Check for updates

Genomic insights and comparative analysis of *Colletotrichum* species associated with anthracnose fruit rot and crown rot of strawberry in North Carolina

Tika B. Adhikari[1]*, Norman Muzhinji[2]*, Ashley N. Philbrick[1] and Frank J. Louws[1,3]

[1]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, United States, [2]Department of Plant Sciences, Plant Pathology Division, University of the Free State, Bloemfontein, South Africa, [3]Department of Horticultural Science, North Carolina State University, Raleigh, NC, United States

Colletotrichum is a large genus of fungal phytopathogens responsible for significant economic losses in numerous crops globally. These pathogens exhibit varying host specificities; some have a broad host range, while others are more limited. To explore the genetic composition and underlying factors of fungal virulence and pathogenicity, we sequenced the genomes of seven isolates of *Colletotrichum* spp.: three from the *C. acutatum* and four from the *C. gloeosporioides*. These isolates were sourced from anthracnose fruit rot and crown rot of strawberry in North Carolina. Phylogenetic and phylogenomic analyses classified the isolates within the *C. acutatum* as *C. nymphaeae*, while those in the *C. gloeosporioides* were identified as *C. siamense*. The genome sizes of the *C. nymphaeae* isolates ranged from 50.3 Mb to 50.7 Mb, with 14,235 to 14,260 predicted protein-coding gene models. In contrast, the genome sizes of the *C. siamense* isolates ranged from 55.7 Mb to 58.6 Mb, with predicted protein-coding gene models ranging from 17,420 to 17,729. The GC content across all genomes spanned from 51.9 to 53.7%. The predicted gene models included effectors (339 to 480), secondary metabolic gene clusters (67 to 90), and carbohydrate-active enzymes (800 to 1,060), with *C. siamense* isolates exhibiting the highest numbers in these categories. The genomic resources from this study will aid in resolving taxonomic challenges associated with *Colletotrichum* spp., elucidate their evolutionary history, and enhance the understanding of fungal biology and ecology, which is crucial for developing effective disease management strategies.

KEYWORDS

anthracnose, crown rot, *Colletotrichum*, whole genome sequencing, comparative genomics, fruit rot, strawberry

Introduction

Strawberry (*Fragaria* × *ananassa* Duch.) is a widely grown small fruit crop, contributing significantly to economic and nutritional aspects. The strawberry industry has experienced notable growth in the United States, establishing itself as a vital element of fruit production (Samtani et al., 2019). However, this expansion faces threats from anthracnose disease caused

by *Colletotrichum* spp. (Crouch et al., 2014). The *Colletotrichum* genus encompasses numerous important plant pathogens, endophytes, and saprobes (Jayawardena et al., 2016), and is categorized into species complexes comprising closely related yet distinct fungal species (Crouch et al., 2014). These species are globally distributed and found on various plants in both the tropical and temperate regions. Notably, species within the *C. gloeosporioides* (Penz.) Penz. & Sacc and *C. acutatum* J. H. Simmonds complexes are responsible for economically significant anthracnose fruit rot (AFR) and crown rot (ACR) diseases in strawberry, respectively (Ji et al., 2022). AFR is characterized by sunken black spots and necrotic lesions on petioles, stolons, and fruit, along with abundant acervuli conidia while crown rot manifests as reddish-brown necrotic areas in the crown (Brooks, 1931).

Identifying *Colletotrichum* species can be challenging due to the similarities in their morphological characteristics, which necessitate expert knowledge and experience (Khodadadi et al., 2020). Advances in molecular techniques, phylogenomic, and whole-genome sequencing have improved the classification and understanding of *Colletotrichum* as a species complex with a shared evolutionary ancestry. Multi-locus sequence analyses classify the genus into 280 species, grouped into 16 main species complexes: *C. acutatum*, *C. boninense*, *C. caudatum*, *C. dematium*, *C. agaves*, *C. destructivum*, *C. dracaenophilum*, *C. gigasporum*, *C. graminicola*, *C. gloeosporioides*, *C. magnum*, *C. orbiculare*, *C. orchidearum*, *C. spaethianum*, *C. truncatum* and *C. bambuscola* (Jayawardena et al., 2016; Cannon et al., 2012; Jayawardena et al., 2021; Chen et al., 2022; Liu et al., 2022). Additionally, 15 singletons have been reported. Within the *C. gloeosporioides* complex (Cg complex), 57 species have been described (Jayawardena et al., 2016; Weir et al., 2012; Talhinhas and Baroncelli, 2021), four of which *C. changpingense*, *C. fructicola*, *C. siamense*, and *C. theobromicola* (syn. *C. fragariae*); have been reported to cause strawberry diseases. In the *C. acutatum* complex (Ca complex), 41 species have been described, with *C. fioriniae*, *C. nymphaeae*, *C. godetiae*, *C. salicis*, and *C. simmondsii* associated with strawberry diseases. Additionally, *C. lineola* from the *C. dematium* species complex and singleton *C. nigrum* have also been reported to be associated with strawberry diseases (Jayawardena et al., 2016; Baroncelli et al., 2015; Wang et al., 2019).

Research on *Colletotrichum* spp. has mainly concentrated on their taxonomy by analyzing multigene families (Khodadadi et al., 2020). With the growing number of genome-sequenced species, numerous essential genes and virulence factors related to host infection and disease development have been identified (Baroncelli et al., 2015). Currently, the NCBI database contains 313 genomes of *Colletotrichum* spp. from various host plants. These whole-genome sequences enhance our understanding of plant-pathogen interactions and the dual lifestyles -pathogen and endophyte - exhibited by *Colletotrichum* (Gan et al., 2019). However, complete genomes have only been sequenced from a limited number of isolates from small fruits, such as *C. fiorinea* and *C. fructicola*, (Baroncelli et al., 2015; Gan et al., 2019; Armitage et al., 2020). This limited genomic representation of *Colletotrichum* from small fruit species hampers comparative genomic analyses both within and across species complexes. Despite this limitation, access to whole-genome sequences is crucial for elucidating the genomic architecture and pathogenicity gene repertoire of *Colletotrichum* spp. in strawberry, significantly enhancing our understanding of their taxonomy, virulence, and pathogenicity. Furthermore, the study of

gene clusters, involved in secondary metabolism, and genome arrangements contributes to insights into fungal genome evolution and ecological interactions between *Colletotrichum* spp. and their plant hosts.

Anthracnose fruit rot (AFR) and anthracnose crown rot (ACR) are the most devastating diseases of strawberry in North Carolina, caused by *C. acutatum* and *C. gloeosporioides*, respectively (Rahman et al., 2015; Rahman et al., 2019). Multilocus sequence analysis (MLSA) classified these fungi into two species complexes: Acutatum and Gloeosporioides (Weir et al., 2012; Baroncelli et al., 2015; Damm et al., 2019). This study aimed to characterize the genome content of seven *Colletotrichum* isolates collected from diseased strawberry fields in North Carolina. These genome resources for *Colletotrichum* isolates from strawberry, which are underrepresented in databases, will facilitate comparative genomic studies on host-pathogen interactions and factors influencing host disease resistance in the host. Furthermore, the findings will aid in developing species-specific diagnostic assays, addressing challenges posed by species complexes.

Materials and methods

Fungal isolations

Seven isolates of *Colletotrichum* were collected from strawberry in North Carolina between 2009 and 2010: three from fruit rot (6Ca, 34Ca, and 40Ca), and four from crown rot (28Cg, 57Cg, 58Cg, and 84Cg). The procedure for isolating these fungi has been outlined in our previous studies (Adhikari et al., 2019; Jacobs et al., 2020). Furthermore, information about the fungal isolation process, the tissue of origin (such as leaf, fruit or crown), identification methods, and the locations where the isolates were collected can be found in our previous publication (Jacobs et al., 2020). Briefly, small pieces of infected fruit and crown rot from strawberry seedlings were washed in sterile water and surface-sterilized with 1% sodium hypochlorite solution to obtain pure cultures. The sterilized pieces were then placed on potato dextrose agar (PDA) and incubated at 25°C for 7 days, using a single spore method to obtain pure cultures. Morphological characteristics and pathogenicity tests were assessed (Gunnell and Gubler, 1992; Smith and Black, 1990). Identification was performed using TaqMan real-time polymerase chain reaction (TaqMan PCR) assays as described previously (Garrido et al., 2009). Some of these isolates play a significant role in strawberry breeding program (Jacobs et al., 2019; Jacobs et al., 2020). Fungal plugs were preserved at −80°C for future use.

DNA extraction and sequencing

Mycelium was scraped from pure cultures of *Colletotrichum* isolates. Total genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Germantown, MD, United States) according to the manufacturer's recommended protocol. The constructed libraries were sequenced on a NovaSeq System (Illumina, USA), producing paired-end reads of 2 × 151 bp. The quality of the sequenced libraries was assessed using FastQC v0.11.7 (Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom). Reads were filtered and trimmed with Fastp (Chen et al., 2018), applying a minimum Phred score of 30.

Genome assembly and mapping

The genomes were trimmed and assembled using SPAdes v3.15.5 (Bankevich et al., 2012). Contigs shorter than 200 bp were filtered using SeqKit v0.10.1 (Shen et al., 2016). The filtered reads were then subjected to scaffolding using the Ragtag pipeline (Alonge et al., 2022), with *Colletotrichum siamense* HNO8[1] as the reference genome. The Burrow-Wheeler Aligner (BWA -v0.7.18 and SAMtools; Li et al., 2009) were then used to sort and index the BAM files. Genome statistics were determined using QUAST v5.0.2 (Gurevich et al., 2013). To evaluate genome completeness, BUSCO analyses were performed using the fungal_odb10 and ascomycota_odb10 lineages (Manni et al., 2021).

Phylogenetic and phylogenomic analysis

To identify the sequenced isolates, we conducted a phylogenetic analysis using five DNA barcoding regions: the 5.8S nuclear ribosomal gene along with its two flanking internal transcribed spacers (ITS), beta-tubulin (TUB2), an intron of glyceraldehyde-3-phosphate dehydrogenase (GAPDH), a partial sequence of the actin gene (ACT), and chitin synthase 1 (CHS1). We extracted the sequences of these genes from the genomes of each of the seven assemblies of *Colletotrichum* isolates. Additionally, we downloaded reference sequences for these gene regions from contemporary phylogenetic studies of *Colletotrichum* (Jayawardena et al., 2021; Liu et al., 2022; Zhang et al., 2023) from NCBI GenBank. The accession numbers for the sequences and the names of the species complexes were also obtained from NCBI GenBank (Supplementary Table S1).

We utilized OrthoVenn 3 (Sun et al., 2023) to identify orthologous protein groups for analyzing protein sequences from 28 species of *Colletotrichum*, which we downloaded from JGI Mycocosm and NCBI GenBank (Supplementary Table S2). We used *Verticillium dahliae* VdLs.17 as an outgroup (Klosterman et al., 2011). The orthologous proteins were clustered, and a species tree was generated for the *Colletotrichum* isolates. Our analysis focused on amino acid sequences corresponding to 9,137 single-copy core genes, which were extracted from the output files of OrthoVenn 3. We concatenated and aligned both trees' DNA and protein sequences using MAFFT v.7.215 (Katoh et al., 2002), followed by manual refinement in MEGA v.11 (Tamura et al., 2021). Maximum-likelihood species phylogeny was inferred from multigene families with RA × ML v.8.2.11, including 1,000 bootstrap replicates, while MEGA v.11 was employed for analyzing multigene families. The resulting trees were visualized using iTOL version 5 (Letunic and Bork, 2021).

Genome alignments and synteny

To assess the alignments and collinearity between genome sequences, we used D-GENIES (Cabanettes and Klopp, 2018) using Minimap2 version 5 (Li, 2018) for alignments. The MCScanX toolkit (Wang et al., 2012) was used to create chromosomal "blocks" that illustrate conserved gene orders (synteny) based on collinearity relationships. For the BlastP analysis, we used eight CPU threads, set an E-value threshold of 1×10^{-10}, and required a minimum of five Blast hits. Finally, we visualized the synteny based on the collinear blocks using SynVisio (Bandi and Gutwin, 2020).

Identification of repetitive elements

Before gene prediction, the repetitive elements in the genome were identified using homology-based and ab initio methods. The ab initio method, RepeatModeler v.1.0.11,[2] was used to predict repetitive sequences across all seven genomes. The predicted repetitive sequences were combined with RepeatMasker v.4.0.9,[3] identifying repeats that matched entries in the Dfam_3.7 and RepBase-20181026 repeat library databases. Furthermore, a Tandem Repeat Finder (Maddi et al., 2022) was employed to identify repeated regions.

Genome annotation

Gene prediction for the genomes of seven isolates of *Colletotrichum* spp. was conducted using the Funannotate v1.8.17 pipeline, following the primary assembly preparation protocol.[4] This process involved cleaning repetitive contigs, sorting the assembly by length, applying repeat masking, and gene prediction. Gene models were generated using AUGUSTUS, an ab initio gene predictor, alongside all fungal proteins available in the SwissProt database, which were manually curated as of March 18, 2024. These proteins were further categorized using InterProScan v5.66–98.0 (Jones et al., 2014), and KEGG (Kyoto Encyclopedia of Genes and Genome; Kanehisa et al., 2016). The ribosomal RNA subunits were analyzed with RNAmmer v1.2 (Lagesen et al., 2007), and tRNA genes were identified utilizing tRNAscan-SE v14 with the RNAweasel interface (Lowe and Eddy, 1997).[5]

Predictions of genes encoding potential secreted proteins were conducted using a combination of SignalP v6, TargetP v1.1, and TMHMM v2.0, integrated into a deep neural network-based software SECRETOOL.[6] The secretome was defined as the aggregated results of SignalP and TargetP, filtered based on TMHMM to include proteins with zero or one transmembrane domain. Additionally, EffectorP v3.0 (Sperschneider and Dodds, 2022) was used to identify potential effector proteins within the predicted secretome. Effector candidate were classified as predicted secreted proteins containing a signal peptide but lacking transmembrane domains or glycosylphosphatidylinositol anchors, and limited to a maximum of 300 amino acids. BLASTP analysis was performed against the Swiss-Prot database with a cutoff E-value of 10^{-5} to assess the similarity of effector candidates to known proteins.

To identify gene clusters associated with secondary metabolite production, we employed the antiSMASH (Antibiotics and Secondary Metabolite Analysis) tool (Blin et al., 2023), focusing exclusively oncore biosynthetic genes (SMKGs). Carbohydrate-active enzymes

---

1 https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_038023885.1/

2 http://www.repeatmasker.org

3 www.repeatmasker.org

4 https://github.com/nextgenusfs/funannotate, accessed April 14, 2024

5 https://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/

6 http://genomes.cicbiogune.es/SECRETOOL

(CAZymes) annotated using the dbCAN2 server (Zheng et al., 2023), which utilized HMMER, DIAMOND, and eCAMI (Yin et al., 2012). Genes were designated as CAZymes when at least two of the three tools provided a positive annotation. To reduce false positives, the identified CAZymes were analyzed using InterProScan 5 with an E-value cutoff of 1e-10, and the member counts for each CAZyme family were adjusted accordingly. Furthermore, the predicted proteins were compared with the Pathogen-Host Interaction (PHI) database, Version 4.7[7] (May 2024), using BLASTP with an E-value cutoff of 1e-5.

Results

The genome sizes of seven *Colletotrichum* spp. isolates ranged from 50.3 Mb for isolate 40Ca to 58.6 Mb for isolate 58Cg (Table 1). The *C. siamense* isolates had larger genome, ranging from 56.1 to 58.6 Mb, while the *C. nymphaeae* isolates had smaller genomes ranging from 50.3 to 50.7 Mb. The contig N50 lengths varied from 167,900 kb to 202,057 kb, and the largest contig sizes ranged from 668,652 kb for isolate 40Ca to 667,855 kb for isolate 28Ca (Table 1). The GC content of the sequenced genomes was between 51.9 and 53.7%, with isolates 6Ca and 84Cg representing the lower and upper extremes, respectively. All genomes exhibited high completeness levels, ranging from 98.2 to 99.1% (Table 1). Isolates of *C. nymphaeae* from the *C. acutatum* showed smaller genome assemblies compared to those of *C. siamense* from the *C. gloeosporioides*. The annotated genome sequences of the *Colletotrichum* isolates have been deposited in GenBank under BioProject PRJNA1090952 and BioSample accession numbers SAMN40582474 and SAMN40582476-8 (Table 1).

Phylogenetic and phylogenomic analyses, using the combined data set of five gene regions, revealed that isolates in the Acutatum complex grouped and clustered with *C. nymphaeae* with a bootstrap support of more than 95% (Supplementary Figure S1). All four isolates in the Gloeosporioides complex formed a sister relationship with *C. siamense* with bootstrap support of more than 96% (Figure 1; Supplementary Figure S1). It is worth noting that the tree topologies of both the multigene family and genomics are identical.

The proportion of repetitive elements in the sequenced *Colletotrichum* genomes did not vary significantly. The repeat content of the isolates ranged from 1.12% in *C. nymphaeae* 6Ca to 1.27% in *C. siamense* 84Ca, with interspersed repeats constituting 0.06% of the total genome (Table 1; Supplementary Table S3). Gene prediction analysis revealed that the average gene count for *C. nymphaeae* isolates within the *C. acutatum* ranged from 14,152 to 14,260. Isolates belonging to the *C. gloeosporioides* showed similar trends. In contrast, *C. siamense* isolates within the *C. gloeosporioides* had more predicted gene models, ranging from 17,420 to 17,729 (Table 2).

Comparative genome analysis and sequence synteny

Colletotrichum genomic sequences were analyzed using MCScanX (Wang et al., 2012) and D-Genies (Cabanettes and Klopp, 2018). In

the D-Genies plot (Figure 2A) the diagonal line indicates regions of high sequence identity, suggesting conserved sequences between the two genomes However, *C. siamense* HN08 and *C. nymphaeae* 34Ca exhibit low sequence similarity with deviations from diagonal line, indicating frequent insertions or deletions (Figure 2B). The genome synteny and collinearity analysis of isolates 40Ca, and 57Cg isolates shows 73.18% shared collinear genes (Supplementary Figure S2).

The protein functions of the seven *Colletotrichum* isolates were evaluated by clustering proteins into homologous families known as orthogroups. Analysis of the protein-coding genes showed that the *Colletotrichum* genomes contained 20,103 orthologous gene families, comprising a total of 112,951 genes. This represented 99.7% of all predicted genes (113,284) (Figure 3; Supplementary Table S4). Among these, 10,278 gene families were shared by all seven isolates and were categorized as core gene families. Orthofinder identified proteins organized into 9,137 single orthogroups, with eight orthogroups being species-specific and encoding 198 genes (Figure 3; Supplementary Table S4).

This study identified and analyzed the secreted proteins, CAZ enzymes, effectors, secondary metabolite synthesis gene clusters, and effectors in the genomes of each strain. The number of secreted proteins ranged from 1,112 to 1,121 for *C. nymphaeae* isolates, while the secretome gene count ranged from 1,336 to 1,354 for *C. siamense* isolates (Table 2). The CAZymes of three isolates belonging to the *C. nymphaeae* ranged from 800 to 1,002, whereas the numbers of these potentially pathogenic genes were significantly higher in *C. siamense* isolates, ranging from 973 to 1,060 CAZymes (Table 2; Figure 4A; Supplementary Table S5). Isolate 28Cg had the highest number, while 40Ca had the lowest number. CAZymes play an essential role in facilitating the acquisition of complex carbohydrate metabolism. Most of the genes identified encode glycoside hydrolases (GHs) (41%), glycosyltransferases (GTs) (16%), auxiliary activity (AAs) (16%), carbohydrate-binding modules (CBMs) (12%), carbohydrate esterases (CEs) 8% and polysaccharide lyases (PLs) 3% (Figure 4A; Supplementary Table S5).

Isolates *C. nymphaeae* 6Ca and *C. siamense* 84Cg have the lowest and highest distribution, respectively. In these families, enzymes that break down lignocellulose and cellulose-like GHs 1, 2, 3, 7, 9, 10, 12, 74, and AA9 were identified. Lignin-modifying enzymes (LMEs) were mainly distributed in families AA1 and AA2. Across all seven genome assemblies, the CAZyme family's AA3 and the GH3 of gluco-oligosaccharide oxidases were the most abundant. CAZyme AA3 is typically found in wood-degrading fungi (Ludwig and Kracher, 2024), whereas the GH3 family is associated with breaking down and debranching hemicellulose, such as glucosidase, xylosidase, and glucanase activities. This family encompasses a variety of functions, including members with β-D-glucosidase, β-D-xylopyranosidase, α-L-arabinofuranosidase, and N-acetyl-β-D-glucosaminide phosphorylase activities, which are essential in degrading oligosaccharides. The Auxiliary Activity CAZyme Families AA1, AA2, AA9, AA3, and AA7 were much more frequent in the 84Cg assembly (Figure 4B). This same trend was observed in the Glycoside hydrolase families, with GH16, GH43, and GH71 showing significantly higher frequencies in 85Cg (Figure 4B).

Effectors predicted via EffectorP-Fungi v.3.0 were classified as apoplastic, cytoplasmic, and dual-predicted apoplastic/cytoplasmic effectors (Figure 4C). More than 70% of the effectors were classified

---

TABLE 1 Summary of the genome statistics for the seven *Colletotrichum* isolates obtained from strawberry in North Carolina.

		C. nymphaeae			*C. siamense*			
	Feature	6Ca	34Ca	40Ca	28Cg	57Cg	58Cg	84Cg
Genome annotation	Genome size (Mb)[a]	50.9	51	50.3	56.1	57.8	58.6	56.5
	Genome coverage	87×	103×	58×	75×	68.0×	98×	80×
	GC (%)[b]	51.98	51.98	52.60	53.50	53.50	52.90	53.70
	# Contigs	958	1,002	958	2,391	2,270	973	1,002
	Largest contigs (kb)	668,652	667,855	668,652	623,307	572,964	816,516	762,990
	N50[c] contig length (kb)	183,114	183,989	183,114	174,300	167,900	202,057	197,661
	L50[d]	88	88	83	93	100	92	90
	N90	47,203	47,203	52,516	51,290	46,765	46,742	51,178
	L90	288	288	265	297	330	331	332
	Repetitive elements (%)	1.12	1.22	1.15	1.25	1.14	1.27	1.15
Genome completeness	Busco genome completeness (%) [e]	99.1	99.1	98.2	99.7	99.4	98.9	99.2
	BioSample Accession number	SAMN40582474	SAMN40582476	SAMN4058247	SAMN40582478	SAMN40582479	SAMN40582480	SAMN40582481
	SRA accession numbers	SRR29996184	SRR29996182	SRR29996181	SRR29996180	SRR29996179	SRR29996178	SRR29996182
	GenBank accession #	-[f]	-	-	GCA_040955545.1	GCA_040955525.1	-	GCA_040955925.1

[a]Genome size is calculated from the total sequence length of the final processed reads.
[b]GC% calculated from the final assembly of processed reads.
[c]N50 is defined as the sequence length of the shortest scaffold at 50% of the total assembly length.
[d]L50 is defined as the number of scaffolds whose summed length equals 50% of the total assembly length.
[e]Benchmarking Universal Single-Copy Orthologs (BUSCO).
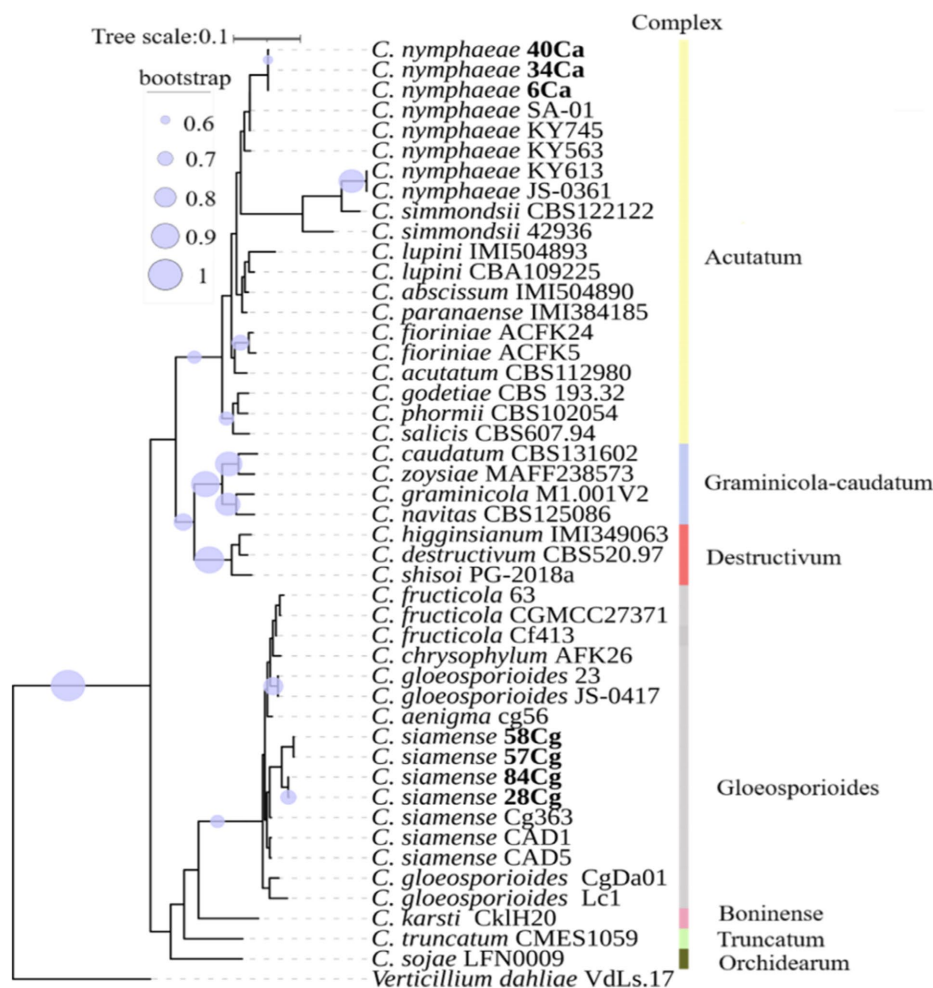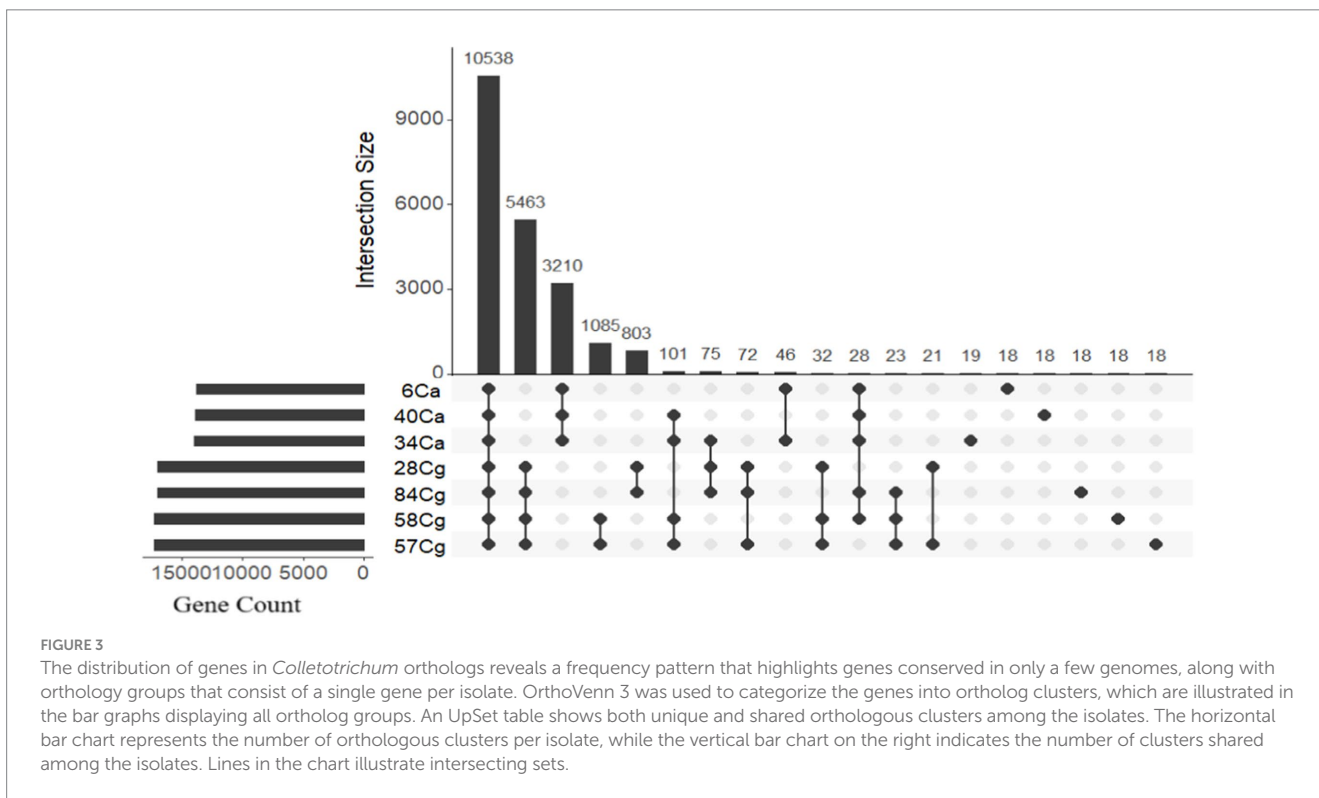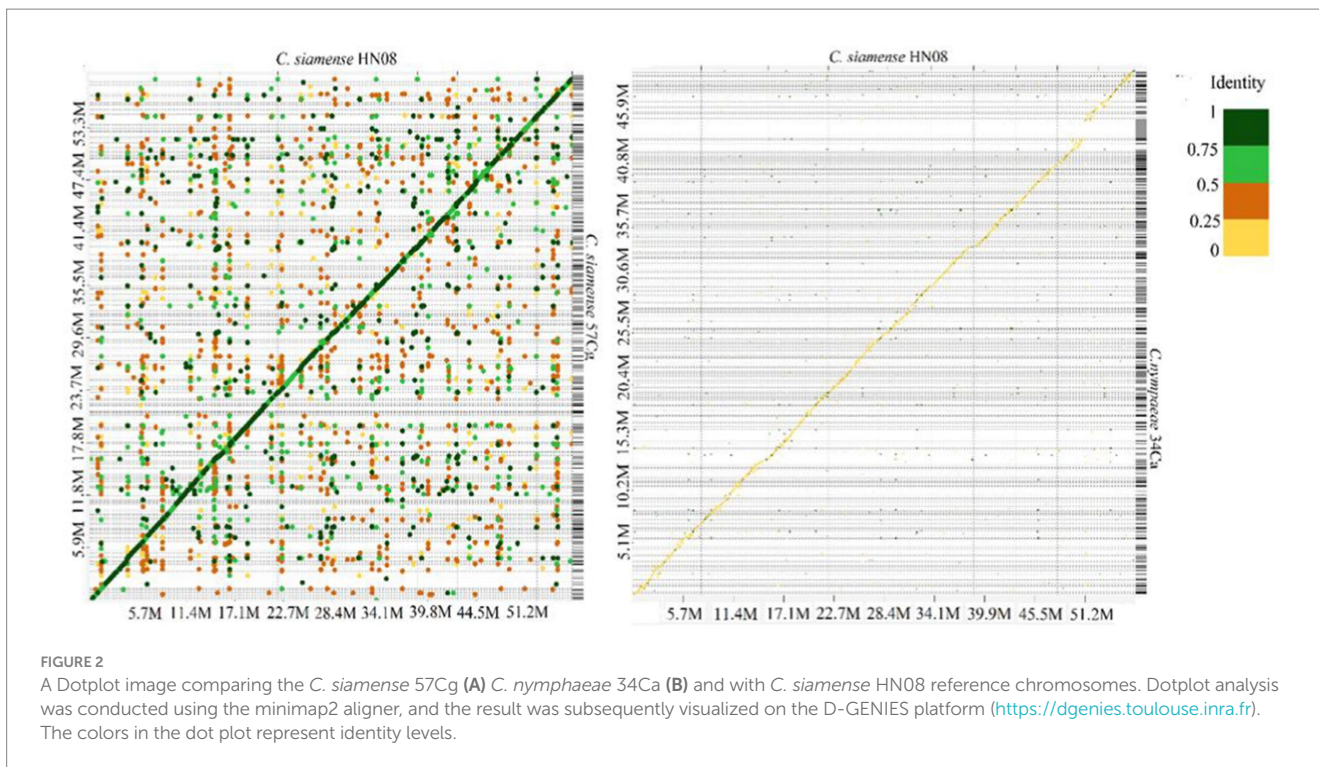[f]GenBank accession numbers not available.

FIGURE 1
A phylogenomic tree constructed using single-copy orthologous genes illustrates the relationship of *Colletotrichum* spp. with additional *Colletotrichum* isolates from NCBI (Supplementary Table S2). The nodes in the tree depict maximum likelihood (ML) bootstrap support values ≥60% from 1,000 replicates. The isolates studied are highlighted in bold. Different color boxes indicate isolates belonging to various *Colletotrichum* spp. complexes.

TABLE 2 Genome assembly and functional annotation statistics of the seven *Colletotrichum* isolates from strawberry in North Carolina.

		C. nymphaeae			*C. siamense*			
	Feature	6Ca	34Ca	40Ca	28Cg	57Cg	58Cg	84Cg
Gene prediction	Predicted protein models	14,235	14,152	14,260	17,420	17,686	17,673	17,729
	Coding sequences	13,901	13,801	13,905	16,922	17,196	17,280	17,318
	tRNA	334	351	355	498	490	393	411
Gene annotation	rRNA	24	24	24	28	33	34	29
	Secretome	1,118	1,112	1,121	1,336	1,351	1,354	1,353
	Predicted CAZymes	800	1,002	958	1,060	1,043	973	1,002
	Predicted effectors	347	352	339	471	478	480	479
	Secondary metabolites gene clusters	67	72	71	90	87	88	91

as apoplastic. Consistent numbers of each effector type were detected in each isolate from the same complex. Generally, isolates belonging to *C. siamense* had more effectors (471 to 480) than *C. nymphaeae* isolates (339 to 347). *C. nymphaeae* isolates shared the same effectors and had no isolate-specific effectors (Table 3). However, 28Cg and
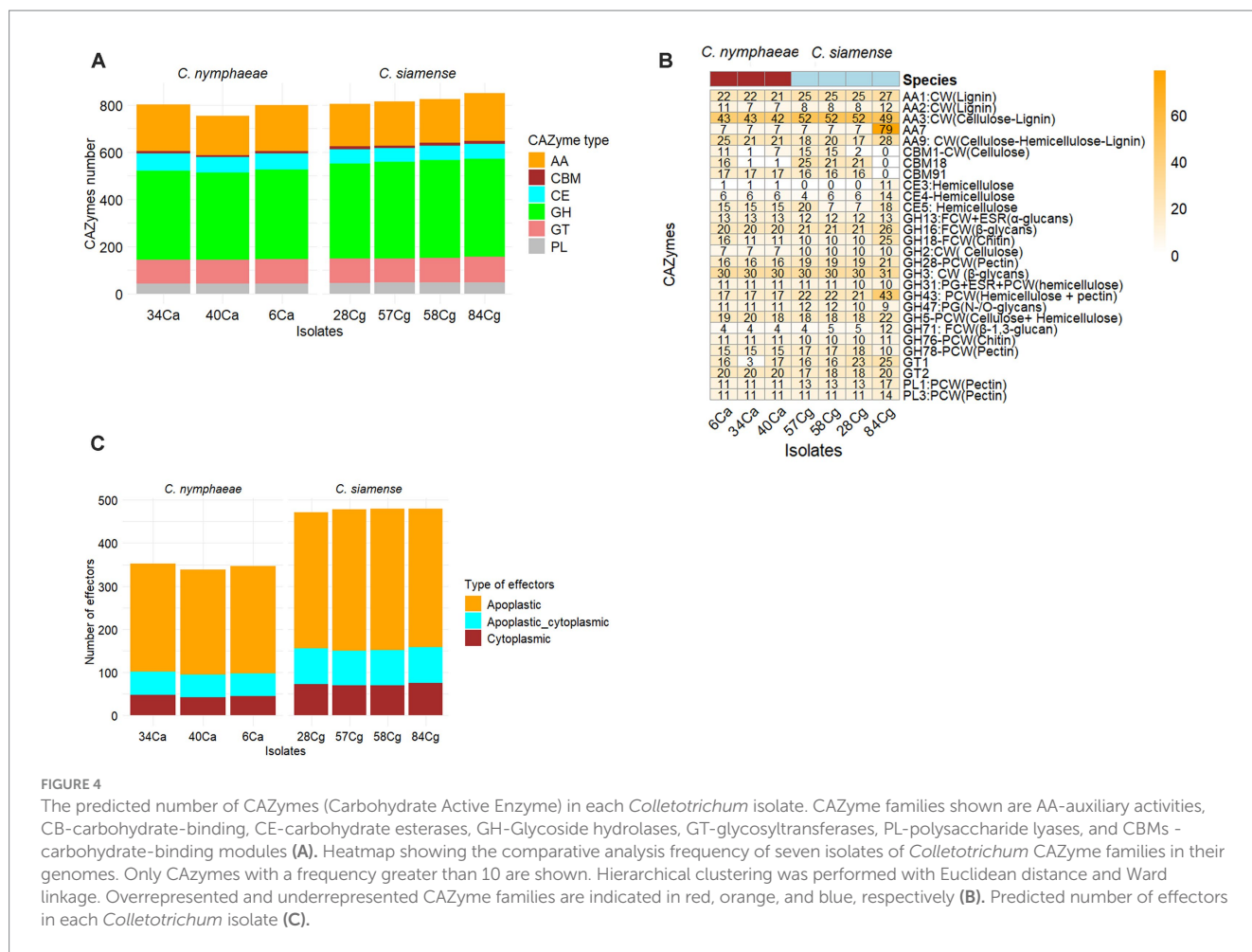
57Cg had isolate-specific effectors, such as the EC51a and Celp0028 effector-like proteins. Isolate 84Cg had seven isolate-specific effectors, five of which were uncharacterized proteins. The other two were annotated as the secreted virulence factor MC69 and the GPI-anchored serine–threonine-rich protein. The *C. nymphaeae* isolates had more

FIGURE 2
A Dotplot image comparing the *C. siamense* 57Cg **(A)** *C. nymphaeae* 34Ca **(B)** and with *C. siamense* HN08 reference chromosomes. Dotplot analysis was conducted using the minimap2 aligner, and the result was subsequently visualized on the D-GENIES platform (https://dgenies.toulouse.inra.fr). The colors in the dot plot represent identity levels.

FIGURE 3
The distribution of genes in *Colletotrichum* orthologs reveals a frequency pattern that highlights genes conserved in only a few genomes, along with orthology groups that consist of a single gene per isolate. OrthoVenn 3 was used to categorize the genes into ortholog clusters, which are illustrated in the bar graphs displaying all ortholog groups. An UpSet table shows both unique and shared orthologous clusters among the isolates. The horizontal bar chart represents the number of orthologous clusters per isolate, while the vertical bar chart on the right indicates the number of clusters shared among the isolates. Lines in the chart illustrate intersecting sets.

NLP (3) domain effectors and fewer LysM domain proteins (33 to 34), compared to the *C. siamense*, which had 1–2 NLP domain and 56–62 LysM domain proteins (Table 3).

To identify potential genes involved in the biosynthesis of secondary metabolites in the *Colletotrichum* genomes, we utilized antiSMASH v7.0 to annotate and analyze the gene clusters

associated with secondary metabolite biosynthesis (BGCs). The BGC clusters ranged from 69 to 71 for *C. nymphaeae* isolates and 87 to 91 for the *C. siamense* isolates (Figure 5). The gene clusters were categorized into 17 BGC types (Figure 5; Supplementary Table S6). Among these, terpenes, Type I polyketide synthase (T1PKS), non-ribosomal peptide synthetase cluster

FIGURE 4
The predicted number of CAZymes (Carbohydrate Active Enzyme) in each *Colletotrichum* isolate. CAZyme families shown are AA-auxiliary activities, CB-carbohydrate-binding, CE-carbohydrate esterases, GH-Glycoside hydrolases, GT-glycosyltransferases, PL-polysaccharide lyases, and CBMs - carbohydrate-binding modules **(A)**. Heatmap showing the comparative analysis frequency of seven isolates of *Colletotrichum* CAZyme families in their genomes. Only CAZymes with a frequency greater than 10 are shown. Hierarchical clustering was performed with Euclidean distance and Ward linkage. Overrepresented and underrepresented CAZyme families are indicated in red, orange, and blue, respectively **(B)**. Predicted number of effectors in each *Colletotrichum* isolate **(C)**.

(NRPS), and fungal ribosomal synthesized and post-translationally modified peptide product cluster (RiPP-like) were the most prevalent BGCs in the *Colletotrichum* genus. NRPS-like fragments, isocyanides, indoles, beta lactones, T3PKS, and hybrids of NRPS (terpene, T1PKS, T3PKS, and fungal-RiPP-like) were also observed, though at lower frequencies. The BGC types T3PKS and its hybrids were specific to C. *nymphaeae*, whereas beta-lactam was absent from the C. *siamense* (Supplementary Table S6).

Secondary metabolite clusters microperfuramone, nectriapyrone, alternaryrone, and xenolozoyerone had 100% similarity with known biosynthetic gene clusters, while polypyrone with 80% similarity was unique to C. *nymphaeae* (Supplementary Table S6). UNII-Y-C20Q1094PT, fusarin, chrysogen, betaenone, clavaric acid, ACT-toxin III, ACE, ipicidin, and bataestatic were unique to C. *siamense* (Supplementary Table S6). Shared secondary metabolite clusters between the two complexes include choline, cichorine, biotin, cercosporin, Sequalistin 1, and ACE (Supplementary Table S6).

The genome-wide annotations of the *Colletotrichum* isolates were analyzed using the Pathogen-Host Interaction Database (PHI). This analysis identified genes associated with reduced virulence, loss of pathogenicity, increased virulence, effector functions, and annotated chemical targets (Table 4). The C. *siamense* isolates contained more lethal genes and genes linked to reduced virulence compared to C. *nymphaeae* isolates.
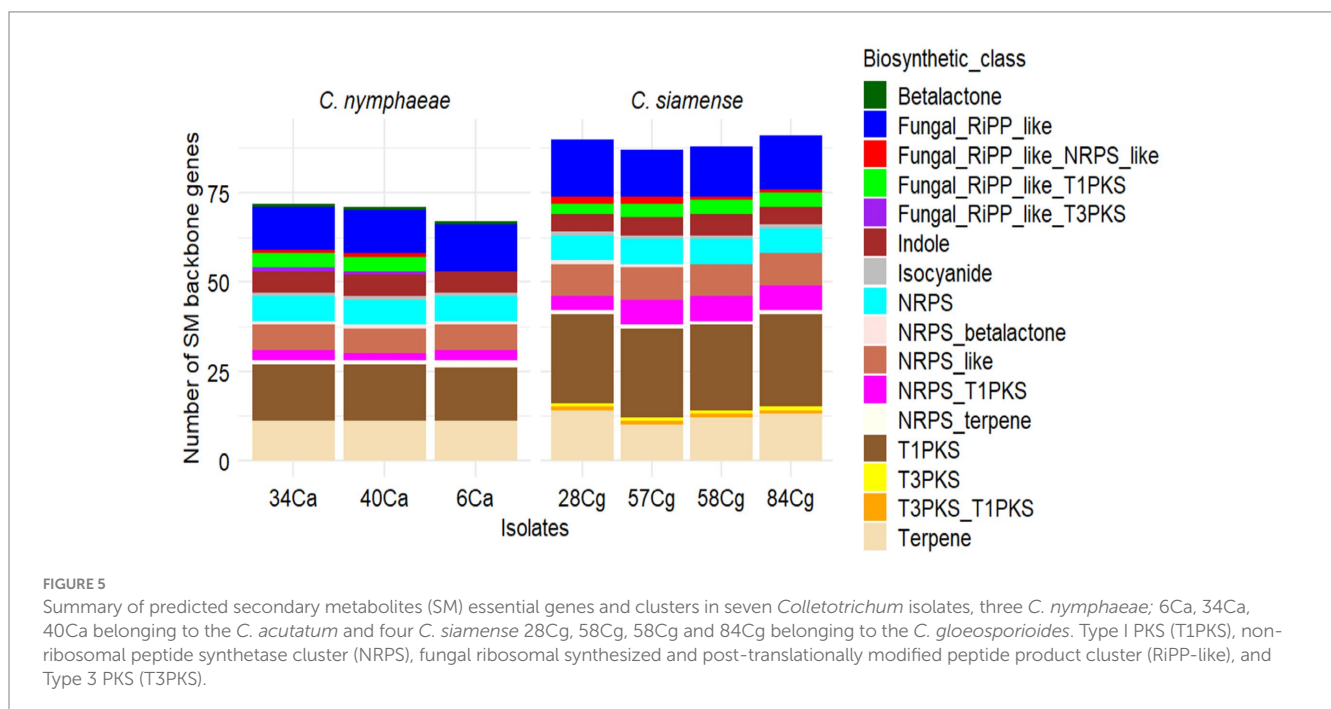
In the KEGG database, 4,429 genes (29.4% of the total predicted genes) were annotated for *C. siamense* isolates, and 4,162 genes (24.9%) were annotated for *C. nymphae* isolates. Most genes were related to genetic information processing (785–794), carbohydrate metabolism (451–488), cellular processes (246–248), and environmental information processing, metabolism, and organismal systems (Figure 6).

Discussion

Strawberry anthracnose, caused by *Colletotrichum* spp., is a significant disease affecting strawberry in the United States and other countries. A detailed analysis of the genome sequence was performed to gain insights into the genomes of anthracnose-causing pathogens of strawberry in North Carolina. This study focussed on the genome profiles of C. *nymphaeae* and C. *siamense* isolates from the *Colletotrichum* complexes, Acutatum, and Gloeosporioides, respectively. Due to the limited availability of whole genome sequence information for *Colletotrichum* isolates from strawberry, our understanding of secretory proteins, including degrading enzymes, remains limited. Therefore, whole genome sequencing and annotation of the *Colletotrichum*-strawberry pathosystem are vital for advancing our understanding of this disease.

TABLE 3 Effectors that are shared between isolates used in this study.

Isolate	Number of effectors	Shared effectors	Isolate specific effectors	BLAST identity	Effectors with NLP domain (PF05630 and IPR008701)	Effectors with LySM motifs
6Ca	347	347	0	–	3	33
34Ca	347	347	0	–	3	34
40Ca	339	339	0	–	3	33
28Cg	471	470	1	EC51a protein	2	58
57Cg	478	477	1	Celp0028 effector-like protein	1	62
58Cg	481	480	1	Uncharacterized proteins 1 Secreted virulence factor MC69	1	59
84Cg	479	472	7	Uncharacterized proteins 1	2	56

FIGURE 5
Summary of predicted secondary metabolites (SM) essential genes and clusters in seven *Colletotrichum* isolates, three *C. nymphaeae;* 6Ca, 34Ca, 40Ca belonging to the *C. acutatum* and four *C. siamense* 28Cg, 58Cg, 58Cg and 84Cg belonging to the *C. gloeosporioides*. Type I PKS (T1PKS), non-ribosomal peptide synthetase cluster (NRPS), fungal ribosomal synthesized and post-translationally modified peptide product cluster (RiPP-like), and Type 3 PKS (T3PKS).

In this study, the genomes of seven *Colletotrichum* isolates were sequenced, assembled, annotated, and analyzed, with a focus on evidence of carbohydrate-active enzymes (CAZymes), effectors, and secondary metabolic clusters. The *C. siamense* isolates exhibited the largest genome sizes, ranging from 55.7 Mb to 58.6 Mb, while the *C. nymphaeae* isolates had genome sizes between 50.3 Mb and 50.7 Mb. These findings align with previous reports on other *Colletotrichum* spp. (Armitage et al., 2020; Becerra et al., 2023; Wang et al., 2023). High N50 values and BUSCO completeness metrics indicate the quality of the genome assembly. The proportion of repetitive sequences in the genomes varied from 1.12% (isolate 6Ca) to 1.27% (isolate 58Cg), compared to 1.03% reported in other *Colletotrichum* species (Lv et al., 2024). This variation was not dependent on genome size and was significantly lower than the proportions reported in other studies, 3.63% for *C. fructicola* (Armitage et al., 2020). This reduced proportion can be attributed to our sequencing and assembly strategy, which excluded scaffolds shorter than 200 bp. This study provides high-quality gene predictions and
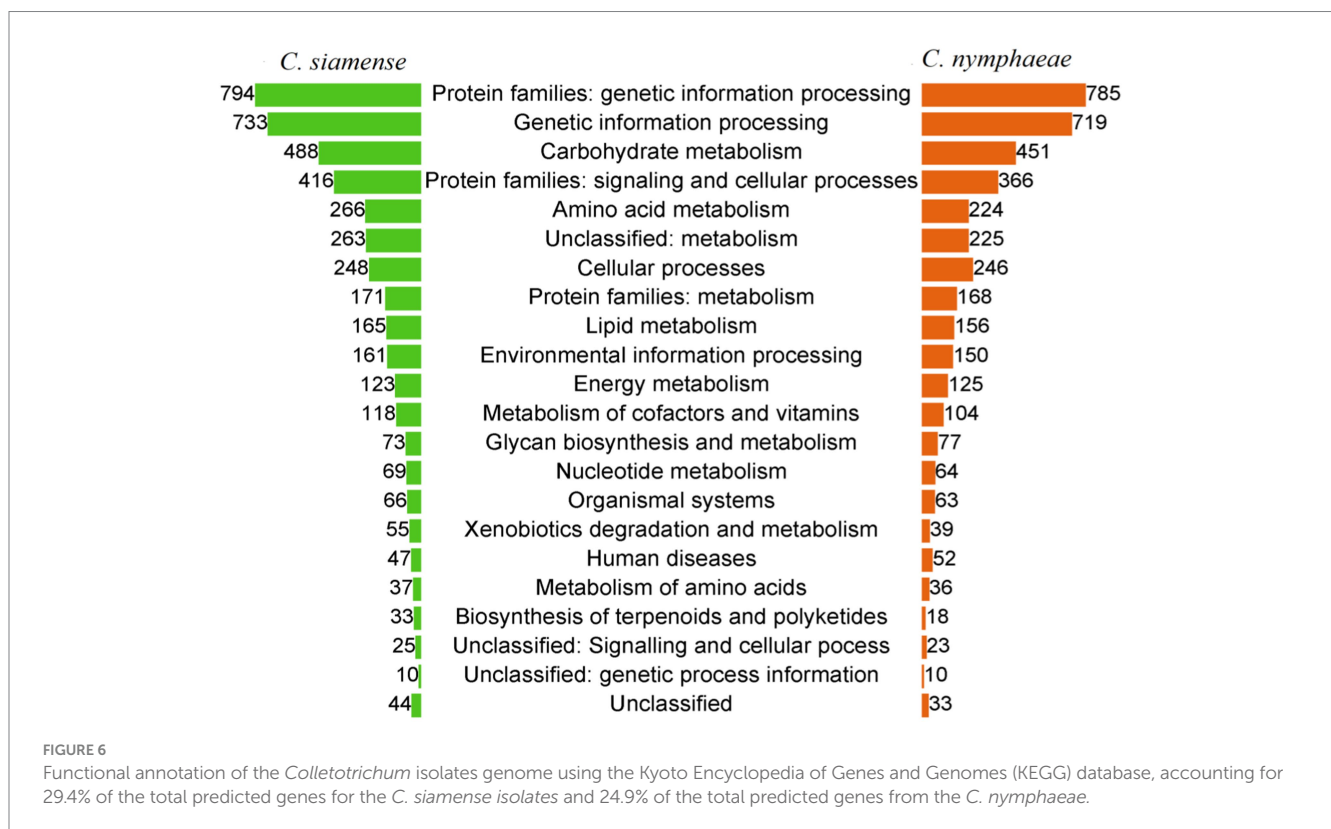
annotations, establishing a robust foundation for future genetic, population genomic, and evolutionary research.

To delineate the boundaries of the *Colletotrichum* species complex, the phylogenomic tree exhibited a topology consistent with that of the five-locus gene tree. Most species complexes formed well-supported clades, corroborating findings from previous studies (Crouch et al., 2014; Jayawardena et al., 2016). The three isolates from the *C. acutatum* formed a robust clade with *C. nymphaeae*. In contrast, isolates from the *C. gloeosporioides* clustered with *C. siamense*. These taxa are prevalent pathogens affecting various plants (Weir et al., 2012; De Silva et al., 2017) and have been documented as infecting strawberry globally (Jayawardena et al., 2016; Baroncelli et al., 2015; Wang et al., 2019; Armitage et al., 2020).

The predicted gene counts for *Colletotrichum* spp. ranged from 14,250 to 17,686, with *C. siamense* isolates displaying the highest predicted gene models. Significant genetic divergences exist between *C. nymphaeae* and *C. siamense* isolates, influencing their lifestyles and pathogenicity (Baroncelli et al., 2015). The predicted gene count for *C. siamense* exceeds that reported by Rehner et al. (2023) but is lower

TABLE 4 Pathogen Host Interaction database annotation result statistics of *Colletotrichum* isolates.

Phenotypic category	C. nymphaeae			C. siamense			
	6Ca	40Ca	34Ca	28Cg	57Cg	58Cg	84Cg
Reduced virulence	4,513	4,413	4,421	4,341	3,789	4,210	3,641
Unaffected pathogenicity	2,649	2,565	2,794	3,406	2,700	2,589	2,438
Planta virulence determinant	730	733	733	777	728	657	698
Loss of pathogenicity	341	331	333	373	372	372	373
Increased virulence	395	375	379	430	381	370	370
Lethal	146	150	147	172	172	172	172
Loss of pathogenicity	68	69	70	70	75	78	78
Chemistry target: resistance to chemical	40	28	28	32	31	29	32

FIGURE 6
Functional annotation of the *Colletotrichum* isolates genome using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, accounting for 29.4% of the total predicted genes for the *C. siamense isolates* and 24.9% of the total predicted genes from the *C. nymphaeae.*

than the 18,324 protein-coding genes identified in *C. lupini* (Baroncelli et al., 2021). Secretome predictions for *C. nymphaeae* isolates ranged from 1,112 to 1,121, while those for *C. siamense* isolates ranged from 1,336 to 1,354. Isolates within the same complex tend to exhibit similar secretome compositions. Previous research indicates that secretome size correlates with phylogenetic relationships (Krijger et al., 2014), and variations in secretomes are associated with fungal lifestyles (Alfaro et al., 2014). Furthermore, isolates within the same species complex that possess similar genome sizes tend to have analogous secretome components. The secretome encompasses fungal virulence factors, including carbohydrate-active enzymes (CAZymes), effector proteins, secondary metabolite gene clusters, and additional effector proteins (Tsushima and Shirasu, 2022).

CAZymes are critical enzymes that facilitate the synthesis and degradation of complex carbohydrates within cells, playing a vital role in fungal biological activity and pathogenicity. The predicted CAZyme

counts in *Colletotrichum* isolates ranged from 800 (6Ca) to 1,060 (28Cg), aligning with similar figures observed in other *Colletotrichum* pathogens (Armitage et al., 2020; Zhang et al., 2023). Notably, the *C. siamense* species complex exhibited the highest average CAZyme count. This abundance may elucidate the capability of species within this complex to infect 283 host plants at various developmental stages, functioning either as pathogens or endophytes (Talhinhas and Baroncelli, 2021). This observation aligns with earlier findings (Baroncelli et al., 2016) indicating that host range diversity is linked to the expansion and contraction of gene families in *Colletotrichum*.

This study identified a diverse array of glycoside hydrolases (GHs) critical for degrading glycosidic bonds in plant cell wall polysaccharides, facilitating cell wall degradation and subsequent infection. Both *Colletotrichum* complexes exhibit comparable enzyme profiles for cellulose degradation, including GH5, GH6, GH7, GH12, GH45, and GH61, and for hemicellulose breakdown,

encompassing CE1, GH10, GH11, GH26, GH29, GH43, GH51, GH53, GH54, GH62, GH67, GH74, and GH93. Additionally, enzymes from the CE10, CBM1, AA9, AA3, AA7, GH5, GH16, GH28, GH43, and PL1 families were detected, underscoring the AA family's role in plant cell wall degradation. Notably, isolate 84Cg demonstrates a high abundance of AA7 enzymes, potentially involved in the conversion or detoxification of lignocellulosic compounds. Enzymes from the AA9 and GH16 families exhibit cellulase activity, while GH28, GH43, and PL1 show pectinase activity. All fungal genomes analyzed encode comparable quantities of glycosyltransferases (GTs) necessary for fungal cellular functions, including cell wall synthesis and carbohydrate metabolism. These observations corroborate findings by O'Connell et al. (2012). The presence of lignin-degrading AA2 family enzymes suggests a role in xylem vessel infection, aiding the translocation of propagules and enhancing secondary infections. The CAZyme profiles of *C. nymphaeae* and *C. siamense* isolates resemble those of hemibiotrophic and necrotrophic plant pathogens, further emphasizing the importance of glycoside hydrolases in the infection processes of *Colletotrichum* spp (Tsushima and Shirasu, 2022).

Colletotrichum spp. exhibit a broad host range and utilize two primary classes of secreted effectors for infection. Apoplastic effectors, located outside plant cells, comprise glycoside hydrolases and necrosis-inducing proteins (NLPs), which are essential for compromising the plant's physical and chemical defenses and providing necessary nutrients during early infection stages (Dong et al., 2012). In contrast, cytoplasmic effectors manipulate host physiology and morphology to increase susceptibility, including suppressing host immunity and promoting programmed cell death. These effectors typically enter plant cells via specialized structures known as haustoria. Genomic analyses reveal that 75% of the effectors are apoplastic, 10% are cytoplasmic, and 15% exhibit characteristics of both. Most effectors are conserved across isolates; however, isolate 28Cg contains a unique effector with 95% identity to a putative Celp0028 effector-like protein, contributing to the virulence and stress tolerance of *Sclerotinia sclerotiorum* (Wei et al., 2022). Isolate 84Cg possesses seven isolate-specific effectors, none shared with other isolates. Among these, five are uncharacterized proteins, while two resemble a GPI-anchored cell wall protein and the secreted protein MC69, as determined by BLAST and InterProScan analyses. The GPI-anchored cell wall protein is critical for maintaining cell wall integrity and serves as a pathogenicity determinant in many fungi. Additionally, the secreted protein MC69 is vital for the infection of both dicots and monocots and is essential for the full virulence of *C. orbiculare* (Irieda et al., 2016).

A BLASTP analysis utilizing the Swiss-Prot database identified several effector candidates containing necrosis-inducing proteins (NLPs) and lysin motifs (LysMs), as determined by the LysM domains identified through InterProScan (Jones et al., 2014). All isolates exhibited effector proteins with NLP activity, identified through searches using IPR (InterPro) and PFAM terms while cataloging domains with supplementary IPR and PFAM annotations. Isolates of *C. nymphaeae* contained three NLP proteins, while *C. siamense* exhibited one. NLPs are known to induce leaf necrosis and activate immune responses in dicotyledonous plants (Pemberton and Salmond, 2004). Furthermore, all isolates possessed LysM-encoding genes, with counts ranging from 33 to 34 in *C. nymphaeae* and from 56 to 62 in *C. siamense*. LysMs interact with chitin and peptidoglycans,

sequestering chitin fragments released during fungal infections in the apoplastic space, thereby inhibiting the detection of these fragments by plant immune receptors (Krijger et al., 2014).

Fungi synthesize various secondary metabolites (SMs) crucial for environmental interactions and organismal communication (Keller et al., 2005). These metabolites encompass polyketides (PKs), non-ribosomal peptides (NRPS), terpenes (TS), isocyanates, indoles, beta-lactones, and fungi-RiPP. They serve as protective agents against stress and significantly influence pathogenicity and virulence. This study identified six prominent groups of secondary metabolites produced by *Colletotrichum* isolates: PKs, NRPS, TS, non-canonical isocyanide synthase, trimodular NRPSs, and fungi-RiPP. Furthermore, PKS, NRPS, and PKS-NRPS hybrids were detected among the isolates. Most secondary metabolite clusters in the sequenced isolates were reminiscent of type I PKs, typically associated with fungal toxin biosynthesis (Keller et al., 2005). The predominant secondary metabolite gene clusters (SMGCs) observed were 15 in *C. nymphaeae* and 25 in *C. siamense*. Unique SM clusters in *C. nymphaeae* included microperfuramone, nectriapyrone, alternaryrone, and xenolozoyerone. Notably, alternaryrone, related to alternariol, exhibits mycotoxin properties and induces oxidative stress in plant cells, thus enhancing the pathogenicity of *C. nymphaeae* (Talhinhas and Baroncelli, 2021).

Colletotrichum siamense isolates produce distinctive secondary metabolites, including UNII-Y-C20Q1094PT, fusarin, chrysogen, betaenone, clavaric acid, and ACT-toxin III. Betaenone exhibits phytotoxic effects, such as chlorosis and necrosis in host plants. Clavaric acid inhibits squalene synthase, disrupting sterol biosynthesis and compromising the host plant's cellular integrity, which facilitates fungal invasion (Baroncelli et al., 2016). ACT-toxin III comprises host-specific toxins that target susceptible plant varieties, resulting in cell death and tissue necrosis, thereby allowing *efficient* infection of specific plant hosts by *C. siamense* (Dowling et al., 2020).

Both species within the *Colletotrichum* complex possess specific clusters of secondary metabolites (SM), including choline, chlorine, biotin, cercosporin, Sequalistin 1, and ACE, which are critical for the survival and virulence of pathogenic fungi (Krijger et al., 2014). The choline and scytalone/T3HN biosynthetic gene clusters reside on chromosome 2. Choline is essential for phospholipid formation, and vital for cell membranes. In *Colletotrichum*, choline may aid in maintaining cellular integrity during infection. Cichorine, a phytotoxic compound, inhibits plant growth and defense responses (Keller et al., 2005). A cercosporin-like biosynthetic gene cluster has also been mapped to chromosome 7. Cercosporin generates reactive oxygen species (ROS) that damage host plant cells, further facilitating infection. Sequalistin 1 and ACE inhibit proteases, enabling *Colletotrichum* to circumvent plant defense mechanisms. Collectively, these secondary metabolite profiles bolster *Colletotrichum*'s hemibiotrophic and necrotrophic lifestyle (Talhinhas and Baroncelli, 2021).

Conclusion

The seven sequenced genomes of *Colletotrichum* isolates augment the genome database and facilitate comparative genomics for

investigating host specificity, pathogenicity mechanisms, and evolutionary history. They also contribute to the development of diagnostic tools and population genetic markers. These resources will bolster future research on strawberry - *Colletotrichum* interactions, providing a foundation for studies on effector biology, secondary metabolites, transposable elements, and evolutionary population genomics.

Data availability statement

The authors state that all the necessary data for confirming the conclusions presented in the article are presented fully within the article and supplemental material. The raw sequencing data for the genome and the assembly reported in the paper is associated with BioProject number PRJNA1090952 and BioSample numbers SAMN40582474, SAMN40582476, SAMN4058247, SAMN40582478, SAMN40582479, SAMN40582480, and SAMN40582481.

Author contributions

TA: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. NM: Data curation, Formal analysis, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AP: Writing – review & editing. FL: Funding acquisition, Methodology, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2025.1515894/full#supplementary-material

References

Adhikari, T. B., Chacon, J. G., Fernandez, G. E., and Louws, F. J. (2019). First report of anthracnose causing both crown and fruit rot of strawberry by *Colletotrichum siamense* in North Carolina. *Plant Dis.* 103:1775. doi: 10.1094/PDIS-02-19-0314-PDN

Alfaro, M., Oguiza, J. A., Ramirez, L., and Pisabarro, A. G. (2014). Comparative analysis of secretomes in basidiomycete fungi. *J. Proteome* 102:43. doi: 10.1016/j.jprot.2014.03.001

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., et al. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 23:258. doi: 10.1186/s13059-022-02823-7

Armitage, A. D., Nellist, C. F., Bates, H. J., Zhang, L., Zou, X., Gao, Q., et al. (2020). Draft genome sequence of the strawberry anthracnose pathogen *Colletotrichum fructicola*. *Microbiol. Resour. Announc.* 9:10.

Bandi, V., and Gutwin, C. (2020). *Interactive exploration of genomic conservation*. In: Proceedings of the 46th Graphics Interface Conference on Proceedings of Graphics Interface 2020 (GI'20). doi: 10.20380/GI2020.09

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Baroncelli, R., Amby, D. B., Zapparata, A., Sarrocco, S., Vannacci, G., Le Floch, G., et al. (2016). Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics* 5:555. doi: 10.1186/s12864-016-2917-6

Baroncelli, R., Pensec, F., Da Lio, D., Boufleur, T., Vicente, I., et al. (2021). Complete genome sequence of the plant-pathogenic fungus *Colletotrichum lupini*. *Mol. Plant-Microbe Interact.* 34, 1461–1464. doi: 10.1094/MPMI-07-21-0173-A

Baroncelli, R., Zapparata, A., Sarrocco, S., Sukno, S. A., Lane, C. R., Thon, M. R., et al. (2015). Molecular diversity of anthracnose pathogen populations associated with UK strawberry production suggests multiple introductions of three different *Colletotrichum* species. *PLoS One* 10:e0129140. doi: 10.1371/journal.pone.0129140

Becerra, S., Baroncelli, T., Boufleur, T. R., Sukno, S. A., and Thon, M. R. (2023). Chromosomes -level analysis of the *Colletotrichum gramininicila* genome reveals the unique characteristics of core and minichromomes. *Front. Microbiol.* 14:1129319. doi: 10.3389/fmicb.2023.1129319

Blin, K., Shaw, S., Augustijn, H. E., Reitz, Z. L., Biermann, F., Alanjary, M., et al. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* 51, W46–W50. doi: 10.1093/nar/gkad344

Brooks, A. N. (1931). Anthracnose of strawberry caused by *Colletotrichum fragariae*, n. sp. *Phytopathology* 21, 739–744.

Cabanettes, F., and Klopp, C. (2018). D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ.* 6:e4958. doi: 10.7717/peerj.4958

Cannon, P. F., Damm, U., Johnston, P. R., and Weir, B. S. (2012). *Colletotrichum* – current status and future directions. *Stud. Mycol.* 73, 181–213. doi: 10.3114/sim0014

Chen, Y. P., Wu, T., Tian, W. H., Ilyukhin, E., Hyde, K. D., and Maharachchikumbura, S. S. N. (2022). Comparative genomics provides new insights into the evolution of *Colletotrichum*. *Mycosphere* 13, 134–187. doi: 10.5943/mycosphere/si/1f/5

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Crouch, J., O'Connell, R., Gan, P., Buiate, E., Torres, M. F., Beirn, L., et al. (2014). "Genomics of plant-associated fungi: monocot pathogens" in The genomics of

Colletotrichum. eds. R. Dean, A. Lichens-Park and C. Kole (Berlin, Heidelberg: Springer), 69–102.

Damm, U., Sato, T., Alizadeh, A., Groenewald, J., and Crous, P. W. (2019). The *Colletotrichum dracaenophilum*, *C. magnum* and *C. orchidearum* species complexes. *Stud. Mycol.* 92, 1–46. doi: 10.1016/j.simyco.2018.04.001

De Silva, D. D., Crous, P. W., Ades, P. K., Hyde, K. D., and Taylor, P. W. (2017). Lifestyles of *Colletotrichum* species and implications for plant biosecurity. *Fungal Biol. Rev.* 31, 155–168. doi: 10.1016/j.fbr.2017.05.001

Dong, S., Kong, G., Qutob, D., Yu, X., Tang, J., Kang, J., et al. (2012). The NLP toxin family in Phytophthora sojae includes rapidly evolving groups that lack necrosis-inducing activity. *Mol. Plant-Microbe Interact.* 5, 896–909. doi: 10.1094/MPMI-01-12-0023-R

Dowling, M., Peres, N., Villani, S., and Schnabel, G. (2020). Managing *Colletotrichum* on fruit crops: a complex challenge. *Plant Dis.* 104, 2301–2316. doi: 10.1094/PDIS-11-19-2378-FE

Gan, P., Tsushima, A., Narusaka, M., Narusaka, Y., Takano, Y., Kubo, Y., et al. (2019). Genome sequence resources for four phytopathogenic fungi from the *Colletotrichum orbiculare* species complex. *Mol. Plant-Microbe Interact.* 32, 1088–1090. doi: 10.1094/MPMI-12-18-0352-A

Garrido, C., Carbú, M., Fernández-Acero, F. J., Boonham, N., Colyer, A., Cantoral, J. M., et al. (2009). Development of protocols for detection of *Colletotrichum acutatum* and monitoring of strawberry anthracnose using real-time PCR. *Plant Pathol.* 58, 43–51. doi: 10.1111/j.1365-3059.2008.01933.x

Gunnell, P. S., and Gubler, W. D. (1992). Taxonomy and morphology of *Colletotrichum* species pathogenic to strawberry. *Mycologia* 84, 157–165. doi: 10.1080/00275514.1992.12026122

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Irieda, H., Ogawa, S., and Takano, Y. (2016). Focal effector accumulation in a biotrophic interface at the primary invasion sites of *Colletotrichum orbiculare* in multiple susceptible plants. *Plant Signal. Behav.* 11:e1137407. doi: 10.1080/15592324.2015.1137407

Jacobs, R. L., Adhikari, T. B., Pattison, J., Yencho, G. C., Fernandez, G. E., and Louws, F. J. (2019). Inheritance of resistance to *Colletotrichum gloeosporioides* and *C. acutatum* in strawberry. *Phytopathology* 109, 428–435. doi: 10.1094/PHYTO-08-18-0283-R

Jacobs, R. L., Adhikari, T. B., Pattison, J., Yencho, G. C., Fernandez, G. E., and Louws, F. J. (2020). Assessing rate-reducing foliar resistance to anthracnose crown rot and fruit rot in strawberry. *Plant Dis.* 104, 398–407. doi: 10.1094/PDIS-04-19-0687-RE

Jayawardena, R. S., Bhunjun, C. S., Hyde, K. D., Gentekaki, E., and Itthayakorn, P. (2021). *Colletotrichum*: lifestyles, biology, morpho-species, species complexes and accepted species. *Mycosphere* 12, 519–669. doi: 10.5943/mycosphere/12/1/7

Jayawardena, R., Hyde, K., Damm, U., Cai, L., Liu, M., Li, X., et al. (2016). Notes on currently accepted species of *Colletotrichum*. *Mycosphere* 7, 1192–1260. doi: 10.5943/mycosphere/si/2c/9

Ji, Y., Li, X., Gao, Q. H., Chunny, G., and Ke, D. (2022). *Colletotrichum* species pathogenic to strawberry: discovery history, global diversity, prevalence in China, and the host range of top two species. *Phytopathol. Res.* 4:42. doi: 10.1186/s42483-022-00147-9

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

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

Keller, N. P., Turner, G., and Bennett, J. W. (2005). Fungal secondary metabolism—from biochemistry to genomics. *Nat. Rev. Microbiol.* 3, 937–947. doi: 10.1038/nrmicro1286

Khodadadi, F., González, J. B., Martin, P. L., Giroux, E., Bilodeau, G. J., Peter, K. A., et al. (2020). Identification and characterization of *Colletotrichum* species causing apple bitter rot in New York and description of *C. noveboracense* sp. nov. *Sci. Rep.* 10:11043. doi: 10.1038/s41598-020-66761-9

Klosterman, S. J., Subbarao, K. V., Kang, S., Veronese, P., Gold, S. E., Thomma, B. P., et al. (2011). Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog.* 7:e1002137. doi: 10.1371/journal.ppat.1002137

Krijger, J. J., Thon, M. R., Deising, H. B., and Wirsel, S. G. (2014). Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. *BMC Genomics* 15:722. doi: 10.1186/1471-2164-15-722

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35, 3100–3108. doi: 10.1093/nar/gkm160

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty19

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAM tools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Liu, F., Ma, Z. Y., Hou, L. W., Diao, Y. Z., Wu, W. P., Damm, U., et al. (2022). Updating species diversity of *Colletotrichum*, with a phylogenomic overview. *Stud. Mycol.* 101, 1–56. doi: 10.3114/sim.2022.101.01

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Ludwig, R., and Kracher, D. (2024). *Auxiliary activity family 3*. https://www.cazypedia.org/index.php/Auxiliary_Activity_Family_3 (Accessed December 12, 2024).

Lv, W., Jiang, H., Li, X., Lu, Q., Wang, X., and Wang, Y. (2024). Draft genome sequence of *Colletotrichum fructicola* causing leaf spot on tea plants (*Camellia sinensis*). *Plant Pathol.* 73, 144–156. doi: 10.1111/ppa.13802

Maddi, A. M. A., Kavousi, K., Arabfard, M., Ohadi, H., and Ohadi, M. (2022). Tandem repeats ubiquitously flank and contribute to translation initiation sites. *BMC Genom. Data* 23:59. doi: 10.1186/s12863-022-01075-5

Manni, M., Berkeley, M. R., Seppey, M., and Zdobnov, E. M. (2021). BUSCO: Assessing genomic data quality and beyond. *Curr. Protocols.* 1:e323. doi: 10.1002/cpz1.323

O'Connell, R. J., Thon, M. R., Hacquard, S., Amyotte, S. G., Kleemann, J., Torres, M. F., et al. (2012). Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat. Genet.* 44, 1060–1065. doi: 10.1038/ng.2372

Pemberton, C. L., and Salmond, G. P. (2004). The Nep1-like proteins-a growing family of microbial elicitors of plant necrosis. *Mol. Plant Pathol.* 5, 353–359. doi: 10.1111/j.1364-3703.2004.00235.x

Rahman, M., Islam, T., Schwegel, R., and Louws, F. J. (2019). Simultaneous detection of *Colletotrichum acutatum* and *C. gloeosporioides* from quiescently infected strawberry foliage by real-time PCR based on high resolution melt curve analysis. *Am. J. Plant Sci.* 10, 382–401. doi: 10.4236/ajps.2019.103028

Rahman, M., Ojiambo, P., and Louws, F. J. (2015). Initial inoculum and spatial dispersal of *Colletotrichum gloeosporioides*, the causal agent of strawberry anthracnose crown rot. *Plant Dis.* 99, 80–86. doi: 10.1094/PDIS-02-13-0144-RE

Rehner, S. A., Gazis, R., Doyle, V. P., Vieira, W. A. S., Campos, P. M., and Shao, J. (2023). Genome resources for the *Colletotrichum* gloeosporioides species complex: 13 tree endophytes from the neotropics and paleotropics. *Microbiol. Resour. Announc.* 12:e01022. doi: 10.1128/mra.01040-22

Samtani, J. B., Rom, C. R., Friedrich, H., Fennimore, S. A., Finn, C. E., and Petran, A. (2019). The status and future of the strawberry industry in the United States. *HortTechnology* 29, 11–24. doi: 10.21273/HORTTECH04135-18

Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11:e0163962. doi: 10.1371/journal.pone.0163962

Smith, B. J., and Black, L. L. (1990). Morphological, cultural and pathogenic variation among *Colletotrichum* species isolated from strawberry. *Plant Dis.* 74, 69–76. doi: 10.1094/PD-74-0069

Sperschneider, J., and Dodds, P. N. (2022). EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol. Plant-Microbe Interact.* 35, 146–156. doi: 10.1094/MPMI-08-21-0201-R

Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., and Wang, Y. (2023). OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* 51, W397–W403. doi: 10.1093/nar/gkad313

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

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Tsushima, A., and Shirasu, K. (2022). Genomic resources of *Colletotrichum* fungi: development and application. *J. Gen. Plant Pathol.* 88, 349–357. doi: 10.1007/s10327-022-01097-y

Wang, N. Y., Forcelini, B. B., and Peres, N. A. (2019). Anthracnose fruit and root necrosis of strawberry are caused by a dominant species within the *Colletotrichum acutatum* species complex in the United States. *Phytopathology* 109, 1293–1301. doi: 10.1094/PHYTO-12-18-0454-R

Wang, H., Huang, R., Ren, J., Tang, L., Huang, S., Chen, X., et al. (2023). The evolution of mini-chromosomes in the fungal genus *Colletotrichum*. *MBio* 14:e00623. doi: 10.1128/mbio.00629-23

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Wei, W., Pierre-Pierre, N., Ellur, V., and Chen, W. (2022). A putative Celp0028 effector-like protein contributes to virulence and stress tolerance of Sclerotinia sclerotiorum. *Phytopathology* 11:69.

Weir, B. S., Johnston, P. R., and Damm, U. (2012). The *Colletotrichum gloeosporioides* species complex. *Stud. Mycol.* 73, 115–180. doi: 10.3114/sim0011

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* W1, W445–W451. doi: 10.1093/nar/gks479

Zhang, Q., Nizamani, M. M., Feng, Y., Yang, Y. Q., Jayawardena, R. S., Hyde, K. D., et al. (2023). Genome-scale and multi-gene phylogenetic analyses of *Colletotrichum* spp.

host preference and associated with medicinal plants. *Mycosphere* 14, 1–106. doi: 10.5943/mycosphere/14/si2/1

Zheng, I., Ge, Q., Yan, Y., Zhang, X., Huang, L., and Yin, Y. (2023). dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res.* W1, W115–W121. doi: 10.1093/nar/gkad328