro2790

Rawlings ND, Barrett AJ, Thomas PD, Huang X et al. 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(D1), D624–D632. Doi 10.1093/nar/gkx1134

Reddy VS, Shlykov MA, Castillo R, Sun EI, Saier Jr MH. 2012 – The major facilitator superfamily (MFS) revisited. The FEBS Journal 279(11), 2022–2035. Doi 10.1111/j.1742-4658.2012.08588.x

Rovenich H, Boshoven JC, Thomma BPHJ. 2014 – Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. Current Opinion in Plant Biology 20, 96–103. Doi 10.1016/j.pbi.2014.05.001

Schena L, Mosca S, Cacciola SO, Faedda R et al. 2014 – Species of the *Colletotrichum gloeosporioides* and *C. boninense* complexes associated with olive anthracnose. Plant Pathology 63 (2), 437–446. Doi 10.1111/ppa.12110

Schenk JJ. 2016 – Consequences of secondary calibrations on divergence time estimates. Plos One 11 (1), e0148228. Doi 10.1371/journal.pone.0148228

Schmitt EK, Hoff B, Kück U. 2004 – Regulation of cephalosporin biosynthesis. In: Brakhage AA (ed) Molecular biotechnolgy of fungal beta-lactam antibiotics and related peptide synthetases. Advances in Biochemical Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg 88, 1–43. Doi 10.1007/b99256

Seidl MF, Faino L, Shi-Kunne X, van den Berg GCM et al. 2015 – The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic relatives. Molecular Plant-Microbe Interactions 28(3), 362–373. Doi 10.1094/MPMI-06-14-0173-R

Shah F, Rineau F, Canbäck B, Johansson T, Tunlid A. 2013 – The molecular components of the extracellular protein-degradation pathways of the ectomycorrhizal fungus *Paxillus involutus*. New Phytologist 200(3), 875–887. Doi 10.1111/nph.12425

Shao F, Han M, Peng Z. 2019 – Evolution and diversity of transposable elements in fish genomes. Scientific Reports 9(1), 15399. Doi 10.1038/s41598-019-51888-1

Shen X-X, Steenwyk JL, LaBella AL, Opulente DA et al. 2020 – Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. Science Advances 6(45), eabd0079. Doi 10.1126/sciadv.abd0079

Shivaprakash MR, Appannanavar SB, Dhaliwal M, Gupta A et al. 2011 – *Colletotrichum truncatum*: an unusual pathogen causing mycotic keratitis and endophthalmitis. Journal of Clinical Microbiology 49(8), 2894–2898. Doi 10.1128/JCM.00151-11

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015 – BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19), 3210–3212. Doi 10.1093/bioinformatics/btv351

Smit A, Hubley R, Green P. 2015 – RepeatMasker Open-4.0. 2013–2015. http://www.repeatmasker.org

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021 – The Dfam community resource of transposable element families, sequence models, and genome annotations. Mobile DNA 12(1), 2. Doi 10.1186/s13100-020-00230-y

The UniProt Consortium TU. 2018 – UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research 47 (D1), D506-D515. Doi 10.1093/nar/gky1049

Talhinhas P, Baroncelli R. 2021 – *Colletotrichum* species and complexes: geographic distribution, host range and conservation status. Fungal Diversity 110(1), 109–198. Doi 10.1007/s13225-021-00491-9

Tanaka A, Tapper BA, Popay A, Parker EJ, Scott B. 2005 – A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory. Molecular Microbiology 57(4), 1036–1050. Doi 10.1111/j.1365-2958.2005.04747.x

Tao G, Liu Z-Y, Liu F, Gao Y-H, Cai L. 2013 – Endophytic *Colletotrichum* species from *Bletilla ochracea* (Orchidaceae), with descriptions of seven new speices. Fungal Diversity 61(1), 139–164. Doi 10.1007/s13225-013-0254-5

Tsushima A, Gan P, Kumakura N, Narusaka M et al. 2019 – Genomic plasticity mediated by transposable elements in the plant pathogenic fungus *Colletotrichum higginsianum*. Genome Biology and Evolution 11(5), 1487–1500. Doi 10.1093/gbe/evz087

van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. 2018 – The third revolution in sequencing technology. Trends in Genetics 34(9), 666–681. Doi 10.1016/j.tig.2018.05.008

van Tuinen M, Hadly EA. 2004 – Error in estimation of rate and time inferred from the early amniote fossil record and avian molecular clocks. Journal of Molecular Evolution 59(2), 267–276. Doi 10.1007/s00239-004-2624-9

van Tuinen M, Torres CR. 2015 – Potential for bias and low precision in molecular divergence time estimation of the Canopy of Life: an example from aquatic bird families. Frontiers in Genetics 6: 203. Doi 10.3389/fgene.2015.00203

Vela-Corcía D, Aditya Srivastava D, Dafa-Berger A, Rotem N et al. 2019 – MFS transporter from *Botrytis cinerea* provides tolerance to glucosinolate-breakdown products and is required for pathogenicity. Nature Communications 10(1), 2886. Doi 10.1038/s41467-019-10860-3

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(3), 364–375. Doi 10.1111/j.1365-313X.2005.02377.x

Voloch CM, Schrago CG. 2012 – Impact of the partitioning scheme on divergence times inferred from mammalian genomic data sets. Evolutionary Bioinformatics 8, EBO.S9627. Doi 10.4137/EBO.S9627

Wang W, de Silva DD, Moslemi A, Edwards J et al. 2021 – *Colletotrichum* species causing anthracnose of citrus in Australia. Journal of Fungi 7(1), 47. Doi 10.3390/jof7010047

Wang Z, Gerstein M, Snyder M. 2009 – RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10(1), 57–63. Doi 10.1038/nrg2484

Yike I. 2011 – Fungal proteases and their pathophysiological effects. Mycopathologia 171(5), 299–323. Doi 10.1007/s11046-010-9386-2

Yu G. 2020 – Using ggtree to visualize data on tree-like structures. Current Protocols in Bioinformatics 69(1), e96. Doi 10.1002/cpbi.96

Zarnowski R, Lewicka T, Pietr SJ. 2000 – Production and secretion of 5-n-Alkylresorcinols by *Fusarium culmorum*. Zeitschrift für Naturforschung C 55(C), 846–848. Doi 10.1515/znc-2000-9-1030

Zhang L, Song L, Xu X, Zou X et al. 2020 – Characterization and fungicide sensitivity of *Colletotrichum* species causing strawberry anthracnose in eastern China. Plant Disease 104(7), 1960–1968. Doi 10.1094/PDIS-10-19-2241-RE

Zhao R-L, Zhou J-L, Chen J, Margaritescu S et al. 2016 – Towards standardizing taxonomic ranks using divergence times – a case study for reconstruction of the *Agaricus* taxonomic system. Fungal Diversity 78(1), 239–292. Doi 10.1007/s13225-016-0357-x

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(1), 274. Doi 10.1186/1471-2164-14-274

Zuckerkandl E. 1962 – Molecular disease, evolution, and genetic heterogeneity. Horizons in Biochemistry, 189–225.

Supplementary Table 1 A total of 270 RNA-seq data used in genome annotation.

Run	Organism	Size (Mb)	BioProject	BioSample	All isoforms			Longest isoforms		
					Count	**AverageLength**	**MaxLength**	**Count**	**AverageLength**	**MaxLength**
SRR12329585	*C. camelliae*	8,276	PRJNA628751	SAMN14751196	33,429	1,165	15,128	30,498	1,046	15,128
SRR12329586	*C. camelliae*	9,353	PRJNA628751	SAMN14751196	34,435	1,219	15,297	31,562	1,086	15,297
SRR12329587	*C. camelliae*	9,704	PRJNA628751	SAMN14751196	35,887	1,427	18,750	31,420	1,174	18,750
SRR12329588	*C. camelliae*	10,916	PRJNA628751	SAMN14751196	182,147	791	13,049	154,847	693	13,049
SRR12329589	*C. camelliae*	10,550	PRJNA628751	SAMN14751196	213,369	757	10,789	178,544	656	10,789
SRR12329590	*C. camelliae*	10,513	PRJNA628751	SAMN14751196	203,474	715	11,123	175,697	627	11,123
SRR12329591	*C. camelliae*	9,481	PRJNA628751	SAMN14751196	40,137	1,816	17,018	34,334	1,512	17,018
SRR12329592	*C. camelliae*	10,268	PRJNA628751	SAMN14751196	42,366	1,745	17,041	36,356	1,438	17,041
SRR12329593	*C. camelliae*	9,866	PRJNA628751	SAMN14751196	37,081	1,369	18,756	32,791	1,135	18,756
SRR12329594	*C. camelliae*	9,725	PRJNA628751	SAMN14751196	42,053	1,715	18,861	36,282	1,433	18,861
SRR12329595	*C. camelliae*	6,588	PRJNA628751	SAMN14751196	96,969	898	15,636	88,376	793	15,636
SRR12329596	*C. camelliae*	6,693	PRJNA628751	SAMN14751196	98,102	917	16,040	89,627	816	16,040
SRR12329597	*C. camelliae*	7,676	PRJNA628751	SAMN14751196	93,006	937	15,491	85,385	823	15,491
SRR12329598	*C. camelliae*	8,850	PRJNA628751	SAMN14751196	34,038	1,163	15,196	31,137	1,045	15,196
SRR12329599	*C. camelliae*	9,766	PRJNA628751	SAMN14751196	42,046	1,237	18,934	36,811	1,012	18,934
SRR13053733	*C. echinochloae*	3,088	PRJNA672254	SAMN16551125	26,088	425	4,432	25,355	429	4,432
SRR13050946	*C. eleusines*	2,476	PRJNA671855	SAMN16552423	20,390	475	5,887	20,135	478	5,887
SRR1765657	*C. falcatum*	9,997	PRJNA272832	SAMN03282447	46,362	1,427	19,809	37,196	1,119	19,809
SRR6041844	*C. fructicola*	1,495	PRJNA407475	SAMN07655939	25,610	1,147	13,197	23,383	1,034	13,197
SRR6041845	*C. fructicola*	1,804	PRJNA407475	SAMN07655938	30,087	1,049	12,390	27,269	939	12,390
SRR6041846	*C. fructicola*	2,114	PRJNA407475	SAMN07655937	31,359	1,051	11,191	28,085	947	11,191
SRR6041847	*C. fructicola*	1,237	PRJNA407475	SAMN07655936	25,455	1,189	9,724	24,076	1,123	9,724
SRR6041848	*C. fructicola*	1,517	PRJNA407475	SAMN07655935	25,836	1,232	9,635	24,330	1,159	9,635
SRR6041849	*C. fructicola*	1,437	PRJNA407475	SAMN07655934	27,177	1,196	11,099	25,549	1,120	11,099
SRR6041850	*C. fructicola*	8,761	PRJNA407475	SAMN07655933	180,490	787	13,831	137,995	596	13,831
SRR6041851	*C. fructicola*	10,067	PRJNA407475	SAMN07655932	232,131	727	15,776	184,347	563	15,776
SRR6041852	*C. fructicola*	9,337	PRJNA407475	SAMN07655943	204,866	766	15,649	157,815	577	15,649
SRR6041853	*C. fructicola*	1,374	PRJNA407475	SAMN07655942	33,670	930	8,218	31,655	882	8,218
SRR6041854	*C. fructicola*	1,281	PRJNA407475	SAMN07655941	32,530	979	8,652	30,359	926	8,652
SRR6041855	*C. fructicola*	1,685	PRJNA407475	SAMN07655940	37,063	936	9,466	34,483	881	9,466
SRR11216269	*C. fructicola*	897	PRJNA476648	SAMN09448926	18,535	656	8,093	18,461	653	8,093
SRR11216271	*C. fructicola*	987	PRJNA476648	SAMN09448926	18,335	676	8,126	18,282	674	8,126
SRR11216272	*C. fructicola*	1,228	PRJNA476648	SAMN09448926	20,033	709	8,102	19,956	706	8,102
SRR11216273	*C. fructicola*	1,097	PRJNA476648	SAMN09448926	22,213	680	7,763	22,131	677	7,763
SRR11216274	*C. fructicola*	967	PRJNA476648	SAMN09448926	20,358	649	7,671	20,312	648	7,671
SRR11216275	*C. fructicola*	1,074	PRJNA476648	SAMN09448926	22,065	652	7,738	22,012	651	7,738

Supplementary Table 1 Continued.

Run	Organism	Size (Mb)	BioProject	BioSample	All isoforms			Longest isoforms		
					Count	**AverageLength**	**MaxLength**	**Count**	**AverageLength**	**MaxLength**
SRR11216276	*C. fructicola*	1,171	PRJNA476648	SAMN09448926	44,644	442	4,854	42,110	435	4,854
SRR11216277	*C. fructicola*	1,217	PRJNA476648	SAMN09448926	46,616	442	3,656	44,071	435	3,656
SRR11216278	*C. fructicola*	1,465	PRJNA476648	SAMN09448926	43,410	441	3,250	41,124	434	3,250
SRR11216279	*C. fructicola*	3,170	PRJNA476648	SAMN09448926	61,120	500	5,789	56,948	491	5,789
SRR11216280	*C. fructicola*	3,897	PRJNA476648	SAMN09448926	64,567	522	8,544	59,259	510	8,544
SRR11216282	*C. fructicola*	3,687	PRJNA476648	SAMN09448926	61,665	500	5,408	56,953	489	5,408
SRR11216283	*C. fructicola*	3,313	PRJNA476648	SAMN09448926	59,306	502	5,146	55,337	491	5,146
SRR11216284	*C. fructicola*	3,444	PRJNA476648	SAMN09448926	730	353	1,349	716	353	1,349
SRR11216285	*C. fructicola*	3,227	PRJNA476648	SAMN09448926	62,759	483	6,343	58,279	474	6,343
SRR11216286	*C. fructicola*	752	PRJNA476648	SAMN09531621	16,829	623	7,758	16,792	621	7,758
SRR11216287	*C. fructicola*	1,102	PRJNA476648	SAMN09531621	19,441	628	7,931	19,399	627	7,931
SRR11216288	*C. fructicola*	1,444	PRJNA476648	SAMN09531621	21,740	705	8,093	21,675	703	8,093
SRR11216289	*C. fructicola*	1,339	PRJNA476648	SAMN09531621	20,476	733	7,735	20,409	730	7,735
SRR11216290	*C. fructicola*	1,053	PRJNA476648	SAMN09531621	21,930	710	7,667	21,877	708	7,667
SRR11216291	*C. fructicola*	974	PRJNA476648	SAMN09531621	21,117	663	7,694	21,073	661	7,694
SRR11216293	*C. fructicola*	1,249	PRJNA476648	SAMN09531621	43,891	448	5,142	41,297	439	5,142
SRR11216294	*C. fructicola*	1,154	PRJNA476648	SAMN09531621	46,534	439	5,185	44,033	432	5,185
SRR11216295	*C. fructicola*	2,114	PRJNA476648	SAMN09531621	54,058	468	5,430	50,496	457	5,430
SRR11216296	*C. fructicola*	3,521	PRJNA476648	SAMN09531621	64,238	491	5,306	58,969	480	5,306
SRR11216297	*C. fructicola*	3,322	PRJNA476648	SAMN09531621	57,826	494	4,958	53,388	484	4,958
SRR11216298	*C. fructicola*	3,129	PRJNA476648	SAMN09531621	50,280	467	4,591	46,741	457	4,591
SRR11216299	*C. fructicola*	3,877	PRJNA476648	SAMN09531621	68,910	508	5,255	62,505	494	5,255
SRR11216300	*C. fructicola*	3,279	PRJNA476648	SAMN09531621	64,306	476	5,459	58,771	465	5,459
SRR11216301	*C. fructicola*	3,523	PRJNA476648	SAMN09531621	53,264	460	3,240	48,745	451	3,240
SRR9118233	*C. fructicola*	5,197	PRJNA528175	SAMN11175340	42,529	1,276	15,359	36,594	1,144	15,359
SRR9118234	*C. fructicola*	4,373	PRJNA528175	SAMN11175341	42,363	1,282	30,366	36,391	1,135	30,366
SRR9118235	*C. fructicola*	4,580	PRJNA528175	SAMN11175260	42,958	1,234	17,138	36,841	1,115	17,138
SRR9118236	*C. fructicola*	3,919	PRJNA528175	SAMN11175275	42,063	1,233	25,694	36,094	1,107	25,694
SRR9118237	*C. fructicola*	4,668	PRJNA528175	SAMN11175342	41,661	1,325	30,777	35,729	1,177	30,777
SRR9118238	*C. fructicola*	4,089	PRJNA528175	SAMN11175343	50,050	1,110	17,100	43,028	989	17,100
SRR12149179	*C. fructicola*	13,576	PRJNA644240	SAMN15455115	183,485	686	12,964	155,847	606	12,964
SRR12149180	*C. fructicola*	8,177	PRJNA644240	SAMN15455114	71,382	1,029	17,948	63,172	873	17,948
SRR12149181	*C. fructicola*	8,063	PRJNA644240	SAMN15455113	41,037	1,328	19,721	36,287	1,107	19,721
SRR12149182	*C. fructicola*	8,375	PRJNA644240	SAMN15455112	44,660	1,296	21,698	39,724	1,073	21,698
SRR12149183	*C. fructicola*	10,169	PRJNA644240	SAMN15455111	202,947	724	15,008	161,562	637	15,008
SRR12149184	*C. fructicola*	11,600	PRJNA644240	SAMN15455110	181,199	751	13,375	147,353	663	13,375
SRR12149185	*C. fructicola*	10,972	PRJNA644240	SAMN15455109	193,340	712	15,289	154,470	621	15,289

Supplementary Table 1 Continued.

Run	Organism	Size (Mb)	BioProject	BioSample	All isoforms			Longest isoforms		
					Count	**AverageLength**	**MaxLength**	**Count**	**AverageLength**	**MaxLength**
SRR12149186	*C. fructicola*	10,402	PRJNA644240	SAMN15455108	55,820	1,048	30,542	50,434	887	30,542
SRR12149187	*C. fructicola*	15,340	PRJNA644240	SAMN15455117	245,283	624	15,176	206,048	558	15,176
SRR12149188	*C. fructicola*	11,278	PRJNA644240	SAMN15455116	222,005	607	11,669	188,987	551	11,669
SRR12149189	*C. fructicola*	8,883	PRJNA644240	SAMN15455107	44,617	1,202	28,239	39,379	994	28,239
SRR12149190	*C. fructicola*	7,940	PRJNA644240	SAMN15455106	31,586	1,445	26,222	27,489	1,208	26,222
SRR12488738	*C. fructicola*	6,328	PRJNA658390	SAMN15870153	171,004	821	16,197	125,440	602	16,197
SRR12488739	*C. fructicola*	6,417	PRJNA658390	SAMN15870152	166,616	837	15,803	120,376	610	15,803
SRR11216259	*C. fructicola*	666	PRJNA476648	SAMN02981487	15,896	588	8,347	15,838	586	8,347
SRR11216260	*C. fructicola*	1,363	PRJNA476648	SAMN02981487	49,882	448	4,514	46,716	440	4,514
SRR11216261	*C. fructicola*	822	PRJNA476648	SAMN02981487	32,000	403	3,234	30,951	398	3,234
SRR11216262	*C. fructicola*	1,531	PRJNA476648	SAMN02981487	47,993	460	4,887	46,797	457	4,887
SRR11216263	*C. fructicola*	921	PRJNA476648	SAMN02981487	40,376	436	6,011	38,671	431	6,011
SRR11216264	*C. fructicola*	945	PRJNA476648	SAMN02981487	40,521	440	7,401	38,791	434	7,401
SRR11216265	*C. fructicola*	882	PRJNA476648	SAMN02981487	40,713	438	5,038	39,024	432	5,038
SRR11216266	*C. fructicola*	955	PRJNA476648	SAMN02981487	41,379	430	4,507	39,695	423	4,507
SRR11216267	*C. fructicola*	886	PRJNA476648	SAMN02981487	42,675	417	3,618	41,070	411	3,618
SRR11216268	*C. fructicola*	712	PRJNA476648	SAMN02981487	38,056	411	3,493	36,763	405	3,493
SRR11216270	*C. fructicola*	3,310	PRJNA476648	SAMN02981487	64,202	510	5,503	57,801	496	5,503
SRR11216281	*C. fructicola*	3,722	PRJNA476648	SAMN02981487	55,135	485	4,378	50,010	474	4,378
SRR11216292	*C. fructicola*	2,844	PRJNA476648	SAMN02981487	53,482	483	3,537	49,087	472	3,537
SRR11216302	*C. fructicola*	1,380	PRJNA476648	SAMN02981487	25,998	705	7,767	25,832	701	7,767
SRR11216303	*C. fructicola*	3,457	PRJNA476648	SAMN02981487	62,306	491	7,805	56,986	480	7,805
SRR11216304	*C. fructicola*	1,079	PRJNA476648	SAMN02981487	21,085	670	8,115	21,006	668	8,115
SRR11216305	*C. fructicola*	1,230	PRJNA476648	SAMN02981487	21,256	705	7,768	21,112	702	7,768
SRR11216306	*C. fructicola*	1,144	PRJNA476648	SAMN02981487	22,720	691	7,695	22,567	688	7,695
SRR11216307	*C. fructicola*	759	PRJNA476648	SAMN02981487	20,942	615	7,669	20,885	614	7,669
SRR11216308	*C. fructicola*	994	PRJNA476648	SAMN02981487	21,808	647	8,585	21,705	645	8,585
SRR11216309	*C. fructicola*	2,164	PRJNA476648	SAMN02981487	57,800	475	4,534	53,160	467	4,534
SRR11216310	*C. fructicola*	2,489	PRJNA476648	SAMN02981487	59,715	486	5,461	54,978	478	5,461
SRR11216311	*C. fructicola*	2,690	PRJNA476648	SAMN02981487	60,383	487	5,194	55,094	479	5,194
SRR11216312	*C. fructicola*	879	PRJNA476648	SAMN02981487	19,638	593	8,471	19,552	590	8,471
SRR11216313	*C. fructicola*	888	PRJNA476648	SAMN02981487	17,576	609	8,304	17,492	605	8,304
SRR11216314	*C. fructicola*	3,204	PRJNA476648	SAMN02981487	67,424	521	5,816	60,689	505	5,816
SRR11216315	*C. fructicola*	3,258	PRJNA476648	SAMN02981487	68,143	513	5,466	61,310	498	5,466
SRR609271	*C. gloeosporioides*	408	PRJNA178310	SAMN01796016	22,230	746	7,918	21,765	732	7,918
SRR609272	*C. gloeosporioides*	671	PRJNA178310	SAMN01796017	24,627	873	8,274	23,714	845	8,274
SRR1171641	*C. gloeosporioides*	4,351	PRJNA238599	SAMN02645567	34,011	1,256	12,565	31,861	1,167	12,565

Supplementary Table 1 Continued.

Run	Organism	Size (Mb)	BioProject	BioSample	All isoforms			Longest isoforms		
					Count	AverageLength	MaxLength	Count	AverageLength	MaxLength
SRR2844617	*C. gloeosporioides*	5,821	PRJNA300348	SAMN04219543	33,857	1,454	13,835	29,935	1,235	13,835
SRR5194993	*C. gloeosporioides*	4,938	PRJNA362519	SAMN06241151	35,788	1,685	23,454	30,990	1,465	23,454
SRR5194994	*C. gloeosporioides*	5,699	PRJNA362519	SAMN06241150	36,327	1,622	15,666	31,829	1,451	15,666
SRR5194995	*C. gloeosporioides*	5,127	PRJNA362519	SAMN06241149	35,187	1,662	25,276	30,700	1,460	25,276
SRR5723440	*C. gloeosporioides*	7,863	PRJNA391239	SAMN07259823	40,732	1,587	17,183	36,277	1,383	17,183
SRR5723441	*C. gloeosporioides*	6,879	PRJNA391239	SAMN07259824	46,563	1,319	15,463	41,853	1,193	15,463
SRR5723442	*C. gloeosporioides*	6,240	PRJNA391239	SAMN07259821	65,510	926	15,558	60,710	845	15,558
SRR5723443	*C. gloeosporioides*	8,160	PRJNA391239	SAMN07259822	49,178	1,440	17,765	43,879	1,236	17,765
SRR5723444	*C. gloeosporioides*	6,610	PRJNA391239	SAMN07259819	58,830	1,025	16,465	53,971	916	16,465
SRR5723445	*C. gloeosporioides*	6,508	PRJNA391239	SAMN07259820	51,931	1,104	13,800	47,345	993	13,800
SRR7767987	*C. gloeosporioides*	6,633	PRJNA488301	SAMN09929771	36,589	1,498	15,959	30,907	1,284	15,959
SRR7767988	*C. gloeosporioides*	6,417	PRJNA488301	SAMN09929770	37,548	1,254	20,298	32,395	1,111	20,298
SRR7767989	*C. gloeosporioides*	6,515	PRJNA488301	SAMN09929773	44,649	1,162	16,827	36,731	943	16,827
SRR7767990	*C. gloeosporioides*	6,574	PRJNA488301	SAMN09929772	45,274	1,459	18,470	35,970	1,124	18,470
SRR7767991	*C. gloeosporioides*	6,448	PRJNA488301	SAMN09929767	37,428	1,391	16,877	32,454	1,246	16,877
SRR7767992	*C. gloeosporioides*	6,171	PRJNA488301	SAMN09929766	36,207	1,450	17,468	31,620	1,287	17,468
SRR7767993	*C. gloeosporioides*	6,531	PRJNA488301	SAMN09929769	34,914	1,281	20,456	30,567	1,135	20,456
SRR7767994	*C. gloeosporioides*	6,525	PRJNA488301	SAMN09929768	38,797	1,432	18,130	33,225	1,259	18,130
SRR7767995	*C. gloeosporioides*	6,843	PRJNA488301	SAMN09929774	47,991	1,251	16,820	40,548	1,018	16,820
SRR9936824	*C. gloeosporioides*	2,166	PRJNA556847	SAMN12369112	49,682	1,516	16,764	37,990	1,028	16,764
SRR9936826	*C. gloeosporioides*	2,119	PRJNA556847	SAMN12369111	40,083	959	12,079	36,108	857	12,079
SRR9936828	*C. gloeosporioides*	2,122	PRJNA556847	SAMN12369114	39,615	1,081	11,279	35,046	960	11,279
SRR9936829	*C. gloeosporioides*	2,172	PRJNA556847	SAMN12369115	44,244	1,203	15,297	37,245	1,016	15,297
SRR9936831	*C. gloeosporioides*	2,152	PRJNA556847	SAMN12369113	42,264	1,005	10,699	37,721	879	10,699
SRR9936832	*C. gloeosporioides*	2,259	PRJNA556847	SAMN12369118	35,850	1,145	15,088	32,743	1,023	15,088
SRR9936833	*C. gloeosporioides*	2,215	PRJNA556847	SAMN12369119	37,705	1,080	12,999	33,799	964	12,999
SRR9936834	*C. gloeosporioides*	2,139	PRJNA556847	SAMN12369116	38,078	1,102	12,060	33,405	960	12,060
SRR9936835	*C. gloeosporioides*	2,165	PRJNA556847	SAMN12369117	42,091	1,146	9,520	36,125	960	9,520
SRR10662693	*C. gloeosporioides*	3,539	PRJNA594803	SAMN13531060	32,997	1,023	14,116	30,847	923	14,116
SRR10662694	*C. gloeosporioides*	3,072	PRJNA594803	SAMN13531059	32,154	947	13,064	30,166	855	13,064
SRR10662696	*C. gloeosporioides*	3,534	PRJNA594803	SAMN13531057	28,335	1,024	29,126	26,573	921	29,126
SRR10662697	*C. gloeosporioides*	3,381	PRJNA594803	SAMN13531056	25,684	1,223	15,304	23,787	1,123	15,304
SRR10662698	*C. gloeosporioides*	3,703	PRJNA594803	SAMN13531055	32,053	1,138	16,619	30,081	1,037	16,619
SRR10662699	*C. gloeosporioides*	3,608	PRJNA594803	SAMN13531054	25,783	1,204	26,368	23,998	1,105	26,368
SRR10662700	*C. gloeosporioides*	3,034	PRJNA594803	SAMN13531053	28,666	1,071	29,014	26,732	970	29,014
SRR10662701	*C. gloeosporioides*	3,278	PRJNA594803	SAMN13531062	23,532	1,315	30,198	21,893	1,201	30,198
SRR10662702	*C. gloeosporioides*	3,204	PRJNA594803	SAMN13531061	24,494	1,182	18,759	23,143	1,090	18,759

Supplementary Table 1 Continued.

Run	Organism	Size (Mb)	BioProject	BioSample	All isoforms			Longest isoforms		
					Count	**AverageLength**	**MaxLength**	**Count**	**AverageLength**	**MaxLength**
SRR10662703	*C. gloeosporioides*	3,148	PRJNA594803	SAMN13531052	20,749	1,605	16,194	18,891	1,407	16,194
SRR10662704	*C. gloeosporioides*	3,661	PRJNA594803	SAMN13531051	23,090	1,834	18,663	20,446	1,492	18,663
SRR391823	*C. graminicola*	2,326	PRJNA151285	SAMN00767940	75,198	792	12,274	62,839	671	12,274
SRR391824	*C. graminicola*	2,485	PRJNA151285	SAMN00767941	81,219	806	14,234	66,845	671	14,234
SRR391825	*C. graminicola*	2,508	PRJNA151285	SAMN00767942	95,197	826	13,046	79,513	717	13,046
SRR391826	*C. graminicola*	2,472	PRJNA151285	SAMN00767943	74,086	901	13,299	65,369	827	13,299
SRR391827	*C. graminicola*	852	PRJNA151285	SAMN00767944	47,447	748	10,941	41,295	669	10,941
SRR391828	*C. graminicola*	878	PRJNA151285	SAMN00767945	48,419	752	10,603	42,036	673	10,603
SRR391829	*C. graminicola*	703	PRJNA151285	SAMN00767946	52,342	672	9,796	46,177	607	9,796
SRR391830	*C. graminicola*	727	PRJNA151285	SAMN00767947	50,505	676	8,032	47,981	651	8,032
SRR391831	*C. graminicola*	761	PRJNA151285	SAMN00767948	46,597	731	10,387	40,603	653	10,387
SRR391832	*C. graminicola*	785	PRJNA151285	SAMN00767949	47,829	739	8,434	41,494	659	8,434
SRR391833	*C. graminicola*	755	PRJNA151285	SAMN00767950	60,327	660	17,417	55,013	609	17,417
SRR391834	*C. graminicola*	825	PRJNA151285	SAMN00767951	55,336	702	7,939	51,944	668	7,939
SRR391835	*C. graminicola*	656	PRJNA151285	SAMN00767952	42,544	697	8,796	38,256	638	8,796
SRR391836	*C. graminicola*	678	PRJNA151285	SAMN00767953	43,564	710	10,097	38,862	645	10,097
SRR391837	*C. graminicola*	773	PRJNA151285	SAMN00767954	58,247	661	8,168	53,023	613	8,168
SRR391838	*C. graminicola*	695	PRJNA151285	SAMN00767955	42,433	718	8,233	40,917	698	8,233
SRR391839	*C. graminicola*	510	PRJNA151285	SAMN00767956	40,554	667	8,439	36,575	612	8,439
SRR391840	*C. graminicola*	494	PRJNA151285	SAMN00767957	40,237	662	8,285	36,366	605	8,285
SRR391841	*C. graminicola*	590	PRJNA151285	SAMN00767958	43,900	680	10,265	39,049	618	10,265
SRR391842	*C. graminicola*	591	PRJNA151285	SAMN00767959	43,424	683	10,466	38,696	620	10,466
SRR391843	*C. graminicola*	628	PRJNA151285	SAMN00767960	43,296	684	10,494	38,986	625	10,494
SRR391844	*C. graminicola*	566	PRJNA151285	SAMN00767961	41,619	668	7,295	37,695	611	7,295
SRR1237827	*C. graminicola*	4,032	PRJNA244494	SAMN02726835	30,297	1,213	9,368	27,981	1,125	9,368
SRR2185181	*C. graminicola*	51	PRJNA293991	SAMN04012971	5,126	428	3,169	5,102	429	3,169
SRR2185182	*C. graminicola*	64	PRJNA293991	SAMN04012972	6,684	430	3,581	6,659	430	3,581
SRR2185183	*C. graminicola*	36	PRJNA293991	SAMN04012973	3,514	381	3,166	3,498	381	3,166
SRR2185184	*C. graminicola*	60	PRJNA293991	SAMN04012974	5,683	386	3,171	5,629	385	3,171
SRR2185185	*C. graminicola*	18	PRJNA293991	SAMN04012975	1,779	359	2,581	1,763	360	2,581
SRR2185186	*C. graminicola*	16	PRJNA293991	SAMN04012976	1,353	353	1,949	1,336	354	1,949
SRR2185187	*C. graminicola*	27	PRJNA293991	SAMN04012977	3,537	386	3,160	3,517	387	3,160
SRR2185188	*C. graminicola*	101	PRJNA293991	SAMN04012978	10,985	485	5,697	10,965	485	5,697
SRR2185189	*C. graminicola*	111	PRJNA293991	SAMN04012979	11,978	491	5,812	11,938	491	5,812
SRR2185190	*C. graminicola*	21	PRJNA293991	SAMN04012980	2,404	366	2,530	2,385	367	2,530
SRR2185191	*C. graminicola*	32	PRJNA293991	SAMN04012981	4,316	384	3,175	4,302	384	3,175
SRR2185192	*C. graminicola*	254	PRJNA293991	SAMN04012982	15,539	660	7,068	15,468	659	7,068

Supplementary Table 1 Continued.

Run	Organism	Size (Mb)	BioProject	BioSample	All isoforms			Longest isoforms		
					Count	**AverageLength**	**MaxLength**	**Count**	**AverageLength**	**MaxLength**
SRR2185193	*C. graminicola*	301	PRJNA293991	SAMN04012983	15,960	725	8,076	15,898	725	8,076
SRR2185194	*C. graminicola*	250	PRJNA293991	SAMN04012984	15,689	661	7,070	15,628	661	7,070
SRR364389	*C. higginsianum*	1,215	PRJNA148307	SAMN00754065	62,857	846	15,545	57,955	769	15,545
SRR364390	*C. higginsianum*	1,197	PRJNA148307	SAMN00754065	62,113	846	14,389	57,203	770	14,389
SRR364391	*C. higginsianum*	1,302	PRJNA148307	SAMN00754066	66,522	870	13,072	60,976	792	13,072
SRR364392	*C. higginsianum*	1,283	PRJNA148307	SAMN00754066	66,532	865	14,119	61,019	789	14,119
SRR364393	*C. higginsianum*	1,349	PRJNA148307	SAMN00754067	68,153	866	15,621	62,903	798	15,621
SRR364394	*C. higginsianum*	1,326	PRJNA148307	SAMN00754067	67,307	869	15,498	62,275	803	15,498
SRR364395	*C. higginsianum*	1,108	PRJNA148307	SAMN00754068	33,850	894	10,102	30,750	823	10,102
SRR364396	*C. higginsianum*	1,160	PRJNA148307	SAMN00754069	34,259	944	10,381	30,652	846	10,381
SRR364397	*C. higginsianum*	1,284	PRJNA148307	SAMN00754070	34,947	965	13,140	31,021	866	13,140
SRR364398	*C. higginsianum*	1,270	PRJNA148307	SAMN00754071	65,375	801	12,820	60,595	734	12,820
SRR364399	*C. higginsianum*	1,269	PRJNA148307	SAMN00754071	64,887	805	15,543	60,341	738	15,543
SRR364400	*C. higginsianum*	1,237	PRJNA148307	SAMN00754072	67,560	816	15,495	62,907	759	15,495
SRR364401	*C. higginsianum*	1,236	PRJNA148307	SAMN00754072	67,328	824	12,975	62,674	763	12,975
SRR364402	*C. higginsianum*	1,281	PRJNA148307	SAMN00754073	67,897	827	11,775	63,170	767	11,775
SRR364403	*C. higginsianum*	1,280	PRJNA148307	SAMN00754073	67,205	832	11,792	62,672	772	11,792
SRR364404	*C. higginsianum*	1,210	PRJNA148307	SAMN00754074	48,274	854	13,751	45,697	809	13,751
SRR364405	*C. higginsianum*	1,146	PRJNA148307	SAMN00754074	48,175	826	11,790	45,704	786	11,790
SRR364406	*C. higginsianum*	1,275	PRJNA148307	SAMN00754075	49,292	838	15,494	46,869	797	15,494
SRR364407	*C. higginsianum*	1,203	PRJNA148307	SAMN00754075	49,129	823	12,753	46,752	781	12,753
SRR364408	*C. higginsianum*	1,205	PRJNA148307	SAMN00754076	44,687	850	11,536	42,311	808	11,536
SRR364409	*C. higginsianum*	1,140	PRJNA148307	SAMN00754076	44,897	830	10,242	42,374	789	10,242
SRR7287107	*C. higginsianum*	27,138	PRJNA475022	SAMN09378644	445,902	469	22,255	428,349	446	22,255
SRR7287108	*C. higginsianum*	23,414	PRJNA475022	SAMN09378672	181,887	473	13,640	161,042	436	13,640
SRR7287109	*C. higginsianum*	21,002	PRJNA475022	SAMN09378889	194,897	475	21,485	151,565	413	21,485
SRR10865243	*C. higginsianum*	10,144	PRJNA600305	SAMN13812491	42,474	1,314	13,740	36,585	1,154	13,740
SRR10865244	*C. higginsianum*	10,695	PRJNA600305	SAMN13812490	44,589	1,315	19,565	38,344	1,136	19,565
SRR10865245	*C. higginsianum*	8,684	PRJNA600305	SAMN13812468	38,116	1,360	16,487	32,902	1,189	16,487
SRR10865246	*C. higginsianum*	10,171	PRJNA600305	SAMN13812467	51,902	1,651	20,901	40,476	1,241	20,901
SRR10865247	*C. higginsianum*	11,612	PRJNA600305	SAMN13812458	53,256	1,624	17,373	41,618	1,237	17,373
SRR10865248	*C. higginsianum*	9,805	PRJNA600305	SAMN13812457	53,890	1,673	17,282	41,867	1,206	17,282
SRR1630188	*C. higginsianum*	653	PRJNA264846	SAMN03143065	13,706	397	10,795	13,687	396	10,795
SRR1630189	*C. higginsianum*	579	PRJNA264846	SAMN03143008	13,853	394	10,795	13,828	394	10,795
SRR1630190	*C. higginsianum*	579	PRJNA264846	SAMN03143058	13,665	383	10,795	13,649	383	10,795
SRR1630191	*C. higginsianum*	660	PRJNA264846	SAMN03143060	15,354	404	10,795	15,332	404	10,795
SRR1630192	*C. higginsianum*	831	PRJNA264846	SAMN03143030	21,075	445	10,795	20,999	444	10,795

Supplementary Table 1 Continued.

Run	Organism	Size (Mb)	BioProject	BioSample	All isoforms			Longest isoforms		
					Count	**AverageLength**	**MaxLength**	**Count**	**AverageLength**	**MaxLength**
SRR1630193	*C. higginsianum*	624	PRJNA264846	SAMN03143086	18,814	433	10,795	18,786	432	10,795
SRR1630194	*C. higginsianum*	704	PRJNA264846	SAMN03143084	19,135	443	10,795	19,098	442	10,795
SRR1630195	*C. higginsianum*	886	PRJNA264846	SAMN03143021	19,328	434	10,795	19,286	433	10,795
SRR1630196	*C. higginsianum*	478	PRJNA264846	SAMN03143041	10,639	386	10,795	10,630	386	10,795
SRR1630197	*C. higginsianum*	453	PRJNA264846	SAMN03143014	7,910	383	10,795	7,896	383	10,795
SRR1630198	*C. higginsianum*	439	PRJNA264846	SAMN03143025	11,485	396	10,795	11,473	395	10,795
SRR1630199	*C. higginsianum*	470	PRJNA264846	SAMN03143069	10,080	389	10,795	10,073	388	10,795
SRR1630200	*C. higginsianum*	424	PRJNA264846	SAMN03143070	13,848	407	10,795	13,838	407	10,795
SRR1630201	*C. higginsianum*	453	PRJNA264846	SAMN03143023	15,254	421	10,795	15,234	421	10,795
SRR1630202	*C. higginsianum*	462	PRJNA264846	SAMN03143028	15,004	416	10,795	14,980	415	10,795
SRR1630203	*C. higginsianum*	453	PRJNA264846	SAMN03143020	16,301	429	10,795	16,282	428	10,795
SRR1630204	*C. higginsianum*	685	PRJNA264846	SAMN03143049	10,290	391	10,795	10,281	391	10,795
SRR1630205	*C. higginsianum*	633	PRJNA264846	SAMN03143032	10,302	391	10,795	10,294	390	10,795
SRR1630206	*C. higginsianum*	693	PRJNA264846	SAMN03143013	13,443	408	10,795	13,426	408	10,795
SRR1630207	*C. higginsianum*	761	PRJNA264846	SAMN03143033	12,379	389	10,795	12,373	389	10,795
SRR1630208	*C. higginsianum*	621	PRJNA264846	SAMN03143010	11,554	395	10,795	11,530	395	10,795
SRR1630209	*C. higginsianum*	951	PRJNA264846	SAMN03143039	15,039	418	10,795	14,983	416	10,795
SRR1630210	*C. higginsianum*	687	PRJNA264846	SAMN03143036	14,367	397	10,795	14,355	397	10,795
SRR1630211	*C. higginsianum*	1,023	PRJNA264846	SAMN03143015	13,272	384	10,795	13,243	384	10,795
SRR1630212	*C. higginsianum*	703	PRJNA264846	SAMN03143004	13,883	396	10,795	13,848	395	10,795
SRR1630213	*C. higginsianum*	809	PRJNA264846	SAMN03143019	16,308	395	10,795	16,272	395	10,795
SRR1630214	*C. higginsianum*	734	PRJNA264846	SAMN03142993	15,728	401	10,795	15,699	400	10,795
SRR1630215	*C. higginsianum*	711	PRJNA264846	SAMN03143071	16,761	398	10,795	16,734	398	10,795
SRR1630216	*C. higginsianum*	726	PRJNA264846	SAMN03143026	12,763	372	10,795	12,722	371	10,795
SRR1630217	*C. higginsianum*	779	PRJNA264846	SAMN03143052	15,951	394	10,795	15,928	394	10,795
SRR1630218	*C. higginsianum*	672	PRJNA264846	SAMN03143000	13,867	377	10,795	13,843	377	10,795
SRR1630219	*C. higginsianum*	439	PRJNA264846	SAMN03143043	14,167	440	10,795	14,152	440	10,795
SRR1630220	*C. higginsianum*	439	PRJNA264846	SAMN03143076	14,178	435	10,795	14,158	434	10,795
SRR1630221	*C. higginsianum*	752	PRJNA264846	SAMN03143042	19,010	486	10,795	18,949	485	10,795
SRR1630222	*C. higginsianum*	452	PRJNA264846	SAMN03143048	12,802	440	10,795	12,785	440	10,795
SRR1630223	*C. higginsianum*	320	PRJNA264846	SAMN03143034	6,815	396	10,795	6,781	396	10,795
SRR2072467	*C. incanum*	3,812	PRJNA287627	SAMN03784846	27,457	1,391	16,835	24,706	1,233	16,835
SRR2072468	*C. incanum*	986	PRJNA287627	SAMN03784846	20,640	1,179	13,744	19,480	1,115	13,744
SRR2072469	*C. incanum*	3,960	PRJNA287627	SAMN03784847	28,144	1,399	16,266	24,958	1,231	16,266
SRR2072470	*C. incanum*	999	PRJNA287627	SAMN03784847	21,878	1,171	13,082	20,608	1,101	13,082
SRR2072471	*C. incanum*	3,822	PRJNA287627	SAMN03784848	27,114	1,410	36,189	24,223	1,262	36,189
SRR2072472	*C. incanum*	925	PRJNA287627	SAMN03784848	21,337	1,157	13,106	20,183	1,091	13,106

Supplementary Table 1 Continued.

Run	**Organism**	**Size (Mb)**	**BioProject**	**BioSample**	**All isoforms**			**Longest isoforms**		
					Count	**AverageLength**	**MaxLength**	**Count**	**AverageLength**	**MaxLength**
SRR5738849	*C. sansevieriae*	16	PRJNA391063	SAMN07267235	921	346	2,775	902	345	2,775
SRR2072452	*C. tofieldiae*	3,934	PRJNA287627	SAMN03784834	29,336	1,471	34,964	26,675	1,303	34,964
SRR2072453	*C. tofieldiae*	5,676	PRJNA287627	SAMN03784835	31,825	1,542	36,144	28,474	1,326	36,144
SRR2072454	*C. tofieldiae*	3,345	PRJNA287627	SAMN03784836	29,033	628	10,795	28,110	607	10,795
SRR3482280	*C. truncatum*	6,951	PRJNA320930	SAMN04961170	169,383	593	13,477	159,062	547	13,477
SRR5576236	*C. truncatum*	6,554	PRJNA380621	SAMN06629161	210,139	727	17,492	173,194	578	17,492
SRR5576237	*C. truncatum*	6,023	PRJNA380621	SAMN06629161	216,084	729	16,870	175,134	565	16,870
SRR5576238	*C. truncatum*	10,827	PRJNA380621	SAMN06629161	40,168	1,640	21,399	33,601	1,312	21,399
SRR5576239	*C. truncatum*	13,495	PRJNA380621	SAMN06629161	51,275	872	13,488	44,566	691	13,488
SRR5576240	*C. truncatum*	6,105	PRJNA380621	SAMN06629161	53,816	785	14,199	47,787	644	14,199
SRR5576241	*C. truncatum*	7,677	PRJNA380621	SAMN06629161	221,767	796	17,644	176,625	604	17,644

Supplementary Table 2 Genomic profiles of 102 *Colletotrichum* genomes.

Assembly	**Platform**[2]	**Organism**	**Strain**	**Complex**	**Genome (bp)**	**GenomeNRE**[3] **(bp)**	**%RE**[4]	**RE (bp)**	**#Gene**[5]	**#Pfam**	**%Pfam**[6]	**#Contig**	**N50**[7]	**%BUSCO**	**%GC**
GCA_020457915.1	HiSeq	*C. acutatum*	VPRI 42878	Acutatum	49,531,067	48,748,476	1.58	782,591	12,932	9,498	73.45	2,061	59,120	96.70	52.28
GCA_020457925.1	HiSeq	*C. acutatum*	VPRI 42879	Acutatum	49,885,583	48,738,215	2.3	1,147,368	12,877	9,475	73.58	2,026	63,304	96.70	52.05
GCA_020457905.1	HiSeq	*C. acutatum*	VPRI 42883	Acutatum	49,697,799	48,380,807	2.65	1,316,992	12,816	9,464	73.85	1,879	62,028	96.80	51.91
GCA_020465895.1	HiSeq	*C. acutatum*	VPRI 43374	Acutatum	50,870,935	48,876,794	3.92	1,994,141	12,853	9,519	74.06	365	523,865	97.90	51.47
GCA_020465855.1	HiSeq	*C. acutatum*	VPRI 43375	Acutatum	51,057,129	49,101,641	3.83	1,955,488	12,911	9,569	74.12	451	475,364	98.00	51.48
GCA_020465875.1	HiSeq	*C. acutatum*	VPRI 43376	Acutatum	51,025,733	49,132,678	3.71	1,893,055	12,902	9,569	74.17	464	613,842	98.00	51.45
GCA_020465835.1	HiSeq	*C. acutatum*	VPRI 43377	Acutatum	50,496,057	48,587,306	3.78	1,908,751	12,763	9,500	74.43	194	571,668	98.00	51.44
GCA_020465815.1	HiSeq	*C. acutatum*	VPRI 43378	Acutatum	51,018,208	48,824,425	4.3	2,193,783	12,816	9,488	74.03	396	493,744	98.20	51.38
GCA_020465795.1	HiSeq	*C. acutatum*	VPRI 43379	Acutatum	50,469,302	48,531,281	3.84	1,938,021	12,744	9,469	74.30	474	447,385	98.10	51.46
GCA_020465775.1	HiSeq	*C. acutatum*	VPRI 43380	Acutatum	50,496,501	48,486,740	3.98	2,009,761	12,747	9,475	74.33	284	628,759	97.70	51.44
GCF_013390185.1	PacBio	*C. aenigma*	Cg56	Gloeosporioides	59,189,926	55,597,097	6.07	3,592,829	15,175	10,776	71.01	79	5,181,253	97.50	52.45
GCA_009806415.1	HiSeq	*C. asianum*	ICMP 18580	Gloeosporioides	64,731,264	56,762,845	12.31	7,968,419	15,432	10,921	70.77	486	389,703	97.90	49.38
GCA_014706365.1	Nanopore	*C. australisinense*	CGMCC 3.18886	Acutatum	55,304,262	51,986,006	6	3,318,256	17,809	9,871	55.43	28	5,698,484	87.30	51.36
GCA_011947485.2	HiSeq	*C. camelliae*	CcLH18	Gloeosporioides	57,836,951	52,440,763	9.33	5,396,188	14,387	10,279	71.45	580	198,072	86.40	49.72

Supplementary Table 2 Continued.

Assembly	Platform[2]	Organism	Strain	Complex	Genome (bp)	GenomeNRE[3] (bp)	%RE[4]	RE (bp)	#Gene[5]	#Pfam	%Pfam[6]	#Contig	N50[7]	%BUSCO	%GC
GCA_018853505.1	PacBio	*C. camelliae*	LS-19	Gloeosporioides	67,471,331	58,808,012	12.84	8,663,319	15,867	10,963	69.09	19	5,851,421	90.70	49.04
GCA_020466075.1	Nanorpore	*C. coccodes*	CcND04	Singleton	50,243,842	49,676,087	1.13	567,755	13,013	9,346	71.82	7,394	28,750	89.70	53.78
GCA_016807845.1	Nanopore and NovaSeq	*C. eleusines*	NJC-16	Graminicola-caudatum	53,507,824	45,058,939	15.79	8,448,885	11,109	8,090	72.82	15	5,135,075	97.90	51.11
GCA_019425465.1	PacBio	*C. falcatum*	CF08	Graminicola-caudatum	56,082,967	49,280,103	12.13	6,802,864	15,856	12,310	77.64	238	488,304	92.50	52.75
GCA_001484525.1	HiSeq	*C. falcatum*	Cf671	Graminicola-caudatum	48,186,431	42,938,929	10.89	5,247,502	11,346	7,821	68.93	4,358	29,329	91.90	51.62
GCA_014705555.1	NextSeq	*C. fioriniae*	HC25	Acutatum	49,841,025	48,515,254	2.66	1,325,771	12,754	9,445	74.06	1,763	64,091	97.20	51.72
GCA_014705545.1	NextSeq	*C. fioriniae*	HC296	Acutatum	49,753,773	48,524,855	2.47	1,228,918	12,787	9,452	73.92	2,350	50,913	97.00	51.79
GCA_002930455.1	HiSeq	*C. fioriniae*	HC89	Acutatum	50,150,864	49,122,771	2.05	1,028,093	12,704	9,458	74.45	315	544,764	97.80	51.91
GCA_002930425.1	HiSeq	*C. fioriniae*	HC91	Acutatum	49,792,461	49,075,450	1.44	717,011	12,688	9,411	74.17	970	468,533	97.80	52.20
GCF_000582985.1	HiSeq	*C. fioriniae*	IMI 504882	Acutatum	49,002,765	48,796,953	0.42	205,812	12,787	9,489	74.21	1,096	137,254	97.70	52.54
GCA_014705275.1	NextSeq	*C. fioriniae*	KY119	Acutatum	49,840,158	48,663,930	2.36	1,176,228	12,827	9,471	73.84	2,238	49,414	97.40	51.72
GCA_014705335.1	NextSeq	*C. fioriniae*	KY323	Acutatum	49,661,969	48,370,758	2.6	1,291,211	12,730	9,409	73.91	1,274	84,093	97.40	51.77
GCA_014705265.1	NextSeq	*C. fioriniae*	KY640	Acutatum	49,691,380	48,618,046	2.16	1,073,334	12,819	9,464	73.83	2,696	49,006	97.10	51.90
GCA_014705245.1	NextSeq	*C. fioriniae*	KY646	Acutatum	50,064,071	48,917,604	2.29	1,146,467	12,829	9,440	73.58	2,156	53,113	97.20	51.66
GCA_020464245.1	HiSeq	*C. fioriniae*	VPRI 42872	Acutatum	49,290,758	48,960,510	0.67	330,248	12,968	9,477	73.08	3,173	33,646	95.70	52.42
GCA_017589655.1	HiSeq	*C. fructicola*	1104-6	Gloeosporioides	56,580,086	55,776,649	1.42	803,437	15,680	11,231	71.63	815	257,042	97.50	53.52
GCA_002887685.1	HiSeq	*C. fructicola*	15060	Gloeosporioides	55,915,688	55,591,377	0.58	324,311	15,481	11,111	71.77	1,049	480,505	97.60	53.24
GCA_013201925.1	HiSeq	*C. fructicola*	Cf245	Gloeosporioides	56,060,367	55,555,824	0.9	504,543	15,401	11,115	72.17	1,194	378,303	97.70	53.15
GCA_013390205.1	PacBio	*C. fructicola*	Cf413	Gloeosporioides	56,528,462	55,137,862	2.46	1,390,600	15,326	11,066	72.20	14	4,902,990	97.40	53.19
GCA_013201905.1	HiSeq	*C. fructicola*	Cf415	Gloeosporioides	56,012,246	55,452,124	1	560,122	15,389	11,087	72.04	805	412,455	98.20	53.20
GCA_013201875.1	HiSeq	*C. fructicola*	CfS4	Gloeosporioides	57,426,314	56,943,933	0.84	482,381	15,767	11,229	71.22	1,588	306,788	97.60	53.17
GCA_012932255.1	PacBio and HiSeq	*C. fructicola*	Cg38 S1	Gloeosporioides	58,952,871	57,166,599	3.03	1,786,272	15,790	11,281	71.44	600	1,273,051	97.60	53.13
GCF_009771025.1	Nanopore and MiSeq	*C. fructicola*	CGMCC 3.17371	Gloeosporioides	58,056,435	56,883,695	2.02	1,172,740	15,787	11,268	71.38	447	905,947	97.70	53.20
GCA_014705485.1	NextSeq	*C. fructicola*	HC540	Gloeosporioides	57,126,425	56,560,873	0.99	565,552	15,725	11,180	71.10	2,411	107,635	97.40	53.19
GCF_000319635.2	HiSeq	*C. fructicola*	MAFF 245243	Gloeosporioides	55,607,143	55,451,443	0.28	155,700	15,742	11,060	70.26	1,241	112,809	94.40	53.58
GCA_000446055.1	Roche 454	*C. gloeosporioides*	Cg-14	Gloeosporioides	53,209,944	53,013,067	0.37	196,877	15,422	10,857	70.40	4,537	25,337	90.30	53.37
GCA_011947415.1	HiSeq	*C. gloeosporioides*	CgLH19	Gloeosporioides	55,563,689	54,035,688	2.75	1,528,001	15,028	10,869	72.32	197	553,822	98.30	52.39
GCA_011428055.1	HiSeq	*C. gloeosporioides*	COLG-95	Gloeosporioides	53,724,932	52,500,004	2.28	1,224,928	15,189	10,723	70.60	3,622	20,957	91.10	52.44

Supplementary Table 2 Continued.

Assembly	Platform[2]	Organism	Strain	Complex	Genome (bp)	GenomeNRE[3] (bp)	%RE[4]	RE (bp)	#Gene[5]	#Pfam	%Pfam[6]	#Contig	N50[7]	%BUSCO	%GC
GCF_000149035.1	Roche 454 and Sanger	*C. graminicola*	M1.001	Graminicola-caudatum	51,604,658	43,977,490	14.78	7,627,168	10,852	8,006	73.77	653	579,194	97.80	49.12
GCA_001951205.1	Roche 454	*C. graminicola*	M5.001	Graminicola-caudatum	59,914,086	43,605,472	27.22	16,308,614	11,314	8,114	71.72	8,455	16,615	94.10	45.47
GCF_001672515.1	PacBio	*C. higginsianum*	IMI 349063	Destructivum	50,716,103	47,962,219	5.43	2,753,884	12,940	9,388	72.55	25	5,201,688	97.80	54.41
GCA_004920355.1	PacBio	*C. higginsianum*	MAFF 305635-RFP	Destructivum	49,786,461	47,570,963	4.45	2,215,498	12,936	9,375	72.47	28	5,064,939	97.10	54.61
GCA_019693695.1	Oxford Nanopore	*C. horii*	FJ-1	Gloeosporioides	74,324,508	58,084,603	21.85	16,239,905	15,604	11,077	70.99	42	3,135,412	97.50	46.79
GCA_001625285.1	HiSeq	*C. incanum*	MAFF 238704	Spaethianum	53,597,417	51,121,216	4.62	2,476,201	12,701	9,292	73.16	2,896	177,164	97.20	52.14
GCA_001855235.1	GAIIx	*C. incanum*	MAFF 238712	Spaethianum	53,254,579	50,693,034	4.81	2,561,545	12,648	9,303	73.55	1,036	292,512	97.30	52.14
GCF_011947395.1	HiSeq	*C. karsti*	CkLH20	Boninense	51,850,041	50,621,195	2.37	1,228,846	14,395	10,402	72.26	127	857,663	97.50	52.69
GCA_003386485.1	Illumina	*C. lentis*	CT-30	Destructivum	56,100,083	45,441,067	19	10,659,016	11,375	8,052	70.79	50	4,891,426	97.40	49.95
GCA_001693015.2	HiSeq	*C. lindemuthianum*	83.501	Orbiculare	97,411,422	45,461,911	53.33	51,949,511	11,998	8,507	70.90	1,857	111,275	97.20	37.56
GCA_001693025.2	HiSeq	*C. lindemuthianum*	89 A2	Orbiculare	99,166,722	47,272,776	52.33	51,893,946	11,951	8,479	70.95	1,276	158,217	96.80	37.32
GCA_015832465.1	Illumina	*C. liriopes*	A2	Spaethianum	53,087,791	50,258,212	5.33	2,829,579	13,045	9,572	73.38	407	288,433	97.80	52.72
GCA_002814275.1	HiSeq	*C. musae*	CMM 4420	Gloeosporioides	49,118,800	48,283,780	1.7	835,020	14,762	9,800	66.39	12,435	7,105	87.20	53.62
GCA_014235935.1	MiSeq	*C. musicola*	IMI 507128	Orchidearum	52,725,698	49,203,621	6.68	3,522,077	14,401	9,992	69.38	2,463	46,381	97.80	54.97
GCA_014705455.1	NextSeq	*C. nymphaeae*	HC646	Acutatum	51,255,053	50,399,094	1.67	855,959	13,251	9,686	73.10	3,322	67,103	96.70	52.31
GCA_001563115.1	Illumina	*C. nymphaeae*	IMI 504889	Acutatum	49,956,273	49,646,544	0.62	309,729	13,158	9,648	73.32	1,884	91,051	97.00	52.75
GCA_014705155.1	NextSeq	*C. nymphaeae*	KY563	Acutatum	50,391,861	49,837,551	1.1	554,310	13,125	9,615	73.26	1,698	131,269	97.70	52.56
GCA_014705095.1	NextSeq	*C. nymphaeae*	KY567	Acutatum	51,509,711	50,412,554	2.13	1,097,157	13,328	9,797	73.51	3,013	127,617	97.30	52.28
GCA_014705085.1	NextSeq	*C. nymphaeae*	KY613	Acutatum	50,577,973	49,475,373	2.18	1,102,600	13,095	9,589	73.23	860	138,789	97.20	52.21
GCA_014705125.1	NextSeq	*C. nymphaeae*	KY745	Acutatum	50,509,097	49,448,406	2.1	1,060,691	13,088	9,584	73.23	1,397	84,913	96.90	52.23
GCF_001831195.1	HiSeq	*C. orchidophilum*	IMI 309357	Singleton	48,556,462	45,842,156	5.59	2,714,306	11,687	8,553	73.18	321	284,473	97.50	51.07
GCA_014235945.1	MiSeq	*C. plurivorum*	LFN00145	Orchidearum	49,703,650	48,987,917	1.44	715,733	14,226	10,023	70.46	984	130,438	97.40	55.86
GCA_001563125.1	Illumina	*C. salicis*	CBS 607.94	Acutatum	48,373,413	47,850,980	1.08	522,433	12,668	9,202	72.64	2,776	46,166	97.40	52.85
GCA_002749775.1	IonTorrent	*C. sansevieriae*	Sa-1-2	Agaves	51,201,343	47,714,532	6.81	3,486,811	12,300	8,446	68.67	8,646	15,119	95.80	50.74
GCA_018906765.1	GAII	*C. scovillei*	Coll-153	Acutatum	50,113,654	49,091,335	2.04	1,022,319	12,922	9,574	74.09	59	3,536,762	97.80	52.04
GCA_018906675.1	GAII	*C. scovillei*	Coll-365	Acutatum	49,921,975	48,998,418	1.85	923,557	12,901	9,563	74.13	59	3,528,503	97.70	52.05

Supplementary Table 2 Continued.

Assembly	Platform[2]	Organism	Strain	Complex	Genome (bp)	GenomeNRE[3] (bp)	%RE[4]	RE (bp)	#Gene[5]	#Pfam	%Pfam[6]	#Contig	N50[7]	%BUSCO	%GC
GCA_018907675.1	HiSeq	*C. scovillei*	Coll-524	Acutatum	51,490,565	50,321,729	2.27	1,168,836	13,016	9,604	73.79	54	3,597,184	97.80	51.98
GCF_011075155.1	Nanopore	*C. scovillei*	TJNH1	Acutatum	52,032,179	50,320,320	3.29	1,711,859	13,101	9,658	73.72	16	4,871,440	97.60	51.76
GCA_006783085.1	HiSeq	*C. shisoi*	JCM 31818	Destructivum	69,667,657	45,102,841	35.26	24,564,816	11,976	8,352	69.74	20,745	14,839	97.20	46.68
GCA_013201865.1	HiSeq	*C. siamense*	CAD1	Gloeosporioides	58,399,603	57,272,491	1.93	1,127,112	15,572	11,093	71.24	308	881,809	97.40	52.33
GCA_013201745.1	HiSeq	*C. siamense*	CAD2	Gloeosporioides	58,148,672	56,607,732	2.65	1,540,940	15,433	11,023	71.42	214	709,060	97.80	52.31
GCA_013201795.1	HiSeq	*C. siamense*	CAD4	Gloeosporioides	58,153,656	56,746,338	2.42	1,407,318	15,435	11,053	71.61	199	823,408	97.90	52.32
GCA_013201755.1	HiSeq	*C. siamense*	CAD5	Gloeosporioides	57,642,425	56,380,056	2.19	1,262,369	15,429	11,041	71.56	294	656,815	97.70	52.41
GCF_013390195.1	PacBio	*C. siamense*	Cg363	Gloeosporioides	62,936,415	58,795,199	6.58	4,141,216	15,684	11,107	70.82	22	5,441,060	97.70	50.90
GCA_011426385.1	HiSeq	*C. siamense*	COLG-34	Gloeosporioides	50,611,973	49,483,326	2.23	1,128,647	14,321	10,299	71.92	562	3,772,317	90.40	52.54
GCA_011426375.1	HiSeq	*C. siamense*	COLG-38	Gloeosporioides	55,426,472	54,306,857	2.02	1,119,615	15,251	10,936	71.71	223	5,443,879	95.90	52.64
GCA_011947465.1	HiSeq	*C. siamense*	CsLH17	Gloeosporioides	60,116,880	58,181,116	3.22	1,935,764	15,673	11,127	70.99	319	441,266	98.00	52.07
GCA_014706345.1	Oxford Nanopore	*C. siamense*	HBCG01	Gloeosporioides	58,357,740	57,546,567	1.39	811,173	15,530	11,118	71.59	24	4,579,047	97.20	52.33
GCA_008520285.1	HiSeq	*C. siamense*	ICMP 18578	Gloeosporioides	55,961,623	54,836,794	2.01	1,124,829	15,157	10,963	72.33	235	927,217	97.80	52.67
GCA_004367935.1	HiSeq	*C. sidae*	CBS 518.97	Orbiculare	86,827,816	47,182,235	45.66	39,645,581	12,690	9,104	71.74	14,826	39,310	97.60	37.96
GCA_001563135.1	Illumina	*C. simmondsii*	CBS 122122	Acutatum	50,474,234	49,353,706	2.22	1,120,528	12,876	9,583	74.43	929	292,136	97.40	51.81
GCA_020457885.1	HiSeq	*C. simmondsii*	VPRI 42936	Acutatum	49,176,323	48,689,477	0.99	486,846	12,980	9,581	73.81	1,987	53,980	96.10	52.52
GCA_014235955.1	MiSeq	*C. sojae*	LFN0009	Orchidearum	49,351,125	48,546,702	1.63	804,423	14,153	9,889	69.87	1,140	104,192	97.70	55.92
GCA_004366825.1	HiSeq	*C. spinosum*	CBS 515.97	Orbiculare	82,734,851	47,398,796	42.71	35,336,055	12,571	9,049	71.98	10,715	93,593	98.00	38.71
GCA_001951195.1	Roche 454	*C. sublineola*	CgSl1	Graminicola-caudatum	64,848,627	49,414,654	23.8	15,433,973	11,613	8,406	72.38	548	339,276	97.80	46.54
GCA_020631755.1	PacBio	*C. sublineola*	CsGL1	Graminicola-caudatum	67,666,679	49,241,042	27.23	18,425,637	11,816	8,531	72.20	74	2,691,276	97.80	46.12
GCA_000696135.1	Illumina ?[1]	*C. sublineola*	TX430BB	Graminicola-caudatum	46,755,813	46,185,392	1.22	570,421	11,742	8,448	71.95	1,625	70,717	97.70	53.60
GCA_005350895.1	Illumina ?	*C. tanaceti*	BRIP 57314	Destructivum	57,912,474	45,414,962	21.58	12,497,512	11,424	8,046	70.43	5,242	103,135	96.90	49.29
GCA_001625265.1	Roche 454 and HiSeq	*C. tofieldiae*	861	Spaethianum	52,836,184	51,303,935	2.9	1,532,249	13,154	9,593	72.93	1,046	209,340	97.80	53.28
GCA_001618715.1	HiSeq	*C. tofieldiae*	CBS 127615	Spaethianum	52,719,552	50,616,042	3.99	2,103,510	12,984	9,542	73.49	791	377,287	97.60	52.89
GCA_001618735.1	HiSeq	*C. tofieldiae*	CBS 130851	Spaethianum	53,216,229	51,396,234	3.42	1,819,995	13,027	9,505	72.96	476	364,528	97.80	52.62
GCA_001618705.1	HiSeq	*C. tofieldiae*	CBS 168.49	Spaethianum	52,962,735	50,976,632	3.75	1,986,103	12,992	9,528	73.34	1,300	294,339	97.70	52.99
GCA_001618725.1	HiSeq	*C. tofieldiae*	CBS 495.85	Spaethianum	53,491,261	51,881,174	3.01	1,610,087	13,244	9,666	72.98	1,647	245,346	97.70	53.26

Supplementary Table 2 Continued.

Assembly	**Platform**[2]	**Organism**	**Strain**	**Complex**	**Genome (bp)**	**GenomeNRE**[3] **(bp)**	**%RE**[4]	**RE (bp)**	**#Gene**[5]	**#Pfam**	**%Pfam**[6]	**#Contig**	**N50**[7]	**%BUSCO**	**%GC**
GCA_004367215.1	PacBio and HiSeq	*C. trifolii*	MAFF 305078	Orbiculare	109,659,959	49,708,859	54.67	59,951,100	12,671	9,018	71.17	10,473	35,434	97.70	36.62
GCF_014235925.1	PacBio and HiSeq	*C. truncatum*	CMES1059	Truncatum	56,102,516	51,552,602	8.11	4,549,914	12,538	9,420	75.13	126	831,709	97.20	50.12
GCA_011947275.1	HiSeq	*C. truncatum*	KLC.C- 4	Truncatum	44,146,251	42,804,205	3.04	1,342,046	10,354	7,860	75.91	60	1,509,490	81.40	49.74
GCA_008131165.1	NextSeq	*C. truncatum*	KLC.C5	Truncatum	59,632,073	57,711,920	3.22	1,920,153	17,094	13,032	76.24	5,157	132,868	97.70	51.43
GCA_002632455.2	PacBio and HiSeq	*C. truncatum*	MTCC 3414	Truncatum	57,912,832	55,445,745	4.26	2,467,087	12,585	9,474	75.28	70	2,268,405	97.40	49.38
GCA_013201765.1	HiSeq	*C. viniferum*	CGW01	Gloeosporioides	68,452,207	56,288,250	17.77	12,163,957	14,989	10,481	69.92	5,864	144,434	97.80	47.13
GCA_020226115.1	PacBio and NovaSeq	*C. viniferum*	CvYL2a	Gloeosporioides	73,406,341	57,088,111	22.23	16,318,230	14,940	10,445	69.91	70	2,103,953	98.00	45.06

Notes:

[1] Illumina ?, no detailed information.

[2] Second-generation sequencing technologies include Illumina (HiSeq, MiSeq, NovaSeq) and Roche 454; Third-generation sequencing technologies includes Nanorpore and Pacbio.

[3] GenomNRE = genome size – repeat size.

[4] %RE = Repeat size / genome size.

[5] # means number; #gene means the number of predicted genes.

[6] % means percentage; %Pfam = #gene with pfam annotation / #gene

[7] N50 means the contig length such that 50% of the de novo assembled genome lies in blocks of this size or larger.

Supplementary Table 3 Details of the outgroup.

Assembly	**Organism**	**Strain**	**Genome (bp)**	**BioSample**	**BioProject**
GCF_003724135.2	*Verticillium nonalfalfae*	VnAa140/NRRL 66861	31,747,483	SAMN10135221	PRJNA493511

Supplementary Table 4 Details of secretomes of 102 *Colletotrichum* genomes.

Organism	**Strain**	**CAZymes**						**Protease**								**Lipase**		**SSP**	**Other**	**Secretome**
		CBM	**PL**	**CE**	**AA**	**GT**	**GH**	**Aspartic**	**Cysteine**	**Glutamic**	**Serine**	**Mixed**	**Metallo**	**Asparagine**	**Unknown**	**GGGX-type**	**GX-type**			
C. nymphaeae	KY563	25	33	62	119	3	184	11	0	1	72	5	42	0	0	4	14	428	304	1,289
C. nymphaeae	IMI 504889	25	33	59	119	3	186	11	0	1	75	5	42	0	0	4	13	431	304	1,294
C. nymphaeae	KY567	25	34	61	119	3	184	11	0	1	73	5	42	0	0	4	14	438	312	1,308
C. nymphaeae	HC646	23	35	60	114	2	189	12	0	1	75	5	43	0	0	3	15	432	302	1,293
C. nymphaeae	KY745	24	33	60	113	2	181	11	0	1	79	5	44	0	0	3	15	425	318	1,296
C. nymphaeae	KY613	24	33	61	112	2	181	11	0	1	79	5	43	0	0	3	15	435	318	1,305

Supplementary Table 4 Continued.

Organism	**Strain**	**CAZymes**						**Protease**								**Lipase**		**SSP**	**Other**	**Secretome**
		CBM	**PL**	**CE**	**AA**	**GT**	**GH**	**Aspartic**	**Cysteine**	**Glutamic**	**Serine**	**Mixed**	**Metallo**	**Asparagine**	**Unknown**	**GGGX-type**	**GX-type**			
C. scovillei	Coll-365	26	32	60	119	4	184	11	0	1	80	5	46	0	0	3	19	436	325	1,329
C. scovillei	Coll-153	26	32	60	119	4	184	11	0	1	80	5	46	0	0	3	19	434	327	1,329
C. scovillei	Coll-524	26	33	60	120	4	184	11	0	1	79	5	45	0	0	3	18	433	324	1,325
C. scovillei	TJNH1	26	33	62	122	4	190	11	0	1	79	5	44	0	0	3	16	437	320	1,334
C. australisinense	CGMCC 3.18886	30	34	53	101	2	154	11	0	1	68	4	37	0	0	3	11	485	283	1,263
C. simmondsii	CBS 122122	24	32	65	119	3	191	10	0	1	73	5	45	0	0	3	15	427	325	1,320
C. simmondsii	VPRI 42936	25	31	64	117	3	194	10	0	0	75	5	44	0	0	5	14	444	314	1,326
C. acutatum	VPRI 43374	25	34	58	120	3	185	11	0	1	80	4	43	0	0	3	18	411	324	1,299
C. acutatum	VPRI 43375	25	34	58	121	3	184	11	0	1	78	4	43	0	0	4	18	413	327	1,302
C. acutatum	VPRI 43380	25	34	58	120	3	183	11	0	1	78	4	43	0	0	3	17	417	325	1,302
C. acutatum	VPRI 42878	24	34	57	110	3	179	11	1	1	77	4	42	0	0	3	17	421	316	1,280
C. acutatum	VPRI 42883	25	34	57	120	3	181	11	0	1	79	4	42	0	0	4	18	416	318	1,291
C. acutatum	VPRI 43377	24	34	58	118	3	185	10	0	1	76	4	43	0	0	4	18	413	330	1,299
C. acutatum	VPRI 43379	24	34	58	120	3	184	10	0	1	76	4	43	0	0	4	17	414	327	1,298
C. acutatum	VPRI43376	25	34	58	122	3	185	10	0	1	78	4	43	0	0	3	17	418	327	1,308
C. acutatum	VPRI43378	24	35	57	113	2	183	11	0	1	78	4	44	0	0	3	17	416	323	1,291
C. acutatum	VPRI42879	24	35	57	115	2	176	11	1	1	77	4	43	0	0	3	17	423	316	1,285
C. fioriniae	HC91	25	37	60	103	1	178	10	1	1	78	4	42	0	0	5	13	425	313	1,278
C. fioriniae	HC89	24	36	60	109	1	179	10	1	1	78	4	43	0	0	5	13	429	316	1,291
C. fioriniae	KY640	24	35	59	103	1	178	10	0	1	76	4	42	0	0	5	12	428	320	1,281
C. fioriniae	HC296	24	35	55	111	2	180	9	0	1	80	4	43	0	0	5	14	432	305	1,281
C. fioriniae	HC25	25	35	56	112	2	180	10	1	1	76	4	43	0	0	5	14	428	311	1,284
C. fioriniae	KY323	24	37	58	110	2	180	9	1	1	78	4	43	0	0	5	13	429	313	1,289
C. fioriniae	KY119	26	36	56	112	2	179	10	0	1	77	4	43	0	0	5	14	423	314	1,283
C. fioriniae	KY646	25	35	59	111	2	176	10	0	1	75	4	45	0	0	5	14	434	312	1,289
C. fioriniae	IMI 504882	26	34	58	118	3	184	12	1	1	81	4	46	0	0	3	14	432	314	1,314
C. fioriniae	VPRI 42872	25	34	60	113	3	185	12	1	1	78	4	44	0	0	4	14	440	302	1,302
C. salicis	CBS 607.94	27	32	54	100	3	152	10	1	1	74	4	39	0	0	1	9	387	280	1,164
C. orchidophilum	IMI 309357	18	28	53	84	3	127	10	0	1	51	3	38	0	0	0	9	324	221	961

Supplementary Table 4 Continued.

Organism	Strain	CAZymes						Protease								Lipase		SSP	Other	Secretome
		CBM	**PL**	**CE**	**AA**	**GT**	**GH**	**Aspartic**	**Cysteine**	**Glutamic**	**Serine**	**Mixed**	**Metallo**	**Asparagine**	**Unknown**	**GGGX-type**	**GX-type**			
C. graminicola	M5.001	17	14	42	77	1	129	7	1	1	46	6	37	0	0	0	5	305	171	854
C. graminicola	M1.001	21	14	44	82	1	138	8	1	1	46	6	36	0	0	0	6	287	191	876
C. falcatum	Cf671	18	9	32	62	3	122	7	1	1	51	4	31	0	0	0	7	348	176	865
C. falcatum	CF08	14	12	38	63	3	135	6	2	1	57	7	39	3	3	1	7	475	399	1,257
C. eleusines	NJC-16	16	13	45	80	5	142	7	2	1	49	7	33	0	1	1	6	299	200	900
C. sublineola	CgSl1	24	16	46	86	3	154	6	2	1	52	3	37	0	0	1	7	353	220	1,003
C. sublineola	CsGL1	24	16	45	86	3	156	6	2	1	52	3	38	0	0	1	7	361	225	1,018
C. sublineola	TX430BB	23	14	45	84	3	154	6	2	1	49	3	33	0	0	1	7	331	218	966
C. tofieldiae	CBS 495.85	42	38	49	109	4	186	9	1	2	83	6	49	0	0	1	13	390	278	1,246
C. tofieldiae	CBS 130851	38	35	50	108	4	181	9	1	2	84	6	47	0	0	1	12	396	271	1,232
C. tofieldiae	CBS 127615	39	37	50	105	4	181	8	1	2	84	6	50	0	0	1	10	392	273	1,232
C. tofieldiae	861	39	37	50	105	4	176	9	1	2	80	6	48	0	0	1	11	395	278	1,230
C. tofieldiae	CBS 168.49	40	36	50	107	4	177	8	1	2	84	6	47	0	0	1	12	400	272	1,234
C. liriopes	A2	38	35	52	113	5	187	9	1	2	79	7	46	0	0	2	14	398	268	1,240
C. incanum	MAFF 238712	30	38	50	101	3	174	9	1	2	69	6	49	0	0	2	10	365	273	1,170
C. incanum	MAFF 238704	33	38	50	101	3	175	8	1	2	72	6	47	0	0	2	10	372	269	1,177
C. higginsianum	IMI 349063	30	37	57	102	5	171	13	1	2	87	5	44	0	0	1	10	431	277	1,262
C. higginsianum	MAFF 305635-RFP	32	35	56	98	5	167	13	1	2	89	5	43	0	0	1	10	421	273	1,240
C. shisoi	JCM 31818	20	28	41	81	2	134	13	1	2	56	7	35	0	0	2	9	345	209	974
C. tanaceti	BRIP 57314	30	33	45	61	2	130	9	2	1	57	4	42	0	0	1	6	325	214	955
C. lentis	CT-30	25	26	44	68	2	138	9	1	1	58	4	38	0	0	1	4	304	218	936
C. coccodes	CcND04	26	36	49	99	6	170	11	1	2	78	7	47	0	0	1	10	389	240	1,161
C. fructicola	1104-6	39	39	76	127	4	223	11	0	1	96	6	48	0	0	4	14	551	375	1,596
C. fructicola	15060	40	39	74	129	4	226	11	1	1	96	6	47	0	0	4	16	549	385	1,608
C. fructicola	Cf413	42	39	70	124	4	225	11	1	1	97	6	48	0	0	4	15	548	383	1,599
C. fructicola	Cf415	36	39	75	126	4	224	12	0	1	94	6	49	0	0	3	17	550	384	1,600
C. fructicola	Cf245	39	39	75	127	4	222	12	0	1	96	6	49	0	0	4	16	548	381	1,599
C. fructicola	MAFF 245243	26	35	68	113	3	216	10	0	0	88	5	45	0	0	4	12	531	349	1,489

Supplementary Table 4 Continued.

Organism	Strain	CAZymes						Protease								Lipase		SSP	Other	Secretome
		CBM	PL	CE	AA	GT	GH	Aspartic	Cysteine	Glutamic	Serine	Mixed	Metallo	Asparagine	Unknown	GGGX-type	GX-type			
C. fructicola	CfS4	37	40	74	124	4	226	11	0	1	96	6	49	0	0	4	13	568	385	1,621
C. fructicola	Cg38S1	37	39	77	128	4	227	11	0	1	95	6	48	0	0	4	14	568	387	1,628
C. fructicola	CGMCC 3.17371	37	39	77	123	6	228	10	0	1	93	7	50	0	0	4	14	569	389	1,629
C. fructicola	HC540	32	39	74	127	4	228	13	0	1	96	5	50	0	0	4	15	564	372	1,605
C. viniferum	CvYL2a	29	32	59	109	7	188	13	1	1	79	6	46	0	0	3	14	513	314	1,397
C. viniferum	CGW01	33	32	57	112	5	195	12	2	1	83	6	44	0	0	4	15	515	337	1,434
C. aenigma	Cg56	39	35	68	114	7	210	11	1	1	95	6	50	0	0	3	14	551	335	1,523
C. musae	CMM 4420	25	25	51	78	3	144	10	0	1	70	4	38	0	0	4	12	461	212	1,122
C. siamense	CAD2	42	34	67	129	4	221	13	1	1	106	5	45	0	0	4	15	543	388	1,599
C. siamense	CAD4	41	34	65	130	4	223	13	1	1	104	5	46	0	0	4	14	543	387	1,597
C. siamense	Cg363	43	38	74	127	8	212	14	1	1	100	5	49	0	0	3	13	563	381	1,616
C. siamense	CsLH17	44	36	74	130	4	208	13	1	1	94	6	49	0	0	3	16	553	385	1,598
C. siamense	COLG-34	31	33	72	115	3	191	11	1	1	89	6	43	0	0	4	15	516	324	1,436
C. siamense	COLG-38	37	35	72	129	4	209	11	1	1	98	6	46	0	0	4	17	545	353	1,547
C. siamense	CAD1	37	39	74	130	4	222	13	1	1	100	5	48	0	0	4	17	562	372	1,608
C. siamense	CAD5	38	39	74	129	4	214	13	1	1	101	6	49	0	0	5	15	542	378	1,589
C. siamense	HBCG01	44	36	71	130	4	216	12	1	1	93	5	54	0	0	4	17	537	370	1,574
C. siamense	ICMP 18578	38	37	77	124	5	220	13	1	1	95	6	48	0	0	4	17	550	378	1,593
C. asianum	ICMP 18580	32	35	69	120	5	214	14	2	1	88	5	50	0	0	2	13	528	356	1,519
C. gloeosporioides	Cg-14	29	39	67	117	1	201	12	0	0	91	6	46	0	0	5	14	541	309	1,459
C. gloeosporioides	COLG-95	30	35	69	119	3	189	13	1	0	90	6	43	0	0	5	14	519	327	1,444
C. gloeosporioides	CgLH19	32	39	75	129	3	212	12	1	1	95	7	47	0	0	5	15	514	349	1,516
C. camelliae	CcLH18	35	38	65	114	4	199	9	1	1	100	7	47	0	0	2	13	503	336	1,459
C. camelliae	LS-19	25	35	72	115	5	200	10	1	0	86	6	41	0	0	3	12	536	330	1,462
C. horii	FJ-1	34	37	71	117	5	200	13	1	1	95	4	50	0	0	4	16	527	359	1,514
C. karsti	CkLH20	34	43	66	122	6	200	11	1	0	91	5	44	0	0	6	18	433	342	1,398
C. sansevieriae	Sa-1-2	15	20	44	52	2	132	5	1	0	52	3	25	0	0	1	11	324	183	858
C. truncatum	KLC.C5	38	44	80	104	3	199	10	5	1	111	9	67	3	4	3	14	656	468	1,802
C. truncatum	KLC.C-4	28	33	56	72	2	157	9	1	0	70	5	41	0	0	2	7	367	250	1,091

Supplementary Table 4 Continued.

Organism	Strain	CAZymes						Protease								Lipase		SSP	Other	Secretome
		CBM	PL	CE	AA	GT	GH	Aspartic	Cysteine	Glutamic	Serine	Mixed	Metallo	Asparagine	Unknown	GGGX-type	GX-type			
C. truncatum	MTCC 3414	37	44	67	105	3	185	10	1	1	86	5	47	0	0	3	12	440	295	1,326
C. truncatum	CMES1059	36	43	62	102	2	192	11	1	1	84	4	45	0	0	2	16	440	301	1,324
C. sidae	CBS 518.97	27	40	64	92	3	175	9	1	2	71	4	43	0	0	0	7	497	267	1,295
C. spinosum	CBS 515.97	28	37	62	93	3	180	9	0	2	76	6	43	0	1	1	5	503	274	1,317
C. trifolii	MAFF 305078	31	37	60	96	5	171	9	2	0	69	4	38	0	0	1	7	486	267	1,275
C. lindemuthianum	89A2	28	27	59	93	4	150	6	0	1	59	6	35	0	0	1	7	429	239	1,136
C. lindemuthianum	83.501	28	27	60	90	5	155	6	0	1	57	7	36	0	0	1	8	443	235	1,150
C. plurivorum	LFN00145	34	38	64	116	6	172	11	0	2	86	4	45	0	0	5	11	509	310	1,397
C. musicola	IMI 507128	33	39	66	115	6	164	12	0	1	80	5	44	0	0	3	12	500	299	1,364
C. sojae	LFN0009	26	38	64	104	6	174	11	0	1	85	4	48	0	0	3	11	501	292	1,354

Supplementary Table 5 Composition of secondary metabolic gene clusters of 102 *Colletotrichum* genomes.

Organism	Strain	phosphonate	resorcinol	siderophore	fungal-RiPP	NRPS-like	T1PKS	transAT-PKS	indole	ectoine	lanthipeptide-class-i	terpene	RiPP-like	LAP	betalactone	NRPS	lanthipeptide-class-ii	NAPAA	arylpolyene	T3PKS	other	Total
C. acutatum	VPRI 42878	0	0	0	0	11	24	0	6	0	0	16	0	0	2	12	0	0	0	1	0	72
C. acutatum	VPRI 42879	0	0	0	0	10	17	1	6	0	0	15	0	0	2	13	0	0	0	1	0	65
C. acutatum	VPRI 42883	0	0	0	0	10	23	0	6	0	0	15	0	0	2	13	0	0	0	1	0	70
C. acutatum	VPRI 43374	0	0	0	0	10	24	0	6	0	0	16	0	0	2	13	0	0	0	1	0	72
C. acutatum	VPRI 43375	0	0	0	0	10	24	0	6	0	0	16	0	0	2	12	0	0	0	1	0	71
C. acutatum	VPRI 43376	0	0	0	0	10	23	0	6	0	0	16	0	0	2	13	0	0	0	1	0	71
C. acutatum	VPRI 43377	0	0	0	0	10	24	0	6	0	0	15	0	0	2	13	0	0	0	1	0	71
C. acutatum	VPRI 43378	0	0	0	0	10	23	0	6	0	0	15	0	0	2	13	0	0	0	1	0	70
C. acutatum	VPRI 43379	0	0	0	0	10	24	0	6	0	0	15	0	0	2	13	0	0	0	1	0	71
C. acutatum	VPRI 43380	0	0	0	0	10	24	0	6	0	0	15	0	0	2	13	0	0	0	1	0	71
C. aenigma	Cg56	0	0	0	0	11	33	0	8	0	0	19	0	0	1	16	0	0	0	2	0	90
C. asianum	ICMP 18580	0	0	0	0	12	37	0	9	0	0	21	0	0	1	17	0	0	0	1	0	98
C. australisinense	CGMCC 3.18886	0	0	0	0	11	22	0	6	0	0	12	0	0	2	12	0	0	0	1	0	66

Supplementary Table 5 Continued.

Organism	**Strain**	**phosphonate**	**resorcinol**	**siderophore**	**fungal-RiPP**	**NRPS-like**	**T1PKS**	**transAT-PKS**	**indole**	**ectoine**	**lanthipeptide-class-i**	**terpene**	**RiPP-like**	**LAP**	**betalactone**	**NRPS**	**lanthipeptide-class-ii**	**NAPAA**	**arylpolyene**	**T3PKS**	**other**	**Total**
C. camelliae	CcLH18	0	0	0	1	11	30	0	8	0	0	19	0	0	1	14	0	0	0	2	1	87
C. camelliae	LS-19	0	0	0	1	10	31	0	8	0	0	20	0	0	1	15	0	0	0	2	1	89
C. coccodes	CcND04	0	0	0	0	11	45	0	8	0	0	11	0	0	0	24	0	0	0	4	0	103
C. eleusines	NJC-16	0	0	0	0	8	26	0	6	0	0	8	0	0	1	12	0	0	0	1	0	62
C. falcatum	CF08	1	1	2	0	8	23	0	7	1	0	9	0	1	2	12	0	1	0	1	0	69
C. falcatum	Cf671	0	0	0	0	6	21	0	7	0	0	8	0	0	0	14	0	0	0	1	0	57
C. fioriniae	HC25	0	0	0	0	9	24	0	6	0	0	12	0	0	2	11	0	0	0	1	0	65
C. fioriniae	HC296	0	0	0	0	9	25	0	6	0	0	12	0	0	2	11	0	0	0	1	0	66
C. fioriniae	HC89	0	0	0	0	10	24	0	6	0	0	12	0	0	2	11	0	0	0	1	0	66
C. fioriniae	HC91	0	0	0	0	10	24	0	6	0	0	12	0	0	2	11	0	0	0	1	0	66
C. fioriniae	IMI 504882	0	0	0	0	10	24	0	6	0	0	11	0	0	2	10	0	0	0	1	0	64
C. fioriniae	KY119	0	0	0	0	9	24	0	6	0	0	12	0	0	2	11	0	0	0	1	0	65
C. fioriniae	KY323	0	0	0	0	9	24	0	6	0	0	12	0	0	2	11	0	0	0	1	0	65
C. fioriniae	KY640	0	0	0	0	10	24	0	6	0	0	12	0	0	2	11	0	0	0	1	0	66
C. fioriniae	KY646	0	0	0	0	10	24	0	6	0	0	12	0	0	2	10	0	0	0	1	0	65
C. fioriniae	VPRI 42872	0	0	0	0	9	24	0	6	0	0	12	0	0	1	10	0	0	0	1	0	63
C. fructicola	1104-6	0	0	0	0	12	35	0	9	0	0	19	0	0	1	21	0	0	0	2	0	99
C. fructicola	15060	0	0	0	0	11	27	0	9	0	0	19	0	0	1	20	0	0	0	2	0	89
C. fructicola	Cf245	0	0	0	0	11	36	0	9	0	0	19	0	0	1	20	0	0	0	2	0	98
C. fructicola	Cf413	0	0	0	0	11	36	0	9	0	0	19	0	0	1	20	0	0	0	2	0	98
C. fructicola	Cf415	0	0	0	0	11	36	0	9	0	0	19	0	0	1	20	0	0	0	2	0	98
C. fructicola	CfS4	0	0	0	0	11	37	0	9	0	0	21	0	0	1	21	0	0	0	2	0	102
C. fructicola	Cg38 S1	0	0	0	0	12	37	0	9	0	0	20	0	0	1	21	0	0	0	2	0	102
C. fructicola	CGMCC 3.17371	0	0	0	0	11	37	0	9	0	0	22	0	0	1	21	0	0	0	2	0	103
C. fructicola	HC540	0	0	0	0	11	36	0	9	0	0	20	0	0	1	21	0	0	0	2	1	101
C. fructicola	MAFF 245243	0	0	0	0	10	37	0	9	0	0	20	0	0	1	21	0	0	0	2	0	100
C. gloeosporioides	Cg-14	0	0	0	1	11	26	0	7	0	0	18	0	0	0	12	0	0	0	2	0	77
C. gloeosporioides	CgLH19	0	0	0	0	11	35	0	7	0	0	18	0	0	1	16	0	0	0	2	0	90
C. gloeosporioides	COLG-95	0	0	0	0	12	37	0	8	0	0	18	0	0	1	16	0	0	0	1	0	93

Supplementary Table 5 Continued.

Organism	**Strain**	**phosphonate**	**resorcinol**	**siderophore**	**fungal-RiPP**	**NRPS-like**	**T1PKS**	**transAT-PKS**	**indole**	**ectoine**	**lanthipeptide-class-i**	**terpene**	**RiPP-like**	**LAP**	**betalactone**	**NRPS**	**lanthipeptide-class-ii**	**NAPAA**	**arylpolyene**	**T3PKS**	**other**	**Total**
C. graminicola	M1.001	0	0	0	0	9	38	0	7	0	0	7	0	0	1	10	0	0	0	1	0	73
C. graminicola	M5.001	0	0	0	0	10	38	0	6	0	0	7	0	0	0	13	0	0	0	1	0	75
C. higginsianum	IMI 349063	0	0	0	2	12	31	0	11	0	0	11	0	0	2	19	0	0	0	1	0	89
C. higginsianum	MAFF 305635-RFP	0	0	0	2	11	36	0	12	0	0	12	0	0	2	16	0	0	0	1	0	92
C. horii	FJ-1	0	0	0	1	12	41	0	8	0	0	19	0	0	1	16	0	0	0	2	1	101
C. incanum	MAFF 238704	0	0	0	0	9	25	0	11	0	0	10	0	0	1	18	0	0	0	1	0	75
C. incanum	MAFF 238712	0	0	0	0	8	26	0	11	0	0	10	0	0	2	15	0	0	0	1	0	73
C. karsti	CkLH20	0	0	0	1	11	23	0	9	0	0	16	0	0	0	9	0	0	0	1	0	70
C. lentis	CT-30	0	0	0	0	7	17	0	6	0	0	9	0	0	1	10	0	0	0	1	0	51
C. lindemuthianum	83.501	0	0	0	1	7	19	0	4	0	0	12	0	0	1	10	0	0	0	1	0	55
C. lindemuthianum	89 A2	0	0	0	1	8	20	0	4	0	0	12	0	0	1	11	0	0	0	1	0	58
C. liriopes	A2	0	0	0	0	12	29	0	9	0	0	8	0	0	3	17	0	0	0	3	0	81
C. musae	CMM 4420	0	0	0	0	9	20	0	9	0	0	13	0	0	0	10	0	0	0	2	1	64
C. musicola	IMI 507128	0	0	0	0	10	32	0	8	0	0	15	0	0	0	15	0	0	0	2	1	83
C. nymphaeae	HC646	0	0	0	0	10	24	0	6	0	0	13	0	0	2	14	0	0	0	1	0	70
C. nymphaeae	IMI 504889	0	0	0	0	10	24	0	7	0	0	13	0	0	2	13	0	0	0	1	0	70
C. nymphaeae	KY563	0	0	0	0	10	24	0	7	0	0	13	0	0	2	13	0	0	0	1	0	70
C. nymphaeae	KY567	0	0	0	0	10	24	0	7	0	0	13	0	0	2	13	0	0	0	1	0	70
C. nymphaeae	KY613	0	0	0	0	9	23	0	6	0	0	14	0	0	2	13	0	0	0	1	0	68
C. nymphaeae	KY745	0	0	0	0	10	23	0	6	0	0	13	0	0	2	13	0	0	0	1	0	68
C. orchidophilum	IMI 309357	0	0	0	0	10	18	0	6	0	0	8	0	0	1	11	0	0	0	1	0	55
C. plurivorum	LFN00145	0	0	0	0	10	27	0	9	0	0	14	0	0	1	10	0	0	0	2	1	74
C. salicis	CBS 607.94	0	0	0	0	7	19	0	6	0	0	10	0	0	1	11	0	0	0	1	0	55
C. sansevieriae	Sa-1-2	0	0	0	0	12	20	0	5	0	0	10	0	0	1	10	0	0	0	2	0	60
C. scovillei	Coll-153	0	0	0	0	10	24	0	6	0	0	15	0	0	2	13	0	0	0	1	0	71
C. scovillei	Coll-365	0	0	0	0	10	24	0	6	0	0	15	0	0	2	13	0	0	0	1	0	71
C. scovillei	Coll-524	0	0	0	0	10	25	0	6	0	0	15	0	0	2	14	0	0	0	1	0	73
C. scovillei	TJNH1	0	0	0	0	10	25	0	6	0	0	14	0	0	2	14	0	0	0	1	0	72
C. shisoi	JCM 31818	0	0	0	0	8	16	0	3	0	0	9	0	0	1	12	0	0	0	1	0	50

Supplementary Table 5 Continued.

Organism	**Strain**	**phosphonate**	**resorcinol**	**siderophore**	**fungal-RiPP**	**NRPS-like**	**T1PKS**	**transAT-PKS**	**indole**	**ectoine**	**lanthipeptide-class-i**	**terpene**	**RiPP-like**	**LAP**	**betalactone**	**NRPS**	**lanthipeptide-class-ii**	**NAPAA**	**arylpolyene**	**T3PKS**	**other**	**Total**
C. siamense	CAD1	0	0	0	0	11	38	0	9	0	0	20	0	0	1	16	0	0	0	2	1	98
C. siamense	CAD2	0	0	0	0	12	39	0	9	0	0	21	0	0	1	17	0	0	0	2	0	101
C. siamense	CAD4	0	0	0	0	12	39	0	9	0	0	21	0	0	1	17	0	0	0	2	0	101
C. siamense	CAD5	0	0	0	1	11	40	0	9	0	0	20	0	0	1	16	0	0	0	2	1	101
C. siamense	Cg363	0	0	0	1	11	38	0	9	0	0	18	0	0	1	16	0	0	0	2	1	97
C. siamense	COLG-34	0	0	0	0	11	30	0	8	0	0	18	0	0	1	17	0	0	0	2	1	88
C. siamense	COLG-38	0	0	0	1	11	30	0	7	0	0	18	0	0	1	17	0	0	0	2	1	88
C. siamense	CsLH17	0	0	0	0	11	39	0	9	0	0	21	0	0	1	16	0	0	0	2	1	100
C. siamense	HBCG01	0	0	0	0	11	39	0	9	0	0	20	0	0	1	16	0	0	0	2	1	99
C. siamense	ICMP 18578	0	0	0	1	11	38	0	9	0	0	17	0	0	1	17	0	0	0	2	0	96
C. sidae	CBS 518.97	0	0	0	1	9	23	0	6	0	0	15	0	0	1	15	0	0	0	1	0	71
C. simmondsii	CBS 122122	0	0	0	0	10	26	0	6	0	0	16	0	0	2	14	0	0	0	1	0	75
C. simmondsii	VPRI 42936	0	0	0	0	10	23	0	6	0	0	15	0	0	2	14	0	0	0	1	0	71
C. sojae	LFN0009	0	0	0	0	10	29	0	9	0	0	15	0	0	0	12	0	0	0	2	1	78
C. spinosum	CBS 515.97	0	0	0	1	9	23	0	6	0	0	13	0	0	1	15	0	0	0	1	0	69
C. sublineola	CgSl1	0	0	0	0	10	34	0	8	0	0	9	0	0	1	20	0	0	0	1	0	83
C. sublineola	CsGL1	0	0	0	0	11	35	0	8	0	0	9	0	0	1	19	0	0	0	1	0	84
C. sublineola	TX430BB	0	0	0	0	9	27	0	6	0	0	7	0	0	1	19	0	0	0	1	0	70
C. tanaceti	BRIP 57314	0	0	0	0	7	21	0	5	0	0	11	0	0	2	13	0	0	0	2	0	61
C. tofieldiae	861	0	0	0	0	8	26	0	11	0	0	9	0	0	2	18	0	0	0	3	0	77
C. tofieldiae	CBS 127615	0	0	0	0	8	25	0	11	0	0	9	0	0	2	18	0	0	0	3	0	76
C. tofieldiae	CBS 130851	0	0	0	0	8	25	0	10	0	0	9	0	0	2	18	0	0	0	3	0	75
C. tofieldiae	CBS 168.49	0	0	0	0	8	25	0	11	0	0	9	0	0	2	16	0	0	0	3	0	74
C. tofieldiae	CBS 495.85	0	0	0	0	8	23	0	11	0	0	9	0	0	2	17	0	0	0	2	0	72
C. trifolii	MAFF 305078	0	0	0	3	10	16	0	5	0	0	13	0	0	1	14	0	0	0	1	0	63
C. truncatum	CMES1059	0	0	0	0	8	41	0	9	0	0	14	0	0	1	21	0	0	0	2	0	96
C. truncatum	KLC.C-4	0	0	0	0	8	31	0	7	0	0	13	0	0	0	19	0	0	0	2	0	80
C. truncatum	KLC.C5	0	0	1	0	11	40	0	9	0	1	15	2	0	1	23	1	0	2	2	0	108
C. truncatum	MTCC 3414	0	0	0	0	10	36	0	9	0	0	15	0	0	0	20	0	0	0	2	0	92

Supplementary Table 5 Continued.

Organism	Strain	phosphonate	resorcinol	siderophore	fungal-RiPP	NRPS-like	T1PKS	transAT-PKS	indole	ectoine	lanthipeptide-class-i	terpene	RiPP-like	LAP	betalactone	NRPS	lanthipeptide-class-ii	NAPAA	arylpolyene	T3PKS	other	Total
C. viniferum	CGW01	0	0	0	0	12	29	0	8	0	0	15	0	0	1	13	0	0	0	2	1	81
C. viniferum	CvYL2a	0	0	0	1	10	29	0	8	0	0	18	0	0	0	15	0	0	0	2	0	83

Supplementary Table 6 GO annotation of 449 significantly changed gene families along the *Colletotrichum* phylogeny.

GO id	GO term	Group	# Involved gene
GO:0055085	transmembrane transport	BP	1
GO:0006725	cellular aromatic compound metabolic process	BP	2
GO:0006629	lipid metabolic process	BP	3
GO:0001522	pseudouridine synthesis	BP	5
GO:0006006	glucose metabolic process	BP	11
GO:0006351	transcription, DNA-templated	BP	12
GO:0019915	lipid storage	BP	12
GO:0007155	cell adhesion	BP	14
GO:0008608	attachment of spindle microtubules to kinetochore	BP	16
GO:0006812	cation transport	BP	18
GO:0009058	biosynthetic process	BP	19
GO:0006508	proteolysis	BP	71
GO:0005975	carbohydrate metabolic process	BP	157
GO:0005737	cytoplasm	CC	1
GO:0005615	extracellular space	CC	1
GO:0005634	nucleus	CC	10
GO:0008290	F-actin capping protein complex	CC	19
GO:0005815	microtubule organizing center	CC	24
GO:0005744	TIM23 mitochondrial import inner membrane translocase complex	CC	25
GO:0005576	extracellular region	CC	184
GO:0016020	membrane	CC	272
GO:0016021	integral component of membrane	CC	607
GO:0004096	catalase activity	MF	1

Supplementary Table 6 Continued.

GO id	GO term	Group	# Involved gene
GO:0004222	metalloendopeptidase activity	MF	1
GO:0004601	peroxidase activity	MF	1
GO:0051118	glucan endo-1,3-alpha-glucosidase activity	MF	1
GO:0016838	carbon-oxygen lyase activity, acting on phosphates	MF	1
GO:0016747	acyltransferase activity, transferring groups other than amino-acyl groups	MF	2
GO:0008270	zinc ion binding	MF	2
GO:0003899	DNA-directed 5'-3' RNA polymerase activity	MF	3
GO:0004843	thiol-dependent deubiquitinase	MF	4
GO:0004386	helicase activity	MF	5
GO:0016788	hydrolase activity, acting on ester bonds	MF	5
GO:0004345	glucose-6-phosphate dehydrogenase activity	MF	6
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	MF	7
GO:0003854	3-beta-hydroxy-delta5-steroid dehydrogenase activity	MF	7
GO:0003779	actin binding	MF	8
GO:0004185	serine-type carboxypeptidase activity	MF	8
GO:0004252	serine-type endopeptidase activity	MF	12
GO:0016846	carbon-sulfur lyase activity	MF	13
GO:0003924	GTPase activity	MF	15
GO:0003723	RNA binding	MF	15
GO:0051537	2 iron, 2 sulfur cluster binding	MF	15
GO:0003676	nucleic acid binding	MF	16
GO:0004637	phosphoribosylamine-glycine ligase activity	MF	16
GO:0008080	N-acetyltransferase activity	MF	18
GO:0008168	methyltransferase activity	MF	20
GO:0008716	D-alanine-D-alanine ligase activity	MF	21
GO:0008239	dipeptidyl-peptidase activity	MF	22
GO:0004061	arylformamidase activity	MF	23
GO:0005525	GTP binding	MF	24
GO:0046983	protein dimerization activity	MF	25
GO:0043531	ADP binding	MF	26
GO:0008194	UDP-glycosyltransferase activity	MF	28
GO:0008061	chitin binding	MF	28
GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	MF	30
GO:0051287	NAD binding	MF	31
GO:0050661	NADP binding	MF	47
GO:0003735	structural constituent of ribosome	MF	49
GO:0000166	nucleotide binding	MF	63
GO:0140359	ABC-type transporter activity	MF	65
GO:0004650	polygalacturonase activity	MF	73
GO:0016787	hydrolase activity	MF	85
GO:0016702	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	MF	96
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	MF	101
GO:0004499	N,N-dimethylaniline monooxygenase activity	MF	102
GO:0000981	DNA-binding transcription factor activity, RNA polymerase II-specific	MF	121
GO:0004672	protein kinase activity	MF	145
GO:0071949	FAD binding	MF	153

Supplementary Table 6 Continued.

GO id	GO term	Group	# Involved gene
GO:0003677	DNA binding	MF	269
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	MF	350
GO:0003824	catalytic activity	MF	534
GO:0005515	protein binding	MF	634
GO:0004497	monooxygenase activity	MF	662
GO:0005524	ATP binding	MF	735
GO:0016491	oxidoreductase activity	MF	951
GO:0022857	transmembrane transporter activity	MF	1,023

Supplementary Table 7 Pfam annotation of 449 significantly changed gene families along the *Colletotrichum* phylogeny.

Pfam accession	Description	# Involved gene
PF00173	Cytochrome b5-like Heme/Steroid binding domain	1
PF00188	Cysteine-rich secretory protein family	1
PF19086	Terpene synthase family 2, C-terminal metal binding	1
PF11954	Domain of unknown function (DUF3471)	1
PF01328	Peroxidase, family 2	1
PF12708	Pectate lyase superfamily protein	1
PF01822	WSC domain	1
PF01693	Caulimovirus viroplasmin	1
PF12171	Zinc-finger double-stranded RNA-binding	1
PF08449	UAA transporter family	1
PF18224	ToxB N-terminal domain	1
PF01040	UbiA prenyltransferase family	1
PF05175	Methyltransferase small domain	1
PF00075	RNase H	1
PF14497	Glutathione S-transferase, C-terminal domain	1
PF16073	Starter unit:ACP transacylase in aflatoxin biosynthesis	1
PF18031	Ubiquitin carboxyl-terminal hydrolases	1
PF07969	Amidohydrolase family	1
PF07287	Acyclic terpene utilisation family protein AtuA	1
PF12146	Serine aminopeptidase, S33	1
PF00702	haloacid dehalogenase-like hydrolase	1
PF00096	Zinc finger, C2H2 type	1
PF01546	Peptidase family M20/M25/M40	1
PF01116	Fructose-bisphosphate aldolase class-II	1
PF00271	Helicase conserved C-terminal domain	1
PF01467	Cytidylyltransferase-like	1
PF01375	Heat-labile enterotoxin alpha chain	1
PF00018	SH3 domain	1
PF12894	Anaphase-promoting complex subunit 4 WD40 domain	1
PF00581	Rhodanese-like domain	1
PF17874	MalT-like TPR region	1
PF00801	PKD domain	1
PF14033	Protein of unknown function (DUF4246)	1
PF01794	Ferric reductase like transmembrane component	1

Supplementary Table 7 Continued.

Pfam accession	Description	# Involved gene
PF02178	AT hook motif	1
PF07693	KAP family P-loop domain	1
PF00199	Catalase	1
PF01088	Ubiquitin carboxyl-terminal hydrolase, family 1	1
PF05899	EutQ-like cupin domain	1
PF13391	HNH endonuclease	1
PF06101	Vacuolar protein sorting-associated protein 62	1
PF04389	Peptidase family M28	1
PF02779	Transketolase, pyrimidine binding domain	1
PF02458	Transferase family	1
PF05128	Domain of unknown function (DUF697)	1
PF13086	AAA domain	2
PF13409	Glutathione S-transferase, N-terminal domain	2
PF05347	Complex 1 protein (LYR family)	2
PF07883	Cupin domain	2
PF00319	SRF-type transcription factor (DNA-binding and dimerisation domain)	2
PF00866	Ring hydroxylating beta subunit	2
PF00533	BRCA1 C Terminus (BRCT) domain	2
PF12678	RING-H2 zinc finger domain	2
PF17106	Sigma domain on NACHT-NTPases	2
PF02843	Phosphoribosylglycinamide synthetase, C domain	2
PF00132	Bacterial transferase hexapeptide (six repeats)	2
PF00471	Ribosomal protein L33	2
PF13483	Beta-lactamase superfamily domain	2
PF13020	Domain of unknown function (DUF3883)	2
PF12224	Putative amidoligase enzyme	2
PF01734	Patatin-like phospholipase	2
PF00682	HMGL-like	2
PF02894	Oxidoreductase family, C-terminal alpha/beta domain	2
PF07367	Fungal fruit body lectin	2
PF00975	Thioesterase domain	2
PF01193	RNA polymerase Rpb3/Rpb11 dimerisation domain	2
PF13646	HEAT repeats	2
PF00756	Putative esterase	2
PF12520	Protein of unknown function (DUF3723)	3
PF10642	Mitochondrial import receptor subunit or translocase	3
PF00385	Chromo (CHRromatin Organisation MOdifier) domain	3
PF07714	Protein tyrosine and serine/threonine kinase	3
PF00176	SNF2-related domain	3
PF03101	FAR1 DNA-binding domain	3
PF07819	PGAP1-like protein	3
PF07143	CrtC N-terminal lipocalin domain	3
PF06609	Fungal trichothecene efflux pump (TRI12)	3
PF06500	Esterase FrsA-like	3
PF01071	Phosphoribosylglycinamide synthetase, ATP-grasp (A) domain	3
PF13193	AMP-binding enzyme C-terminal domain	3
PF13023	HD domain	3

Supplementary Table 7 Continued.

Pfam accession	Description	# Involved gene
PF14831	Domain of unknown function (DUF4484)	4
PF13577	SnoaL-like domain	4
PF00441	Acyl-CoA dehydrogenase, C-terminal domain	4
PF02656	Domain of unknown function (DUF202)	4
PF08570	Protein of unknown function (DUF1761)	4
PF04135	Nucleolar RNA-binding protein, Nop10p family	4
PF05630	Necrosis inducing protein (NPP1)	4
PF14634	zinc-RING finger domain	4
PF13520	Amino acid permease	5
PF03946	Ribosomal protein L11, N-terminal domain	5
PF13671	AAA domain	5
PF02212	Dynamin GTPase effector domain	5
PF02781	Glucose-6-phosphate dehydrogenase, C-terminal domain	5
PF13401	AAA domain	5
PF09458	H-type lectin domain	5
PF05792	Candida agglutinin-like (ALS)	5
PF13606	Ankyrin repeat	5
PF12698	ABC-2 family transporter protein	5
PF09804	DENN domain-containing protein 11	6
PF04419	4F5 protein related disordered region	6
PF00241	Cofilin/tropomyosin-type actin-binding protein	6
PF00240	Ubiquitin family	6
PF02844	Phosphoribosylglycinamide synthetase, N domain	6
PF03221	Tc5 transposase DNA-binding domain	6
PF05033	Pre-SET motif	6
PF00389	D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain	6
PF00856	SET domain	6
PF04828	Glutathione-dependent formaldehyde-activating enzyme	7
PF03702	Anhydro-N-acetylmuramic acid kinase	7
PF00708	Acylphosphatase	7
PF13191	AAA ATPase domain	7
PF00298	Ribosomal protein L11, RNA binding domain	7
PF00378	Enoyl-CoA hydratase/isomerase	7
PF01073	3-beta hydroxysteroid dehydrogenase/isomerase family	7
PF02798	Glutathione S-transferase, N-terminal domain	7
PF13649	Methyltransferase domain	7
PF03911	Sec61beta family	8
PF13656	RNA polymerase Rpb3/Rpb11 dimerisation domain	8
PF00135	Carboxylesterase family	8
PF00450	Serine carboxypeptidase	8
PF00294	pfkB family carbohydrate kinase	8
PF00206	Lyase	8
PF16455	Ubiquitin-binding domain	9
PF01694	Rhomboid family	9
PF02771	Acyl-CoA dehydrogenase, N-terminal domain	9
PF13847	Methyltransferase domain	9
PF06775	Putative adipose-regulatory protein (Seipin)	9

Supplementary Table 7 Continued.

Pfam accession	**Description**	**# Involved gene**
PF13639	Ring finger domain	9
PF00646	F-box domain	9
PF05057	Putative serine esterase (DUF676)	9
PF17107	N-terminal domain on NACHT_NTPase and P-loop NTPases	9
PF16508	Second BRCT domain on Nijmegen syndrome breakage protein	9
PF02770	Acyl-CoA dehydrogenase, middle domain	9
PF07992	Pyridine nucleotide-disulphide oxidoreductase	9
PF00155	Aminotransferase class I and II	10
PF01184	GPR1/FUN34/yaaH family	10
PF12937	F-box-like	10
PF13535	ATP-grasp domain	10
PF08520	Fungal protein of unknown function (DUF1748)	10
PF13602	Zinc-binding dehydrogenase	10
PF08592	Domain of unknown function (DUF1772)	10
PF14833	NAD-binding of NADP-dependent 3-hydroxyisobutyrate dehydrogenase	11
PF00857	Isochorismatase family	11
PF01593	Flavin containing amine oxidoreductase	11
PF01423	LSM domain	11
PF07985	SRR1	11
PF02597	ThiS family	11
PF00190	Cupin	11
PF00479	Glucose-6-phosphate dehydrogenase, NAD binding domain	11
PF16155	Domain of unknown function (DUF4863)	12
PF10397	Adenylosuccinate lyase C-terminus	12
PF12697	Alpha/beta hydrolase family	12
PF13472	GDSL-like Lipase/Acylhydrolase family	12
PF11917	Protein of unknown function (DUF3435)	12
PF12680	SnoaL-like domain	12
PF13857	Ankyrin repeats (many copies)	12
PF00076	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	13
PF01738	Dienelactone hydrolase family	13
PF08883	Dopa 4,5-dioxygenase family	13
PF01926	50S ribosome-binding GTPase	13
PF00355	Rieske [2Fe-2S] domain	13
PF02133	Permease for cytosine/purines, uracil, thiamine, allantoin	14
PF08650	DASH complex subunit Dad4	14
PF01267	F-actin capping protein alpha subunit	14
PF13302	Acetyltransferase (GNAT) domain	14
PF00098	Zinc knuckle	14
PF18130	ATP-grasp N-terminal domain	14
PF02386	Cation transport protein	15
PF00293	NUDIX domain	15
PF13087	AAA domain	15
PF13238	AAA domain	15
PF10551	MULE transposase domain	15
PF02826	D-isomer specific 2-hydroxyacid dehydrogenase, NAD binding domain	16

Supplementary Table 7 Continued.

Pfam accession	**Description**	**# Involved gene**
PF11702	Protein of unknown function (DUF3295)	16
PF03981	Ubiquinol-cytochrome C chaperone	16
PF09994	Uncharacterized alpha/beta hydrolase domain (DUF2235)	16
PF00931	NB-ARC domain	16
PF07993	Male sterility protein	17
PF13374	Tetratricopeptide repeat	17
PF13376	Bacteriocin-protection, YdeI or OmpD-Associated	17
PF00043	Glutathione S-transferase, C-terminal domain	17
PF13517	FG-GAP-like repeat	17
PF03572	Peptidase family S41	17
PF03061	Thioesterase superfamily	17
PF00107	Zinc-binding dehydrogenase	17
PF07478	D-ala D-ala ligase C-terminus	17
PF17390	Bacterial alpha-L-rhamnosidase C-terminal domain	18
PF00023	Ankyrin repeat	18
PF08550	Fungal protein of unknown function (DUF1752)	18
PF05699	hAT family C-terminal dimerisation region	18
PF01063	Amino-transferase class IV	18
PF17186	Lipocalin-like domain	18
PF00350	Dynamin family	19
PF13417	Glutathione S-transferase, N-terminal domain	19
PF14226	non-haem dioxygenase in morphine synthesis N-terminal	19
PF13878	zinc-finger of acetyl-transferase ESCO	19
PF08240	Alcohol dehydrogenase GroES-like domain	19
PF08534	Redoxin	19
PF10237	Probable N6-adenine methyltransferase	19
PF05686	Glycosyl transferase family 90	19
PF08530	X-Pro dipeptidyl-peptidase C-terminal non-catalytic domain	19
PF08531	Alpha-L-rhamnosidase N-terminal domain	20
PF08294	TIM21	20
PF08031	Berberine and berberine like	20
PF04199	Putative cyclase	20
PF00118	TCP-1/cpn60 chaperonin family	21
PF08719	NADAR domain	21
PF01284	Membrane-associating domain	21
PF13637	Ankyrin repeats (many copies)	21
PF07989	Centrosomin N-terminal motif 1	21
PF03443	Auxiliary Activity family 9 (formerly GH61)	22
PF00253	Ribosomal protein S14p/S29e	22
PF02129	X-Pro dipeptidyl-peptidase (S15 family)	22
PF17667	Fungal protein kinase	22
PF00201	UDP-glucoronosyl and UDP-glucosyl transferase	23
PF04248	Domain of unknown function (DUF427)	23
PF11915	Protein of unknown function (DUF3433)	23
PF14856	Pathogen effector; putative necrosis-inducing factor	24
PF20150	2EXR family	24
PF14388	Domain of unknown function (DUF4419)	24

Supplementary Table 7 Continued.

Pfam accession	Description	# Involved gene
PF03881	Fructosamine kinase	24
PF01031	Dynamin central region	24
PF00698	Acyl transferase domain	25
PF13424	Tetratricopeptide repeat	26
PF09801	Integral membrane protein S linking to the trans Golgi network	26
PF07876	Stress responsive A/B Barrel Domain	26
PF17389	Bacterial alpha-L-rhamnosidase 6 hairpin glycosidase domain	26
PF00734	Fungal cellulose binding domain	27
PF08242	Methyltransferase domain	27
PF00651	BTB/POZ domain	27
PF00187	Chitin recognition protein	27
PF00383	Cytidine and deoxycytidylate deaminase zinc-binding region	28
PF00544	Pectate lyase	28
PF17100	N-terminal domain of NWD NACHT-NTPase	28
PF12706	Beta-lactamase superfamily domain	29
PF05592	Bacterial alpha-L-rhamnosidase concanavalin-like domain	29
PF01544	CorA-like Mg2+ transporter protein	29
PF00668	Condensation domain	29
PF00291	Pyridoxal-phosphate dependent enzyme	31
PF08659	KR domain	32
PF17784	Sulfotransferase domain	32
PF10528	GLEYA domain	33
PF00153	Mitochondrial carrier protein	33
PF17111	Fungal N-terminal domain of STAND proteins	34
PF16197	Ketoacyl-synthetase C-terminal extension	34
PF00149	Calcineurin-like phosphoesterase	34
PF00109	Beta-ketoacyl synthase, N-terminal domain	35
PF00082	Subtilase family	36
PF03446	NAD binding domain of 6-phosphogluconate dehydrogenase	36
PF00498	FHA domain	36
PF11374	Protein of unknown function (DUF3176)	36
PF10373	Est1 DNA/RNA binding domain	37
PF14765	Polyketide synthase dehydratase	38
PF18271	Glycoside hydrolase 131 catalytic N-terminal domain	38
PF12417	Zinc finger protein	39
PF05199	GMC oxidoreductase	40
PF14269	Arylsulfotransferase (ASST)	41
PF00400	WD domain, G-beta repeat	42
PF03171	2OG-Fe(II) oxygenase superfamily	43
PF00732	GMC oxidoreductase	44
PF02801	Beta-ketoacyl synthase, C-terminal domain	46
PF05721	Phytanoyl-CoA dioxygenase (PhyH)	46
PF00550	Phosphopantetheine attachment site	47
PF13854	Kelch motif	49
PF19055	ABC-2 type transporter	49
PF04616	Glycosyl hydrolases family 43	50
PF11951	Fungal specific transcription factor domain	50

Supplementary Table 7 Continued.

Pfam accession	**Description**	**# Involved gene**
PF14510	ABC-transporter N-terminal	50
PF00004	ATPase family associated with various cellular activities (AAA)	52
PF00501	AMP-binding enzyme	52
PF01408	Oxidoreductase family, NAD-binding Rossmann fold	52
PF14479	Prion-inhibition and propagation	54
PF01042	Endoribonuclease L-PSP	54
PF00704	Glycosyl hydrolases family 18	55
PF05730	CFEM domain	56
PF17851	Beta xylosidase C-terminal Concanavalin A-like domain	57
PF00295	Glycosyl hydrolases family 28	59
PF03033	Glycosyltransferase family 28 N-terminal domain	59
PF01636	Phosphotransferase enzyme family	62
PF00933	Glycosyl hydrolase family 3 N terminal domain	62
PF01915	Glycosyl hydrolase family 3 C-terminal domain	63
PF14310	Fibronectin type III-like domain	66
PF06422	CDR ABC transporter	67
PF11905	Domain of unknown function (DUF3425)	68
PF01061	ABC-2 type transporter	69
PF00753	Metallo-beta-lactamase superfamily	71
PF13450	NAD(P)-binding Rossmann-like domain	72
PF00324	Amino acid permease	72
PF03055	Retinal pigment epithelial membrane protein	73
PF07519	Tannase and feruloyl esterase	74
PF01425	Amidase	77
PF00743	Flavin-binding monooxygenase-like	78
PF01370	NAD dependent epimerase/dehydratase family	80
PF00248	Aldo/keto reductase family	83
PF03211	Pectate lyase	84
PF13460	NAD(P)H-binding	90
PF00171	Aldehyde dehydrogenase family	101
PF00172	Fungal Zn(2)-Cys(6) binuclear cluster domain	102
PF01476	LysM domain	105
PF04479	RTA1 like protein	113
PF00069	Protein kinase domain	118
PF00664	ABC transporter transmembrane region	121
PF01494	FAD binding domain	126
PF01048	Phosphorylase superfamily	134
PF13561	Enoyl-(Acyl carrier protein) reductase	139
PF01565	FAD binding domain	153
PF12796	Ankyrin repeats (3 copies)	161
PF05368	NmrA-like family	200
PF00264	Common central domain of tyrosinase	209
PF04082	Fungal specific transcription factor domain	210
PF00106	short chain dehydrogenase	239
PF05729	NACHT domain	271
PF00005	ABC transporter	299
PF06985	Heterokaryon incompatibility protein (HET)	362

Supplementary Table 7 Continued.

Pfam accession	Description	# Involved gene
PF00083	Sugar (and other) transporter	385
PF00067	Cytochrome P450	536
PF13489	Methyltransferase domain	700
PF07690	Major Facilitator Superfamily	833

Supplementary Table 8 GO annotation of 423 significantly expanded gene families in *Colletotrichum* lineage.

GO ID	GO term	Group	# Involved gene
GO:0006725	cellular aromatic compound metabolic process	BP	1
GO:0000462	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	BP	2
GO:0006397	mRNA processing	BP	3
GO:0006886	intracellular protein transport	BP	10
GO:0051321	meiotic cell cycle	BP	14
GO:0008272	sulfate transport	BP	15
GO:0006807	nitrogen compound metabolic process	BP	16
GO:0006694	steroid biosynthetic process	BP	17
GO:0008654	phospholipid biosynthetic process	BP	17
GO:0006629	lipid metabolic process	BP	21
GO:0006364	rRNA processing	BP	24
GO:0006812	cation transport	BP	24
GO:0016192	vesicle-mediated transport	BP	28
GO:0006508	proteolysis	BP	69
GO:0005975	carbohydrate metabolic process	BP	285
GO:0017119	Golgi transport complex	CC	12
GO:0005634	nucleus	CC	15
GO:0005815	microtubule organizing center	CC	15
GO:0005789	endoplasmic reticulum membrane	CC	16
GO:0005680	anaphase-promoting complex	CC	21
GO:0016020	membrane	CC	377
GO:0005576	extracellular region	CC	395
GO:0016021	integral component of membrane	CC	748
GO:0008757	S-adenosylmethionine-dependent methyltransferase activity	MF	2
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	MF	2
GO:0097367	carbohydrate derivative binding	MF	2
GO:0003877	ATP adenylyltransferase activity	MF	4
GO:0008194	UDP-glycosyltransferase activity	MF	4
GO:0005525	GTP binding	MF	5
GO:0008061	chitin binding	MF	5
GO:0003854	3-beta-hydroxy-delta5-steroid dehydrogenase activity	MF	6
GO:0003735	structural constituent of ribosome	MF	7
GO:0030246	carbohydrate binding	MF	7
GO:0003676	nucleic acid binding	MF	10
GO:0003885	D-arabinono-1,4-lactone oxidase activity	MF	13
GO:0000062	fatty-acyl-CoA binding	MF	14
GO:0004096	catalase activity	MF	14

Supplementary Table 8 Continued.

GO ID	GO term	Group	# Involved gene
GO:0004359	glutaminase activity	MF	14
GO:0004565	beta-galactosidase activity	MF	16
GO:0000030	mannosyltransferase activity	MF	17
GO:0003700	DNA-binding transcription factor activity	MF	17
GO:0004061	arylformamidase activity	MF	17
GO:0043531	ADP binding	MF	17
GO:0004334	fumarylacetoacetase activity	MF	18
GO:0008239	dipeptidyl-peptidase activity	MF	20
GO:0010181	FMN binding	MF	20
GO:0015109	chromate transmembrane transporter activity	MF	20
GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	MF	21
GO:0008171	O-methyltransferase activity	MF	22
GO:0016757	glycosyltransferase activity	MF	23
GO:0035673	oligopeptide transmembrane transporter activity	MF	23
GO:0004559	alpha-mannosidase activity	MF	25
GO:0140359	ABC-type transporter activity	MF	26
GO:0000049	tRNA binding	MF	27
GO:0016836	hydro-lyase activity	MF	29
GO:0004177	aminopeptidase activity	MF	30
GO:0004181	metallocarboxypeptidase activity	MF	30
GO:0016791	phosphatase activity	MF	31
GO:0004474	malate synthase activity	MF	35
GO:0008168	methyltransferase activity	MF	36
GO:0051118	glucan endo-1,3-alpha-glucosidase activity	MF	38
GO:0004411	homogentisate 1,2-dioxygenase activity	MF	39
GO:0005506	iron ion binding	MF	43
GO:0050661	NADP binding	MF	45
GO:0004185	serine-type carboxypeptidase activity	MF	48
GO:0003743	translation initiation factor activity	MF	53
GO:0004571	mannosyl-oligosaccharide 1,2-alpha-mannosidase activity	MF	53
GO:0004601	peroxidase activity	MF	55
GO:0000166	nucleotide binding	MF	60
GO:0004222	metalloendopeptidase activity	MF	60
GO:0005507	copper ion binding	MF	62
GO:0016702	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	MF	72
GO:0003924	GTPase activity	MF	75
GO:0051287	NAD binding	MF	77
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	MF	78
GO:0008080	N-acetyltransferase activity	MF	95
GO:0004672	protein kinase activity	MF	124
GO:0000981	DNA-binding transcription factor activity, RNA polymerase II-specific	MF	140
GO:0004650	polygalacturonase activity	MF	148
GO:0004499	N,N-dimethylaniline monooxygenase activity	MF	160
GO:0016787	hydrolase activity	MF	164
GO:0071949	FAD binding	MF	183
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	MF	220

Supplementary Table 8 Continued.

GO ID	GO term	Group	# Involved gene
GO:0003677	DNA binding	MF	228
GO:0005515	protein binding	MF	301
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	MF	512
GO:0005524	ATP binding	MF	532
GO:0003824	catalytic activity	MF	536
GO:0004497	monooxygenase activity	MF	578
GO:0016491	oxidoreductase activity	MF	1,299
GO:0022857	transmembrane transporter activity	MF	1,436