Genomic streamlining of seagrass-associated *Colletotrichum* sp. may be related to its adaptation to a marine monocot host

Cassandra L. Ettinger[1,#], Jonathan A. Eisen[2,3,4], Jason E. Stajich[1,5]

[1]Department of Microbiology and Plant Pathology, University of California, Riverside, CA, United States
[2]Genome Center, University of California, Davis, CA, United States
[3]Department of Evolution and Ecology, University of California, Davis, CA, United States
[4]Department of Medical Microbiology and Immunology, University of California, Davis, Davis, CA, United States
[5]Institute for Integrative Genome Biology, University of California, Riverside, Riverside, CA, United States

[#]**Corresponding Author:** Cassandra L. Ettinger, cassande@ucr.edu

Keywords: *Zostera marina*, marine fungi, eelgrass, genome assembly, genome annotation, *Colletotrichum*, endophyte, plant-associated, comparative genomics, hemibiotroph

Abstract:

Colletotrichum spp. have a complicated history of association with land plants. Perhaps most well-known as plant pathogens for the devastating effect they can have on agricultural crops, some *Colletotrichum* spp. have been reported as beneficial plant endophytes. However, there have been only a handful of reports of *Colletotrichum* spp. isolated from aquatic plant hosts and their ecological role in the marine ecosystem is underexplored. To address this, we present the draft genome and annotation of *Colletotrichum* sp. CLE4, previously isolated from rhizome tissue from the seagrass *Zostera marina*. This genome (48.03 Mbp in length) is highly complete (BUSCO ascomycota: 98.8%) and encodes 12,015 genes, of which 5.7% are carbohydrate-active enzymes (CAZymes) and 12.6% are predicted secreted proteins. Phylogenetic placement puts *Colletotrichum* sp. CLE4 within the *C. acutatum* complex, closely related to *C. godetiae*. We found a 8.69% smaller genome size, 21.90% smaller gene count, and the absence of 591 conserved gene families in *Colletotrichum* sp. CLE4 relative to other members of the *C. acutatum* complex, suggesting a streamlined genome possibly linked to its specialized ecological niche in the marine ecosystem. Machine learning analyses using CATAStrophy on CAZyme domains predict this isolate to be a hemibiotroph, such that it has a biotrophic phase where the plant is kept alive during optimal environmental conditions followed by a necrotrophic phase where the fungi actively serves a pathogen. While future work is still needed to definitively tease apart the lifestyle strategy of *Colletotrichum* sp. CLE4, this study provides foundational insight and a high-quality genomic resource for starting to understand the evolutionary trajectory and ecological adaptations of marine-plant associated fungi.

Introduction:

Colletotrichum is a diverse genus of plant-associated fungi well known as both pathogens and endophytes of terrestrial plants [1]. Many *Colleotrichum* exhibit a complex hemibiotrophic lifestyle, meaning they have an initial biotrophic phase where the plant host is kept alive, followed by a necrotrophic phase where the fungi actively harms host tissues [2]. During this necrotrophic phase, *Colleotrichum* spp. cause a significant number of diseases, known as anthracnose, in many agricultural crops worldwide, and thus has been named one of the ten most important fungal pathogens [3]. As a result of their complex lifestyle *Colleotrichum* spp. are highly adaptable, associating with a large host range of over 3,200 species of monocot and eudicot plants [4]. While some *Colleotrichum* species have high host specificity, including one-to-one associations, others can infect a wide variety of plant hosts [2,4–7]. Evolutionary analyses suggest that the ancestor of *Colletotrichum* diverged in parallel with the diversification of flowering plants on land, likely beginning with an association with eudicot plants before adapting to other host types [6].

While *Colletotrichum* spp. are predominantly known for their associations with land plants, there have been recent reports documenting their presence as endophytes of aquatic plants [8–10]. Notably, *Colletotrichum* species have been isolated as endophytes in seagrasses, including the ecologically important species *Zostera marina* [11]. *Z. marina* is an early diverging marine monocot that serves as a foundation species in coastal ecosystems across the Northern Hemisphere, with critical roles providing habitat, stabilizing sediment, and contributing to carbon sequestration [12–14].

Previous amplicon-based surveys and culture-dependent studies have reported that *Colletotrichum* spp. are abundant members of the fungal community associated with *Z. marina*, particularly on leaf tissues [11,15]. These studies have found *Colletotrichum* spp. to be dominant on and within healthy *Z. marina* leaves and present in rhizomes, suggesting a possible endophytic relationship. Furthermore, global mycobiome surveys of *Z. marina* have predicted *Colletotrichum* spp. members to be dispersal-limited and exhibit patterns of endemism to specific locations, such as California and Japan [16]. Additionally, *Colletotrichum* spp. have been isolated from leaves and rhizomes of another seagrass species, *Thalassia testudinum*, further supporting their potential role as endophytes in marine environments [17,18].

Given the pathogenic potential of *Colletotrichum* spp. in terrestrial plants, where they cause black lesions characteristic of anthracnose, it is crucial to understand their ecological role in seagrass environments. While no true fungi have yet been reported to cause widespread disease in seagrasses [19], *Colletotrichum* spp. lesions might appear morphologically similar to and be mistaken for those caused by the heterokont pathogen *Labyrinthula zosterae*, which is responsible for seagrass wasting disease [20]. As climate change continues to impact marine ecosystems, it is critical to understand the role of fungi like *Colletotrichum* sp. CLE4 in seagrass health and disease dynamics.

To start to investigate the ecology of seagrass-associated *Colletotrichum* species, we generated a draft genome and annotation for *Colletotrichum* sp. CLE4, previously isolated as an endophyte from the seagrass *Zostera marina* in Ettinger & Eisen [11]. We used this genome to refine taxonomic understanding of this isolate through whole-genome phylogenetic placement among close relatives. We further conducted comparative genomic analyses to identify genes that might have been gained or lost during adaptation to a marine monocot host, leveraging the genome annotation to explore potential ecological roles of *Colletotrichum* sp. CLE4 in the marine environment.

Methods:

Molecular methods

Colletotrichum sp. CLE4 was previously isolated from healthy *Z. marina* rhizome tissues collected in May 2018 from Bodega Bay, CA using Potato Dextrose Agar with 0.45 μM Millipore filtered natural aged seawater as described in Ettinger & Eisen [11] (Figure 1A). Briefly, in that work, the isolate was propagated on the same solid media and DNA was extracted from tissue using a MoBio PowerSoil DNA Extraction kit. The isolate was then identified through phylogenetic analysis using ITS-LSU regions obtained through Sanger sequencing (GenBank Accession: MN543905). In this work, that same DNA was provided to the UC Davis Genome Center DNA Technologies Core for genomic library preparation and sequencing. DNA libraries were sequenced on an Illumina HiSeq4000 to generate 150 bp paired-end reads.

Assembly and annotation

Reads were assembled using the Automatic Assembly of the Fungi (AAFTF) pipeline v. 0.2.5 [21]. This pipeline trims and filters reads using BBTools v. 38.95 [22]. Then, AAFTF assembles these trimmed reads with SPAdes v. 3.14.1 [23] using default parameters. AAFTF screens the resulting assembly for contaminant vectors using BLAST and then uses sourmash v. 3.5.0 [24] to identify and remove any additional contaminant contigs. AAFTF then identifies duplicate contigs for removal using Minimap2 v. 2.17 [25]. Finally, AAFTF runs Pilon v. 1.22 [26] with three rounds of polishing to produce short-read corrected contigs in the assembly.

Repetitive regions were identified and masked prior to genome annotation using RepeatModeler v. 2.0.1 [27] and RepeatMasker v. 4-1-1 [28] with default options to produce a *de novo* library of elements plus the elements from Fungi in Repbase to mask [29]. The repeat landscape of *Colletotrichum* sp. CLE4 was visualized in R v. 4.3.0 [30] using the tidyverse v. 2.0.0 [31]. We then used the Funannotate pipeline v. 1.8.8 to annotate the draft *Colletotrichum* sp. CLE4 genome assembly [32]. Funannotate uses a combination of software to predict gene models including Augustus v. 3.3.3, GlimmerHMM v. 3.0.4, GeneMark-ETS v. 4.62, and SNAP v. 2013_11_29 [33–37], and produces consensus gene model predictions using EVidenceModeler v. 1.1.1 [38]. Funannotate additionally predicts tRNAs using tRNAscan v. 1.3.1 [39]. Funannotate then annotates consensus gene models based on similarity to Pfam-A v. 35.0 [40] and dbCAN v. 9.0 [41,42] using HMMER v.3 [43] and similarity to MEROPS v. 12.0 [44], eggNOG v. 2.1.9 [45], InterProScan v. 5.51-85.0 [46], and UniProt v. 2022_05 [47] using diamond BLASTP v. 2.0.8 [48]. Additionally, Funannotate uses Phobius v. 1.01 [49] to predict transmembrane proteins and SignalP v. 5.0b [50] to predict secreted proteins. AntiSMASH v. 6.1.1 was used to further identify biosynthetic gene clusters [51]. EffectorP v. 3.0 was run on predicted secreted proteins to predict plant effectors [52].

The draft assembly and predicted gene models were assessed for completion using BUSCO v. 5.0.0 [53] in ‘genome’ and ‘protein’ mode, respectively, with the eukartyota_odb10, fungi_odb10 and ascomycota_odb10 sets. To assess genome size and ploidy, we used jellyfish v. 2.3.0 [54] with a *k*-mer size of 21 to produce a *k*-mer frequency histogram, which we supplied to GenomeScope v. 2.0 [55] to predict haploid genome size and heterozygosity.

We used CATAStrophy v. 0.1.0 [56], a classification method based on carbohydrate-active enzyme (CAZyme) patterns from filamentous fungal plant pathogens, to predict the possible lifestyle strategy of *Colletotrichum* sp. CLE4. CATAStrophy was run in a Google Collab implementation using dbCAN v. 10 [41,42].

Comparative genomics

Predicted gene models from annotated genomes were downloaded from either NCBI or JGI for use in comparative analyses. Completion of downloaded predicted gene models was assessed using BUSCO v. 5.0.0 in 'protein' mode with the eukartyota_odb10, fungi_odb10 and ascomycota_odb10 sets. To be included in analysis, the annotated protein sets needed to have >90% completion. In total the final dataset represented 110 annotated genomes [6,57–99] (Table S1). Briefly in addition to *Colletotrichum* sp. CLE4, the dataset included an in-group of 60 *Colletotrichum* genomes and an outgroup representing 49 genomes across six taxonomic orders including Diaporthales (n=2), Glomerellales (n=2), Hypocreales (n=18), Ophiostomatales (n=2), Sordariales (n=5), and Xylariales (n=20). Downloaded gene models were annotated using InterProScan v. 5.51-85.0 [46].

Phylogenomic placement of *Colletotrichum* sp. CLE4 was performed using the PHYling_unified (https://github.com/stajichlab/PHYling_unified) pipeline to generate a protein alignment of all species in the final dataset. This pipeline utilizes HMMER v.3 [43] and ClipKIT [100] to search for, build, and trim an alignment based on the BUSCO fungi_odb10 gene set. A maximum likelihood phylogeny was built from this alignment using IQ-TREE2 v.2.2.6 [101], with the -p option to indicate gene partitions [102] and the -m option to run ModelFinder Plus which identifies the optimal evolutionary model for each partition based on BIC [103]. The resulting phylogenetic tree was imported into R and visualized using ggtree v. 3.8.2 [104]. We also used fastANI v. 1.33 to compare the average nucleotide identity of the draft *Colletotrichum* sp. CLE4 genome to the genome of its closest sister taxa based on the whole genome phylogeny.

Phylogenetic hierarchical orthogroups (HOGs) were identified using OrthoFinder v. 2.5.4 [105]. We focused analyses on the phylogenetic node containing all *Colletotrichum* spp. and then compared orthogroup detection and frequency between *Colletotrichum* sp. CLE4 and other members of the *C. actuatum* clade. HOGs were visualized in R using the UpSetR v. 1.4.0 [106] and pheatmap v. 1.0.12 packages [107].

Results

Genome structure, repeat landscape, and functional potential of Colletotrichum sp. CLE4

The genome of *Colletotrichum* sp. CLE4 was 48.03 Mbp in total length with 408x coverage, distributed across 168 contigs with an N50 of 506,655 bp and an L50 of 32, indicating a relatively contiguous assembly (Table 1). The genome is haploid, with a predicted genome size based on k-mer frequency profiles of 48.3 Mbp (Figure 1B). BUSCO estimates for the genome using the fungi_odb10 dataset reveal that it is highly complete with 98.8% of the expected single-copy orthologs present and complete. Only 0.3% of the BUSCO genes were fragmented, and 0.9% were missing. Based on these results, we believe that the genome of *Colletotrichum* sp. CLE4 represents a high-quality resource for understanding the ecology and evolution of this isolate.

Repeat content in the genome was relatively low, representing only 2.87% of the total genome, with LTR and unknown elements being most prevalent (Figure 1C). The genome repeat landscape indicates that elements have accumulated gradually through time in this species and also exposes a possible historical expansion of repeat content, corresponding to ~11-12% divergence (Figure 1D). While similar patterns of LTR and unknown element expansion have been observed in *Colletotrichum* spp., the overall percent repeat content here is less than what has been reported in other species (e.g., 6.08% in *C. truncatum* [108], 5.86% in *C. incanum* [78]).

Annotation of the *Colletotrichum* sp. CLE4 genome using Funannotate identified a total of 12,015 gene models, including 11,678 mRNA genes and 337 tRNA genes, with 95.53% of gene models having EggNog database annotation hits (Table 1). Additionally, 683 (5.68%) genes were identified as having CAZyme domains (Figure S1). Approximately 11.97% of the genes were predicted to encode secreted proteins, including 469 predicted effectors, of which 141 were predicted to be cytoplasmic effectors and 328 were predicted to be apoplastic effectors.

Based on CAZyme content, CATAStrophy predicted that *Colletotrichum* sp. CLE4 was most likely a hemibiotroph, specifically an extracellular (non-appressorial) mesotroph. This classification is described as representing facultative biotrophic species that have longer latent periods than necrotrophs and that invade extracellular host tissues [56]. This classification group includes members that grow biotrophically under optimal environmental conditions, but under variable conditions can cause disease [109,110]. Thus, this assignment is consistent with potential for an opportunistic pathogenic lifestyle.

Whole-genome phylogenetic placement of Colletotrichum sp. CLE4 and genomic similarity to close relatives

Whole genome phylogenetic approaches place *Colletotrichum* sp. CLE4 in the *C. acutatum* complex (Figure 2), whose common ancestor was dated at 14.5 mya [6]. Within this complex, *Colletotrichum* sp. CLE4 is placed sister to *C. godetiae,* which is best known for causing disease in terrestrial eudicot plants [5,7]. Average nucleotide identity (ANI) between *Colletotrichum* sp. CLE4 and its sister *C. godetiae* was relatively high at 98.98%. While ANI to other members of its immediate clade were lower at 93.50% for *C. salcis* and 93.77% for *C. phormii*. Although ANI species boundaries in fungi have yet to be used extensively to delineate species boundaries, with such high similarity it's possible that *Colletotrichum* sp. CLE4 may represent a new marine strain of *C. godetiae* that infects monocot plants.

Colletotrichum sp. CLE4 has the smallest reported genome size and fewest number of gene models of any of the members in the *C. acutatum* complex looked at in this study. The genome size of *Colletotrichum* sp. CLE4 is 8.69% smaller than the average genome size in the *C. acutatum* complex (52.6 Mbp). While the gene content is 21.90% less in comparison to the average number of genes in the complex (15,383). Further, despite high ANI similarity, *Colletotrichum* sp. CLE4 has a genome size that is 7.02% smaller compared to the genome of its sister taxa *C. godetiae* (51.7 Mbp) and has 25.24% less gene content (16,071). Interestingly, the *Colletotrichum* sp. CLE4 genome is only slightly smaller than the only reported endophyte isolate in this complex, *C. fioriniae* [83], with only a 2.86% smaller genome size (49.4 Mbp in *C. fioriniae*) and a 1.34% smaller gene content (12,178 in *C. fioriniae*). Zooming out to the genus overall, while *Colletotrichum* sp. CLE4 is still among the smallest assemblies for genome size and number of gene models, it is not the smallest for either metric. Additionally, the average BUSCO completeness for the *C. acutatum* complex was 98.38% and for the *Colletotrichum* genus was 97.39%. Thus, while smaller in genome size and gene content, the draft genome assembly for *Colletotrichum* sp. CLE4 has a similarly high completion rate (98.8%).

Gene family reductions in Colletotrichum sp. CLE4

To further explore gene family gain or loss in *Colletotrichum* sp. CLE4, we performed OrthoFinder analysis, comparing *Colletotrichum* sp. CLE4 with other members of the genus *Colletotrichum* with a focus on comparisons to members of the *C. acutatum* complex. In total, we identified 38,833 phylogenetically hierarchical orthogroups (HOGs) among all *Colletotrichum* spp. and 23,298 HOGs among members of the *C. acutatum* complex.

Of these, 9197 HOGs were conserved across all members of the *C. acutatum* complex and *Colletotrichum* sp. CLE4 (Figure 3A). Interestingly, *Colletotrichum* sp. CLE4 appears to be missing 591 HOGs that are shared between all other members of the *C. acutatum* complex, which we infer to be most likely due to gene loss in CLE4. The main functional domains of the missing HOGs include hypothetical domains, transcription factors, transporters, cytochrome p450s, FAD-binding domains, and heterokaryon incompatibility protein domains (Figure 3B).

This set also included a handful of HOGs with CAZyme domains including GH3, GH43, GH76, GT25, GT15, and GT90.

In total, only 18 HOGs were exclusively present in *Colletotrichum* sp. CLE4 and absent in other members of the *C. acutatum* complex, many of which represented uncharacterized proteins (Figure 4). However, only two of these HOGs were exclusively present in *Colletotrichum* sp. CLE4 relative to all other *Colletotrichum* spp, an uncharacterized protein and a short-chain dehydrogenase (Figure S2). Seven HOGs were detected as shared with multiple members of the *C. gloeosporioides* complex, which includes many reported endophytic species. Two HOGs were predicted to be apoplastic plant effectors, representing a multicopper oxidase and a FAD-binding domain protein. Further and of particular note, one HOG was predicted to be a NodB homology domain-containing protein.

Discussion

Genome and annotation provide critical resource for marine-plant fungi work

This study provides the first draft genome and annotation of a marine *Colletotrichum* sp. and is a valuable resource for future investigations into the evolution and ecology of this highly adaptable fungal genus. At 98.8% BUSCO completion, this genome is comparable to other members of the *C. acutatum* complex and the broader *Colletotrichum* genus. This high level of completeness indicates that the assembly accurately reflects the genetic content of *Colletotrichum* sp. CLE4 and can serve as a foundation for understanding the genomic basis of its adaptation to a marine environment.

Interestingly, the repeat content of the *Colletotrichum* sp. CLE4 genome (2.87%) is lower than that reported for other *Colletotrichum* species [78,108,111]. This could be due to the use of short-read sequencing, which may collapse repetitive regions, thereby underestimating repeat content. However, another study has reported a significant positive correlation between genome size and repeat content in *Colletotrichum* spp. [111]. Further, in *C. tanaceti*, higher repeat content was suggested as a possible mechanism for expansion of pathogenicity genes [89]. Further studies using long-read sequencing could help clarify the repeat landscape of *Colletotrichum* sp. CLE4 and explore its potential role in genomic streamlining and adaptation.

Phylogenetic placement indicates likely recent evolutionary association

Whole genome phylogenetic approaches place this isolate it in *C. acutatum* complex, which evolved 14.5 mya [6], and place it sister to *C. godetiae,* which is best known for causing disease in a broad range of terrestrial eudicot plants and having a global distribution [5,7]. Interestingly in multi-locus phylogenies, *C. godetiae* places sister to *C. lauri* (no publicly available genome) which has been reported once in association with neither a monocot or eudicot, but instead a magnoliid plant [7,112]. The evolutionary history of the *Colletotrichum* genus suggests an ancestral association with eudicot hosts, with subsequent diversification and independent adaptation to monocot hosts as flowering plants diversified [6]. Seagrasses, such as *Z. marina*, are early branching monocots whose ancestors recolonized the marine ecosystem 70 - 100 mya [113]. The more recent evolution of the *C. acutatum* complex suggests that *Colletotrichum* sp. CLE4's relationship with seagrass likely occurred after the return of the ancestor of *Z. marina* to the ocean, as opposed to co-evolving with *Z. marina*. This is similar to the timing of other monocot-host jumps in *C. orchidophilum* and *C. phormii*, which both are predicted to have transitioned at a date after the speciation of their host [6]. Ultimately, this means that the ancestor of *Colletotrichum* sp. CLE4 needed to adapt to both a monocot host and the marine ecosystem simultaneously.

Adaptations by seagrasses to the marine ecosystem may pose additional challenges as well. For example, seagrasses have lost all the required genes to form stomata [114], and their cell wall contains polyanionic, low-methylated pectins and sulfated galactans, in addition the polysaccharides typical of land plants [115]. The plant cell wall is considered the first line of plant defense against microbial invasion and the stomata are often how fungi initially invade plants [116,117]. These physiological modifications, as a result of adaptation to the marine ecosystem, would likely make it harder for *Colletotrichum* sp. CLE4 to invade and proliferate in *Z. marina*, possibly requiring a new or divergent invasion strategy relative to the strategies and genes utilized by close relatives for land plant colonization.

Genomic streamlining during adaptation to a marine monocot host

Comparative genomic analyses revealed smaller genome size, gene count, and absence of conserved gene families (i.e., HOGs) in *Colletotrichum* sp. CLE4 compared to other members of the *C. acutatum* complex. Specifically, we found that the genome size of *Colletotrichum* sp. CLE4 was 8.69% smaller, and its gene content was 21.90% lower than the average for the *C. acutatum* complex. This combined with the absence of 591 gene families that are conserved among all other *C acutatum* complex members leads us to conclude that the smaller genome, gene count and reduced orthologous groups may be the result of genome reduction and streamlining. Genome streamlining is a well-documented phenomenon in microbial adaptation to marine environments, as seen in bacteria [118], and has been recently described in fungi [119]. Streamlining removes non-essential genes and non-coding DNA to improve efficiency, often at the cost of metabolic versatility. For example, the marine fungus *Rhodotorula sphaerocarpa* exhibits a 10% smaller genome size compared to its terrestrial relatives, largely due to a decrease in transporter genes, particularly Major Facilitator Superfamily transporters, which are key for cross-membrane transport of organic solutes [119].

Genomic streamlining in *Colletotrichum* sp. CLE4 could also be linked to adaptation to a monocot host or an endophytic lifestyle. Pathogens often have expanded or unique secreted enzymes related to host-specialization and virulence in comparison to non-pathogens, and similar patterns have been reported for hemibiotrophic fungi compared to biotrophs [2,92,120–

122]. However, gene family differences in some studies have been reported to be more about relatedness than trophic lifestyle [123]. Previous work in *Colletotrichum* species have observed gains and losses of CAZyme and protease encoding genes in species that have a more narrow host range and have suggested that switching to a new host involves gene losses coupled with expansions in lineage-specific genes [6,61,78,124].

Colletotrichum sp. CLE4 had 683 CAZyme domain predictions, which is lower than most other *Colletotrichum* species. Monocot infecting *Colletotrichum* spp. are generally reported to have a smaller number of CAZymes compared to eudicot infecting species (741 vs. 867 on average) [6]. This supports that some gene content reduction may be due to specialization to a monocot host. However, *Colletotrichum* sp. CLE4 has a 7.82% smaller CAZyme content when compared to the average for other monocot infecting species, indicating that host specialization alone may not fully explain the extent of its smaller gene content.

Retention and loss of gene families provides functional insight into adaptation

In comparison to other *C. acutatum* complex members, *Colletotrichum* sp. CLE4 has lost 591 conserved gene families. While the majority of these had no predicted function, there was an enrichment in the loss of transcription factors, transporters, cytochrome p450s, and FAD-binding domains, as well as some specific CAZymes. The loss of gene families with transporter domain annotations, particularly Major Facilitator Superfamily transporters, is similar to the reports from genomic streamlining in marine *Rhodotorula* in response to marine adaptation [119].

While the reduction in transporters may relate to adaptation to the marine realm, the loss of other gene families, such as cytochrome P450s and CAZymes, could relate to specialization to a monocot host or endophytic lifestyle. It's been suggested that having a diverse set of cytochrome p450s play a role in the colonization success or virulence of plant pathogenic fungi [122,125]. In *Colletotrichum* spp., contractions in cytochrome p450 diversity have been suggested to relate to host range and specificity [6,61,126,127].

Similarly, the loss of certain CAZyme gene families further highlights the specialization of *Colletotrichum* sp. CLE4, potentially reflecting adaptations not just to marine life but also to the unique defenses of its monocot host and lifestyle strategy. For example, GH3, which can help detoxify plant antifungal commands, may be unnecessary depending on the specific defenses of *Z. marina* [128]*.* GH43 has been suggested to be important for plant-host interaction or plant tissue degradation in other *Colletotrichum* spp. [6] and expanded across distantly related pathogenic lineages [78]. While GH76 expansions in *Colletotrichum* have been associated with host-specificity towards woody plants [124].

A study comparing gene family differences between monocot and eudicot-infecting *Colletotrichum* species found that pathogenic eudicot species retained three unique gene families that were lost in monocot-infecting species, including a secreted β-glucosidase (GH3); a secreted protein with a FAD-binding domain, and an α-1,2-mannosidase (GH92) [6]. The loss of GH3 and FAD-binding domains in *Colletotrichum* sp. CLE4 aligns with this pattern, suggesting that these losses may relate to specialization to a monocot host.

In comparison to other *C. acutatum* complex members, *Colletotrichum* sp. CLE4 had only 18 unique gene families, of which only two were truly unique, while several others were shared with endophytic members of *C. gloeosporioides* species complex. Two gene families were predicted to be apoplastic plant effectors, representing a multicopper oxidase and a FAD-binding domain protein. While often associated with pathogenicity, effectors are also important for beneficial plant-fungal interactions [129–132] and for biotrophic lifestyles, where fungi may still need to suppress host defenses and evade recognition [133,134]. Further and of particular note, one gene family was predicted to be a NodB homology domain-containing protein. NodB genes are chitin deacetylases involved in the production of signaling molecules, most famously in legume-rhizobia symbiosis [135,136], and homologous signaling pathways have been used by symbiotic fungi [137]. The retention of gene families shared with *C. gloeosporioides* endophytes aligns with previous studies suggesting that such shared gene families across distant *Colletotrichum* species result from recent independent acquisitions or rapid losses during host specialization [2,6,61,78].

Pathogen or endophyte - deconvoluting a complex hemibiotrophic lifestyle

Colletotrichum spp. isolated from healthy, undamaged seagrass tissues, such as *Zostera marina*, have not been associated with anthracnose or other known pathogenic symptoms in marine plants, suggesting they may act primarily as endophytes in these environments. However, given the well-documented pathogenic potential of *Colletotrichum* spp. in terrestrial plants and their hemibiotrophic lifestyle, we used machine learning with CATAStrophy to predict the lifestyle of *Colletotrichum* sp. CLE4.

Perhaps unsurprisingly, CATAStrophy predicted that *Colletotrichum* sp. CLE4 is a hemibiotroph, specifically an extracellular mesotroph. This lifestyle involves a biotrophic phase, where the fungus maintains the host plant's viability under favorable conditions, followed by a necrotrophic phase under stress, where it acts as a pathogen. This prediction aligns with CATAStrophy's classification of most other *Colletotrichum* spp. as mesotrophs. This dual capacity for both benign and pathogenic behavior is similar to fungi like *Cladosporium fulvum*, which exhibits a biotrophic lifestyle under controlled, optimal conditions but can turn pathogenic in response to environmental stress [109,110].

The predicted potential of *Colletotrichum* sp. CLE4 to switch between endophytic and pathogenic roles suggests that it could remain a benign endophyte under stable conditions but become harmful under environmental stressors or when *Z. marina* is compromised. This flexibility supports a growing view of many fungi as multi-niche organisms that can form either beneficial or pathogenic associations depending on context [138]. Thus, while *Colletotrichum* sp. CLE4 likely functions as an endophyte in *Z. marina*, it may retain the genomic potential to become pathogenic under adverse conditions, highlighting the importance of environmental factors in shaping fungal-host dynamics, and a need for further work in understanding the exact nature of its ecological role when associated with *Z. marina.*

Conclusion

We report the first high quality draft genome and annotation of a marine monocot infecting *Colletotrichum* sp. The genome is near-complete (98.8%) and encodes 12,015 genes, of which 5.7% are CAZymes and 12.6% are predicted secreted proteins. Whole-genome phylogenetic analyses place *Colletotrichum* sp. CLE4 within the *C. acutatum* complex, most closely related to *C. godetiae*, which infects terrestrial eudicot plants. Overall, we found evidence of a streamlined genome, with an 8.69% reduction in genome size, 21.90% reduction in gene content, and a loss of 591 conserved gene families compared to other members of the *C. acutatum* complex. This streamlining is likely due to adaptation to both the marine ecosystem and a monocot host. We also identified unique gene families some of which were shared with members of the *C. gloeosporioides* complex which includes several endophytes, as well as NodB homology containing domain protein. Machine learning analyses predicted that *Colletotrichum* sp. CLE4 has an extracellular mesotroph lifestyle, which may indicate it still has capacity to serve as an opportunistic pathogen of *Z. marina*. This study provides a foundational insight into understanding the evolutionary trajectory and ecological adaptations of marine-plant associated *Colletotrichum* spp. Further work is needed, including challenge experiments and transcriptomics, to assess whether CLE4 is endophyte of ZM only under optimal conditions and whether new environmental stressors such as a changing climate might trigger opportunistic pathogenicity.

DNA Deposition

The raw sequences were deposited at GenBank under accession no. PRJNA1140278. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JBFUUI000000000. The version described in this paper is version JBFUUI010000000. All code used in this work has been deposited on Github (casett/ZM_Colletotrichum_sp_Genome) and archived in Zenodo (DOI: 10.5281/zenodo.14207532).

COI

Jonathan A. Eisen is on the Scientific Advisory Board of Zymo Research, Inc. Jason E. Stajich is a scientific consultant for Michroma, Inc.

Funding sources

The sequencing data in this work was generated by grants from the UC Davis H. A. Lewin Family Fellowship and the UC Davis Center for Population Biology to CLE. CLE was supported by the National Science Foundation (NSF) under a NSF Ocean Sciences Postdoctoral Fellowship (Award No. 2205744). JES is a CIFAR Fellow in the program Fungal Kingdom: Threats and Opportunities and was partially supported by NSF awards EF-2125066 and IOS-2134912. Computations were performed using the computer clusters and data storage resources of the UC Riverside HPCC, which were funded by grants from NSF (MRI-2215705, MRI-1429826) and NIH (1S10OD016290-01A1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

CLE conceived and designed the experiments, performed sampling, analyzed the data, prepared figures and/or tables, wrote and reviewed drafts of the paper. JAE and JES reviewed drafts of the paper.

Acknowledgements

We would like to thank the following undergraduate researchers for their contributions to this work including help with sample collection, fungal isolation and DNA extraction: Tess McDaniel, Katelin Jones, Katie Somers and Neil Brahmbhatt. We would also like to thank John J. Stachowicz for use of his scientific sampling permit, California Department of Fish and Wildlife Scientific Collecting Permit # SC 4874.

References

1. De Silva DD, Crous PW, Ades PK, Hyde KD, Taylor PWJ. Life styles of *Colletotrichum* species and implications for plant biosecurity. Fungal Biol Rev. 2017;31: 155–168. doi:10.1016/j.fbr.2017.05.001
2. O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, et al. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat Genet. 2012;44: 1060–1065. doi:10.1038/ng.2372
3. Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, et al. The Top 10 fungal pathogens in molecular plant pathology. Mol Plant Pathol. 2012;13: 414–430. doi:10.1111/j.1364-3703.2011.00783.x
4. Cannon PF, Damm U, Johnston PR, Weir BS. *Colletotrichum* - current status and future directions. Stud Mycol. 2012;73: 181–213. doi:10.3114/sim0014
5. Damm U, Cannon PF, Woudenberg JH, Crous PW. The *Colletotrichum acutatum* species complex. Stud Mycol. 2012;73. doi:10.3114/sim0010
6. Baroncelli R, Cobo-Díaz JF, Benocci T, Peng M, Battaglia E, Haridas S, et al. Genome evolution and transcriptome plasticity is associated with adaptation to monocot and dicot plants in *Colletotrichum* fungi. Gigascience. 2024;13. doi:10.1093/gigascience/giae036
7. Talhinhas P, Baroncelli R. *Colletotrichum* species and complexes: geographic distribution, host range and conservation status. Fungal Divers. 2021;110: 109–198. doi:10.1007/s13225-021-00491-9

8. Zheng H, Yu Z, Jiang X, Fang L, Qiao M. Endophytic *Colletotrichum* Species from Aquatic Plants in Southwest China. Journal of Fungi. 2022;8. doi:10.3390/jof8010087

9. Li J, Zheng H, Yu ZF, Qiao M. *Colletotrichum litangense* sp. nov., Isolated as an Endophyte of *Hippuris vulgaris*, an Aquatic Plant in Sichuan, China. Curr Microbiol. 2022;79. doi:10.1007/s00284-022-02846-0

10. Johnson DA, Carris LM, Rogers JD. Morphological and molecular characterization of *Colletotrichum nymphaeae* and *C. nupharicola* sp. nov. on water-lilies (*Nymphaea* and *Nuphar*). Mycol Res. 1997;101: 641–649. doi:10.1017/S0953756296003115

11. Ettinger CL, Eisen JA. Fungi, bacteria and oomycota opportunistically isolated from the seagrass, *Zostera marina*. PLoS One. 2020;15: e0236135. doi:10.1371/journal.pone.0236135

12. Hemminga MA, Duarte CM. Seagrass Ecology. Cambridge University Press; 2000. doi:10.1017/CBO9780511525551

13. Fourqurean JW, Duarte CM, Kennedy H, Marbà N, Holmer M, Mateo MA, et al. Seagrass ecosystems as a globally significant carbon stock. Nature Geoscience. 2012;5: 505–509. doi:10.1038/ngeo1477

14. Orth RJ, Carruthers TJB, Dennison WC, Duarte CM, Fourqurean JW, Heck KL, et al. A Global Crisis for Seagrass Ecosystems. BioScience. 2006;56: 987. doi:10.1641/0006-3568(2006)56[987:agcfse]2.0.co;2

15. Ettinger CL, Eisen JA. Characterization of the Mycobiome of the Seagrass, *Zostera marina*, Reveals Putative Associations With Marine Chytrids. Front Microbiol. 2019;10: 2476. doi:10.3389/fmicb.2019.02476

16. Ettinger CL, Vann LE, Eisen JA. Global Diversity and Biogeography of the Mycobiome. Appl Environ Microbiol. 2021;87: e0279520. doi:10.1128/AEM.02795-20

17. Rodriguez GM. Potential of fungal endophytes from *Thalassia testudinum* Bank ex K.D. koenig as producers of bioactive compounds. Master's thesis, University of Puerto Rico, Mayaguez. 2008.

18. Venkatachalam A, Thirunavukkarasu N, Trichur SS. Distribution and diversity of endophytes in seagrasses. Fungal Ecol. 2015;13: 60–65. doi:10.1016/j.funeco.2014.07.003

19. Sullivan BK, Trevathan-Tackett SM, Neuhauser S, Govers LL. Review: Host-pathogen dynamics of seagrass diseases under future global change. Mar Pollut Bull. 2018;134: 75–88. doi:10.1016/j.marpolbul.2017.09.030

20. Muehlstein LK, Porter D, Short FT. Labyrinthula Zosterae Sp. Nov., The Causative Agent of Wasting Disease of Eelgrass, *Zostera Marina*. Mycologia. 1991. doi:10.1080/00275514.1991.12025994

21. Stajich J, Palmer J. stajichlab/AAFTF: v0.2.5. 2021. doi:10.5281/zenodo.3437300

22. Bushnell B. BBTools. 2022. Available: https://sourceforge.net/projects/bbmap/

23. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo

Assembler. Curr Protoc Bioinformatics. 2020;70: e102. doi:10.1002/cpbi.102

24. Brown CT, Titus Brown C, Irber L. sourmash: a library for MinHash sketching of DNA. The Journal of Open Source Software. 2016;1: 27. doi:10.21105/joss.00027

25. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34: 3094–3100. doi:10.1093/bioinformatics/bty191

26. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9: e112963. doi:10.1371/journal.pone.0112963

27. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117: 9451–9457. doi:10.1073/pnas.1921046117

28. Smit AFA, Hubley R, Green P. RepeatMasker. 2013-2015. Available: http://www.repeatmasker.org

29. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 2015;6. doi:10.1186/s13100-015-0041-9

30. The R Project for Statistical Computing. Available: https://www.r-project.org/

31. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. Journal of Open Source Software. 2019;4: 1686. doi:10.21105/joss.01686

32. Palmer JM, Stajich J. Funannotate v1.8.1: Eukaryotic genome annotation. 2020. Available: http://doi.org/10.5281/zenodo.4054262

33. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 2006;34: W435–9. doi:10.1093/nar/gkl200

34. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. Genome Research. 2008;18: 1979–1990. doi:10.1101/gr.081612.108

35. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. Bioinformatics. 2004;20: 2878–2879. doi:10.1093/bioinformatics/bth315

36. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5: 59. doi:10.1186/1471-2105-5-59

37. Testa AC, Hane JK, Ellwood SR, Oliver RP. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. BMC Genomics. 2015;16: 170. doi:10.1186/s12864-015-1344-4

38. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9: R7. doi:10.1186/gb-2008-9-1-r7

39. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25: 955–964. doi:10.1093/nar/25.5.955

40. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42: D222–30. doi:10.1093/nar/gkt1223

41. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42: D490–5. doi:10.1093/nar/gkt1178

42. Huang L, Zhang H, Wu P, Entwistle S, Li X, Yohe T, et al. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Res. 2018;46: D516–D521. doi:10.1093/nar/gkx894

43. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7: e1002195. doi:10.1371/journal.pcbi.1002195

44. Rawlings ND, Barrett AJ, Bateman A. Using the MEROPS Database for Proteolytic Enzymes and Their Inhibitors and Substrates. Curr Protoc Bioinformatics. 2014;48: 1.25.1–33. doi:10.1002/0471250953.bi0125s48

45. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016;44: D286–93. doi:10.1093/nar/gkv1248

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

47. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. Methods Mol Biol. 2016;1374: 23–54. doi:10.1007/978-1-4939-3167-5_2

48. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12: 59–60. doi:10.1038/nmeth.3176

49. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004;338: 1027–1036. doi:10.1016/j.jmb.2004.03.016

50. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37: 420–423. doi:10.1038/s41587-019-0036-z

51. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39: W339–46. doi:10.1093/nar/gkr466

52. Sperschneider J, Dodds PN. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. Mol Plant Microbe Interact. 2022;35: 146–156. doi:10.1094/MPMI-08-21-0201-R

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

54. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27: 764–770. doi:10.1093/bioinformatics/btr011

55. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020;11: 1432. doi:10.1038/s41467-020-14998-3

56. Hane JK, Paxman J, Jones DAB, Oliver RP, de Wit P. “CATAStrophy,” a Genome-Informed Trophic Classification of Filamentous Plant Pathogens - How Many Different Types of Filamentous Plant Pathogens Are There? Front Microbiol. 2019;10: 3088. doi:10.3389/fmicb.2019.03088

57. Alkan N, Meng X, Friedlander G, Reuveni E, Sukno S, Sherman A, et al. Global aspects of pacC regulation of pathogenicity genes in *Colletotrichum gloeosporioides* as revealed by transcriptome analysis. Mol Plant Microbe Interact. 2013;26: 1345–1358. doi:10.1094/MPMI-03-13-0080-R

58. Amlacher S, Sarges P, Flemming D, van Noort V, Kunze R, Devos DP, et al. Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. Cell. 2011;146: 277–289. doi:10.1016/j.cell.2011.06.039

59. Baroncelli R, Sreenivasaprasad S, Sukno SA, Thon MR, Holub E. Draft Genome Sequence of *Colletotrichum acutatum* Sensu Lato (*Colletotrichum fioriniae*). Genome Announc. 2014;2. doi:10.1128/genomeA.00112-14

60. Baroncelli R, Sanz-Martín JM, Rech GE, Sukno SA, Thon MR. Draft Genome Sequence of *Colletotrichum sublineola*, a Destructive Pathogen of Cultivated Sorghum. Genome Announc. 2014;2. doi:10.1128/genomeA.00540-14

61. Baroncelli R, Amby DB, Zapparata A, Sarrocco S, Vannacci G, Le Floch G, et al. Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. BMC Genomics. 2016;17: 555. doi:10.1186/s12864-016-2917-6

62. Baroncelli R, Sukno SA, Sarrocco S, Cafà G, Le Floch G, Thon MR. Whole-Genome Sequence of the Orchid Anthracnose Pathogen *Colletotrichum orchidophilum*. Mol Plant Microbe Interact. 2018;31: 979–981. doi:10.1094/MPMI-03-18-0055-A

63. Becerra S, Baroncelli R, Boufleur TR, Sukno SA, Thon MR. Chromosome-level analysis of the genome reveals the unique characteristics of core and minichromosomes. Front Microbiol. 2023;14: 1129319. doi:10.3389/fmicb.2023.1129319

64. Franco MEE, Wisecaver JH, Arnold AE, Ju Y-M, Slot JC, Ahrendt S, et al. Ecological generalism drives hyperdiversity of secondary metabolite gene clusters in xylarialean

endophytes. New Phytol. 2022;233: 1317–1330. doi:10.1111/nph.17873

65. Mesny F, Miyauchi S, Thiergart T, Pickel B, Atanasova L, Karlsson M, et al. Genetic determinants of endophytism in the *Arabidopsis* root mycobiome. Nat Commun. 2021;12: 7227. doi:10.1038/s41467-021-27479-y

66. Norrbom AL, Rodriguez EJ, Steck GJ, Sutton BA, Nolazco N. New Species and Host Plants of *Anastrepha* (Diptera: Tephritidae) Primarily from Peru and Bolivia. 2015. doi:10.11646/zootaxa.4041.1.1

67. Rehner SA, Gazis R, Doyle VP, Vieira WAS, Campos PM, Shao J. Genome Resources for the *Colletotrichum gloeosporioides* Species Complex: 13 Tree Endophytes from the Neotropics and Paleotropics. Microbiology resource announcements. 2023;12. doi:10.1128/mra.01040-22

68. Rogério F, Boufleur TR, Ciampi-Guillardi M, Sukno SA, Thon MR, Ns MJ, et al. Genome Sequence Resources of *Colletotrichum truncatum, C. plurivorum, C. musicola,* and *C. sojae*: Four Species Pathogenic to Soybean (*Glycine max*). Phytopathology. 2020;110. doi:10.1094/PHYTO-03-20-0102-A

69. Hsieh D-K, Chuang S-C, Chen C-Y, Chao Y-T, Lu M-YJ, Lee M-H, et al. Comparative Genomics of Three Strains and Genetic Analysis Revealed Genes Involved in Fungal Growth and Virulence on Chili Pepper. Front Microbiol. 2022;13: 818291. doi:10.3389/fmicb.2022.818291

70. Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, et al. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. Nat Commun. 2016;7: 11362. doi:10.1038/ncomms11362

71. Crouch JA, Dawe A, Aerts A, Barry K, Churchill ACL, Grimwood J, et al. Genome Sequence of the Chestnut Blight Fungus *Cryphonectria parasitica* EP155: A Fundamental Resource for an Archetypical Invasive Plant Pathogen. Phytopathology. 2020;110. doi:10.1094/PHYTO-12-19-0478-A

72. Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, Pietro A D, et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science. 2007;317. doi:10.1126/science.1143708

73. David AS, Haridas S, LaButti K, Lim J, Lipzen A, Wang M, et al. Draft Genome Sequence of *Microdochium bolleyi*, a Dark Septate Fungal Endophyte of Beach Grass. Genome Announc. 2016;4. doi:10.1128/genomeA.00270-16

74. de Man TJ, Stajich JE, Kubicek CP, Teiling C, Chenthamara K, Atanasova L, et al. Small genome of the fungus *Escovopsis weberi*, a specialized disease agent of ant agriculture. Proc Natl Acad Sci U S A. 2016;113. doi:10.1073/pnas.1518501113

75. Fu FF, Hao Z, Wang P, Lu Y, Xue LJ, Wei G, et al. Genome Sequence and Comparative Analysis of *Colletotrichum gloeosporioides* Isolated from Liriodendron Leaves. Phytopathology. 2020;110. doi:10.1094/PHYTO-12-19-0452-R

76. Gan P, Tsushima A, Hiroyama R, Narusaka M, Takano Y, Narusaka Y, et al. *Colletotrichum shisoi* sp. nov., an anthracnose pathogen of *Perilla frutescens* in Japan: molecular

phylogenetic, morphological and genomic evidence. Sci Rep. 2019;9. doi:10.1038/s41598-019-50076-5

77. Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, et al. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytol. 2013;197: 1236–1249. doi:10.1111/nph.12085

78. Gan P, Narusaka M, Kumakura N, Tsushima A, Takano Y, Narusaka Y, et al. Genus-Wide Comparative Genome Analyses of *Colletotrichum* Species Reveal Specific Gene Family Losses and Gains during Adaptation to Specific Infection Lifestyles. Genome Biol Evol. 2016;8: 1467–1481. doi:10.1093/gbe/evw089

79. Gan P, Narusaka M, Tsushima A, Narusaka Y, Takano Y, Shirasu K. Draft Genome Assembly of *Colletotrichum chlorophyti*, a Pathogen of Herbaceous Plants. Genome Announc. 2017;5. doi:10.1128/genomeA.01733-16

80. Gao Q, Jin K, Ying SH, Zhang Y, Xiao G, Shang Y, et al. Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. PLoS Genet. 2011;7. doi:10.1371/journal.pgen.1001264

81. Gómez-Cortecero A, Harrison RJ, Armitage AD. Draft Genome Sequence of a European Isolate of the Apple Canker Pathogen *Neonectria ditissima*. Genome Announc. 2015;3. doi:10.1128/genomeA.01243-15

82. Haridas S, Wang Y, Lim L, Massoumi AS, Jackman S, Docking R, et al. The genome and transcriptome of the pine saprophyte *Ophiostoma piceae*, and a comparison with the bark beetle-associated pine pathogen *Grosmannia clavigera*. BMC Genomics. 2013;14. doi:10.1186/1471-2164-14-373

83. Hill R, Levicky Q, Pitsillides F, Junnonen A, Arrigoni E, Bonnin JM, et al. Tapping Culture Collections for Fungal Endophytes: First Genome Assemblies for Three Genera and Five Species in the Ascomycota. Genome Biol Evol. 2023;15. doi:10.1093/gbe/evad038

84. Huo J, Wang Y, Hao Y, Yao Y, Wang Y, Zhang K, et al. Genome Sequence Resource for *Colletotrichum scovillei*, the Cause of Anthracnose Disease of Chili. Mol Plant Microbe Interact. 2021;34. doi:10.1094/MPMI-03-20-0055-A

85. Khoulassa S, Elmoualij B, Benlyas M, Meziani R, Bouhlali EDT, Houria B, et al. High-Quality Draft Nuclear and Mitochondrial Genome Sequence of *Fusarium oxysporum* f. sp. albedinis strain 9, the Causal Agent of Bayoud Disease on Date Palm. Plant Dis. 2022;106. doi:10.1094/PDIS-01-22-0245-A

86. Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BP, et al. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog. 2011;7. doi:10.1371/journal.ppat.1002137

87. Steindorff AS, Ramada MH, Coelho AS, Miller RN, Pappas GJ, Ulhoa CJ, et al. Identification of mycoparasitism-related genes against the phytopathogen Sclerotinia sclerotiorum through transcriptome and expression profile analysis in *Trichoderma harzianum*. BMC Genomics. 2014;15. doi:10.1186/1471-2164-15-204

88. Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Pietro A D, et al.

Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature. 2010;464. doi:10.1038/nature08850

89. Lelwala RV, Korhonen PK, Young ND, Scott JB, Ades PK, Gasser RB, et al. Comparative genome analysis indicates high evolutionary potential of pathogenicity genes in *Colletotrichum tanaceti*. PLoS One. 2019;14: e0212248. doi:10.1371/journal.pone.0212248

90. Meng Y, Gleason ML, Zhang R, Sun G. Genome Sequence Resource of the Wide-Host-Range Anthracnose Pathogen *Colletotrichum siamense*. Mol Plant Microbe Interact. 2019;32. doi:10.1094/MPMI-01-19-0010-A

91. Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, et al. Widespread adenine N6-methylation of active genes in fungi. Nat Genet. 2017;49. doi:10.1038/ng.3859

92. Morales-Cruz A, Amrine KC, Blanco-Ulate B, Lawrence DP, Travadon R, Rolshausen PE, et al. Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. BMC Genomics. 2015;16. doi:10.1186/s12864-015-1624-z

93. Proctor RH, McCormick SP, Kim HS, Cardoza RE, Stanley AM, Lindo L, et al. Evolution of structural diversity of trichothecenes, a family of toxins produced by plant pathogenic and entomopathogenic fungi. PLoS Pathog. 2018;14. doi:10.1371/journal.ppat.1006946

94. Terfehr D, Dahlmann TA, Specht T, Zadra I, Kürnsteiner H, Kück U. Genome Sequence and Annotation of *Acremonium chrysogenum*, Producer of the β-Lactam Antibiotic Cephalosporin C. Genome Announc. 2014;2. doi:10.1128/genomeA.00948-14

95. Tsushima A, Gan P, Kumakura N, Narusaka M, Takano Y, Narusaka Y, et al. Genomic Plasticity Mediated by Transposable Elements in the Plant Pathogenic Fungus *Colletotrichum higginsianum*. Genome Biol Evol. 2019;11. doi:10.1093/gbe/evz087

96. Xiao G, Ying SH, Zheng P, Wang ZL, Zhang S, Xie XQ, et al. Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. Sci Rep. 2012;2. doi:10.1038/srep00483

97. Yang D, Pomraning K, Kopchinskiy A, Karimi AR, Atanasova L, Chenthamara K, et al. Genome Sequence and Annotation of *Trichoderma parareesei*, the Ancestor of the Cellulase Producer *Trichoderma reesei*. Genome Announc. 2015;3. doi:10.1128/genomeA.00885-15

98. Zheng P, Xia Y, Xiao G, Xiong C, Hu X, Zhang S, et al. Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional Chinese medicine. Genome Biol. 2011;12. doi:10.1186/gb-2011-12-11-r116

99. Zampounis A, Pigné S, Dallery J-F, Wittenberg AHJ, Zhou S, Schwartz DC, et al. Genome Sequence and Annotation of *Colletotrichum higginsianum*, a Causal Agent of Crucifer Anthracnose Disease. Genome Announc. 2016;4. doi:10.1128/genomeA.00821-16

100. Steenwyk JL, Buida TJ 3rd, Li Y, Shen X-X, Rokas A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. 2020;18: e3001007. doi:10.1371/journal.pbio.3001007

101. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37: 1530–1534. doi:10.1093/molbev/msaa015

102. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 2016;65: 997–1008. doi:10.1093/sysbio/syw037

103. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14: 587–589. doi:10.1038/nmeth.4285

104. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8: 28–36. doi:10.1111/2041-210X.12628

105. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20: 238. doi:10.1186/s13059-019-1832-y

106. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33: 2938–2940. doi:10.1093/bioinformatics/btx364

107. Kolde R. pheatmap: Pretty Heatmaps. 2018. Available: https://github.com/raivokolde/pheatmap

108. Rao S, Sharda S, Oddi V, Nandineni MR. The Landscape of Repetitive Elements in the Refined Genome of Chilli Anthracnose Fungus *Colletotrichum truncatum*. Front Microbiol. 2018;9. doi:10.3389/fmicb.2018.02367

109. de Wit PJGM. *Cladosporium fulvum* Effectors: Weapons in the Arms Race with Tomato. Annu Rev Phytopathol. 2016;54: 1–23. doi:10.1146/annurev-phyto-011516-040249

110. Collemare J, Griffiths S, Iida Y, Karimi Jashni M, Battaglia E, Cox RJ, et al. Secondary metabolism and biotrophic lifestyle in the tomato pathogen *Cladosporium fulvum*. PLoS One. 2014;9: e85877. doi:10.1371/journal.pone.0085877

111. Wang H, Huang R, Ren J, Tang L, Huang S, Chen X, et al. The evolution of mini-chromosomes in the fungal genus. MBio. 2023;14: e0062923. doi:10.1128/mbio.00629-23

112. Hyde KD, Norphanphoun C, Abreu VP, Bazzicalupo A, Thilini Chethana KW, Clericuzio M, et al. Fungal diversity notes 603–708: taxonomic and phylogenetic notes on genera and species. Fungal Divers. 2017;87: 1–235. doi:10.1007/s13225-017-0391-3

113. Les DH, Cleland MA, Waycott M. Phylogenetic Studies in Alismatidae, II: Evolution of Marine Angiosperms (Seagrasses) and Hydrophily. Syst Bot. 1997;22: 443. doi:10.2307/2419820

114. Olsen JL, Rouzé P, Verhelst B, Lin YC, Bayer T, Collen J, et al. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. Nature. 2016;530. doi:10.1038/nature16548

115. Pfeifer L. The Cell Wall of Seagrasses: Fascinating, Peculiar and a Blank Canvas for Future Research. Front Plant Sci. 2020;11. doi:10.3389/fpls.2020.588754

116. Malinovsky FG, Fangel JU, Willats WGT. The role of the cell wall in plant immunity. Front Plant Sci. 2014;5. doi:10.3389/fpls.2014.00178

117. Wu J, Liu Y. Stomata-pathogen interactions: over a century of research. Trends Plant Sci. 2022;27. doi:10.1016/j.tplants.2022.07.004

118. Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, et al. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci U S A. 2013;110: 11463–11468. doi:10.1073/pnas.1304246110

119. Lane DM, Valentine DL, Peng X. Genomic analysis of the marine yeast *Rhodotorula sphaerocarpa* ETNP2018 reveals adaptation to the open ocean. BMC Genomics. 2023;24. doi:10.1186/s12864-023-09791-7

120. Zhao Z, Liu H, Wang C, Xu JR. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics. 2013;14. doi:10.1186/1471-2164-14-274

121. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal effectors and plant susceptibility. Annu Rev Plant Biol. 2015;66. doi:10.1146/annurev-arplant-043014-114623

122. Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, et al. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS One. 2008;3. doi:10.1371/journal.pone.0002300

123. Krijger JJ, Thon MR, Deising HB, Wirsel SG. Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. BMC Genomics. 2014;15. doi:10.1186/1471-2164-15-722

124. Meng F, Tian C. Gene Family Expansion during the Adaptation of *Colletotrichum gloeosporioides* to Woody Plants. Journal of Fungi. 2023;9. doi:10.3390/jof9121185

125. Li L, Zhu XM, Zhang YR, Cai YY, Wang JY, Liu MY, et al. Research on the Molecular Interaction Mechanism between Plants and Pathogenic Fungi. Int J Mol Sci. 2022;23. doi:10.3390/ijms23094658

126. Ortiz-Álvarez J, Becerra S, Baroncelli R, Hernández-Rodríguez C, Sukno SA, Thon MR. Evolutionary history of the cytochrome P450s from *Colletotrichum* species and prediction of their putative functional roles during host-pathogen interactions. BMC Genomics. 2024;25: 56. doi:10.1186/s12864-023-09858-5

127. Liang X, Wang B, Dong Q, Li L, Rollins JA, Zhang R, et al. Pathogenic adaptations of *Colletotrichum* fungi revealed by genome wide gene family evolutionary analyses. PLoS One. 2018;13: e0196303. doi:10.1371/journal.pone.0196303

128. Bradley EL, Ökmen B, Doehlemann G, Henrissat B, Bradshaw RE, Mesarich CH. Secreted Glycoside Hydrolase Proteins as Effectors and Invasion Patterns of Plant-Associated Fungi and Oomycetes. Front Plant Sci. 2022;13: 853106. doi:10.3389/fpls.2022.853106

129. Wawra S, Fesel P, Widmer H, Timm M, Seibel J, Leson L, et al. The fungal-specific β-glucan-binding lectin FGB1 alters cell-wall composition and suppresses glucan-triggered immunity in plants. Nat Commun. 2016;7. doi:10.1038/ncomms13188

130. Martin F, Aerts A, Ahrén D, Brun A, Danchin EG, Duchaussoy F, et al. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature. 2008;452. doi:10.1038/nature06556

131. Plett JM, Kemppainen M, Kale SD, Kohler A, Legué V, Brun A, et al. A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. Curr Biol. 2011;21. doi:10.1016/j.cub.2011.05.033

132. Voß S, Betz R, Heidt S, Corradi N, Requena N. RiCRN1, a Crinkler Effector From the Arbuscular Mycorrhizal Fungus *Rhizophagus irregularis*, Functions in Arbuscule Development. Front Microbiol. 2018;9. doi:10.3389/fmicb.2018.02068

133. Koeck M, Hardham AR, Dodds PN. The role of effectors of biotrophic and hemibiotrophic fungi in infection. Cell Microbiol. 2011;13. doi:10.1111/j.1462-5822.2011.01665.x

134. Tanaka S, Kahmann R. Cell wall-associated effectors of plant-colonizing fungi. Mycologia. 2021;113. doi:10.1080/00275514.2020.1831293

135. Schmidt J, Röhrig H, John M, Wieneke U, Stacey G, Koncz C, et al. Alteration of plant growth and development by Rhizobium nodA and nodB genes involved in the synthesis of oligosaccharide signal molecules. Plant J. 1993;4: 651–658. doi:10.1046/j.1365-313X.1993.04040651.x

136. John M, Röhrig H, Schmidt J, Wieneke U, Schell J. Rhizobium NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. Proc Natl Acad Sci U S A. 1993;90: 625. doi:10.1073/pnas.90.2.625

137. Oldroyd GED. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nat Rev Microbiol. 2013;11: 252–263. doi:10.1038/nrmicro2990

138. Selosse M-A, Schneider-Maunoury L, Martos F. Time to re-think fungal ecology? Fungal ecological niches are often prejudged. New Phytol. 2018;217: 968–972. doi:10.1111/nph.14983

Figure 1. General characteristics of *Colletotrichum* sp. CLE4. (A) Photograph of *Colletotrichum* sp. CLE4. (B) GenomeScope profile depicting the *k*-mer frequency histogram used to calculate genome size, ploidy and heterozygosity. (C) A bar plot representing the percent of the genome composed of repetitive elements from each repeat class. (D) A stacked bar plot representing the percent of the genome made of repeat elements from each repeat class binned by 1% sequence divergence (CpG adjusted Kimura divergence). For (C) and (D) bars are colored repeat class (LINE = pink, SINE = orange, LTR = green, DNA = light blue, and Unknown = yellow). Abbreviations: long-interspersed nuclear element (LINE), small-interspersed nuclear element (SINE), long-terminal repeat retrotransposon (LTR), and DNA transposons (DNA).

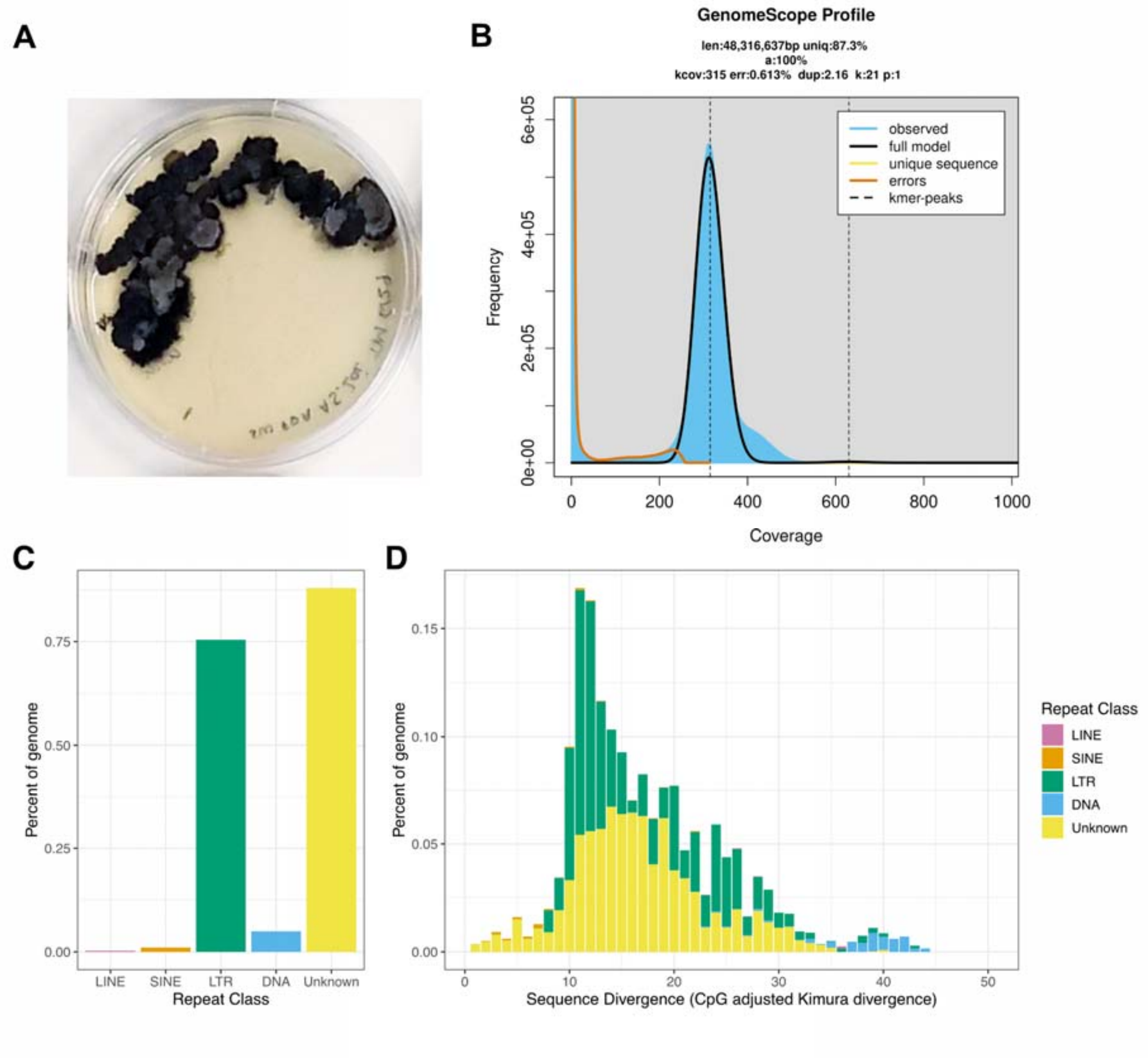


Figure 2. Phylogenetic placement and genomic content comparison to other *Colletotrichum* spp. From left to right, first, a maximum likelihood phylogeny that shows the relationship of *Colletotrichum* sp. CLE4 to other *Colletotrichum* spp. lineages. This tree was generated using IQ-TREE2 on an alignment of BUSCO fungi_odb10 HMMs constructed using the PHYling_unified pipeline. Taxon labels in the phylogeny are shown colored by their assigned *Colletotrichum* species complex. Next (from left to right), in association with this phylogeny, a bar chart of BUSCO "protein" completion status for the ascomycota_odb10 set is shown. Bars show the percentage of genes found in each genome annotation as a percentage of the total gene set and are colored by BUSCO status (missing = gray, fragmented = yellow, complete and duplicated = green, complete and single copy = blue). Next is a bar chart of predicted gene counts for each taxon with counts colored by fungal host plant ecotype reported during isolate deposition (eudicot = purple, gymnosperm = pink, magnoliid = orange, monocot = yellow). Finally, there is a bar chart of the draft genome size (Mbp) for each taxon with genome size colored by fungal guild (endophyte = black, plant pathogen = grey).

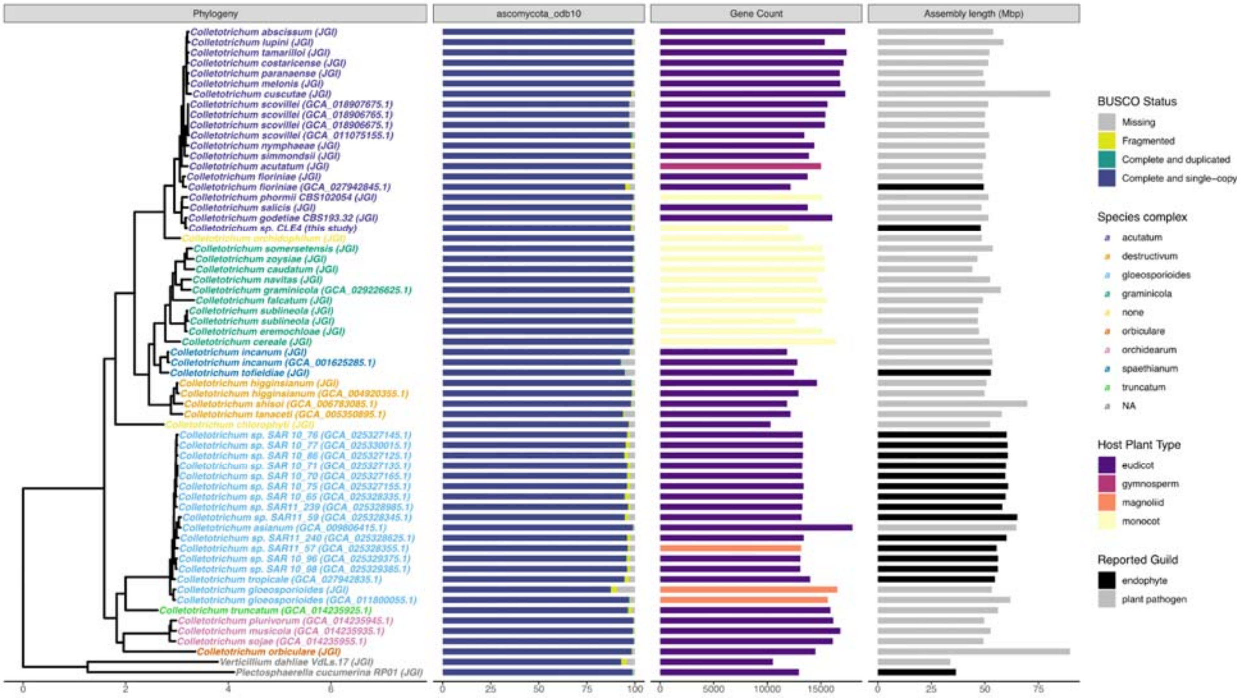


Figure 3. *Colletotrichum* sp. CLE4 is missing many *C. acutatum* complex conserved phylogenetically hierarchical orthogroups (HOGs). (A) UpSet plot depicting shared HOGs within the *C. acutatum* complex with *Colletotrichum* spp. organized by their phylogenetic relationship from Figure 2. Vertical bars display the counts of HOGs in each group while the dots indicate the organism present in each group. Horizontal bars display the number of HOGs in the genome of each organism. We have highlighted three groups of HOGs: unique HOGs to *Colletotrichum* sp. CLE4 (blue), conserved HOGs shared by the *C. acutatum* complex (purple), and conserved HOGs missing in *Colletotrichum* sp. CLE4, but shared by the rest of the *C. acutatum* complex (yellow). (B) Bar plot depicting the distribution of PFAM annotations across the conserved HOGs missing in *Colletotrichum* sp. CLE4 but shared by the rest of the *C. acutatum* complex. To simplify visualization, only annotations that were observed across ≥ 4 HOGs are depicted.

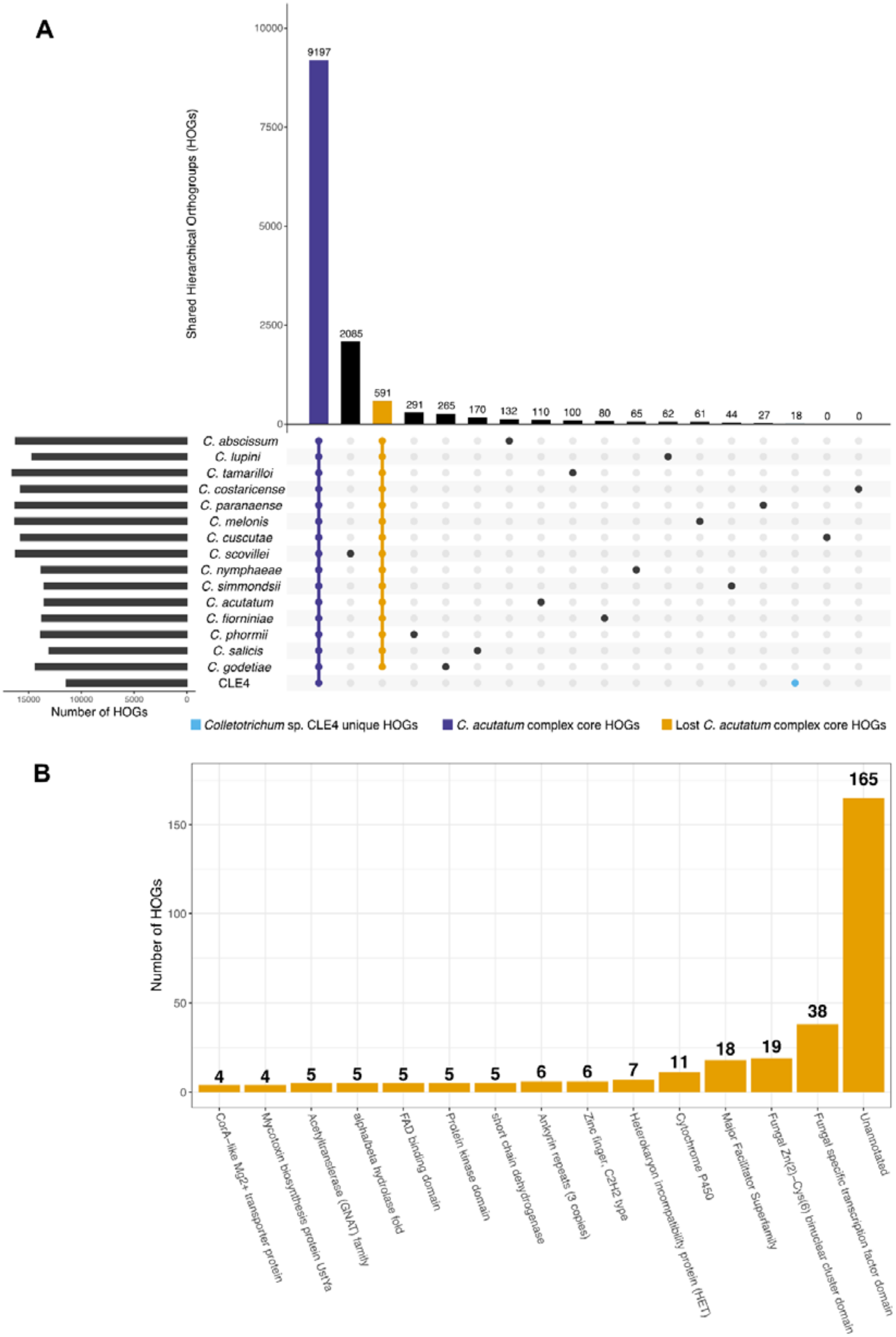

A
Shared Hierarchical Orthogroups (HOGs)
10000
7500
5000
2500
0
9197
2085
591
291
265
170
132
110
100
80
65
62
61
44
27
18
0
0
C. abscissum
C. lupini
C. tamarilloi
C. costaricense
C. paranaense
C. melonis
C. cuscutae
C. scovillei
C. nymphaeae
C. simmondsii
C. acutatum
C. fiorniniae
C. phormii
C. salicis
C. godetiae
CLE4
15000
10000
5000
0
Number of HOGs
Colletotrichum sp. CLE4 unique HOGs
C. acutatum complex core HOGs
Lost C. acutatum complex core HOGs
B
Number of HOGs
150
100
50
0
165
38
19
18
11
7
6
6
5
5
5
5
5
4
4
CorA-like Mg2+ transporter protein
Mycotoxin biosynthesis protein UstYa
Acetyltransferase (GNAT) family
alpha/beta hydrolase fold
FAD binding domain
Protein kinase domain
short chain dehydrogenase
Ankyrin repeats (3 copies)
Zinc finger, C2H2 type
Heterokaryon incompatibility protein (HET)
Cytochrome P450
Major Facilitator Superfamily
Fungal Zn(2)-Cys(6) binuclear cluster domain
Fungal specific transcription factor domain
Unannotated

Figure 4. Majority of unique HOGs shared with endophytic *C. gloeosporiodes* complex. A heatmap is depicted here visualizing the copy number of phylogenetically hierarchical orthogroups (HOGs) that were unique to *Colletotrichum* sp. CLE4 relative to other members of the *C. actutatum* complex. Further, the dendrogram to the left of the heatmap clusters the HOGs by similarity in counts between the different HOGs, while the dendrogram above clusters genomes based on similarity in counts between the different genomes.

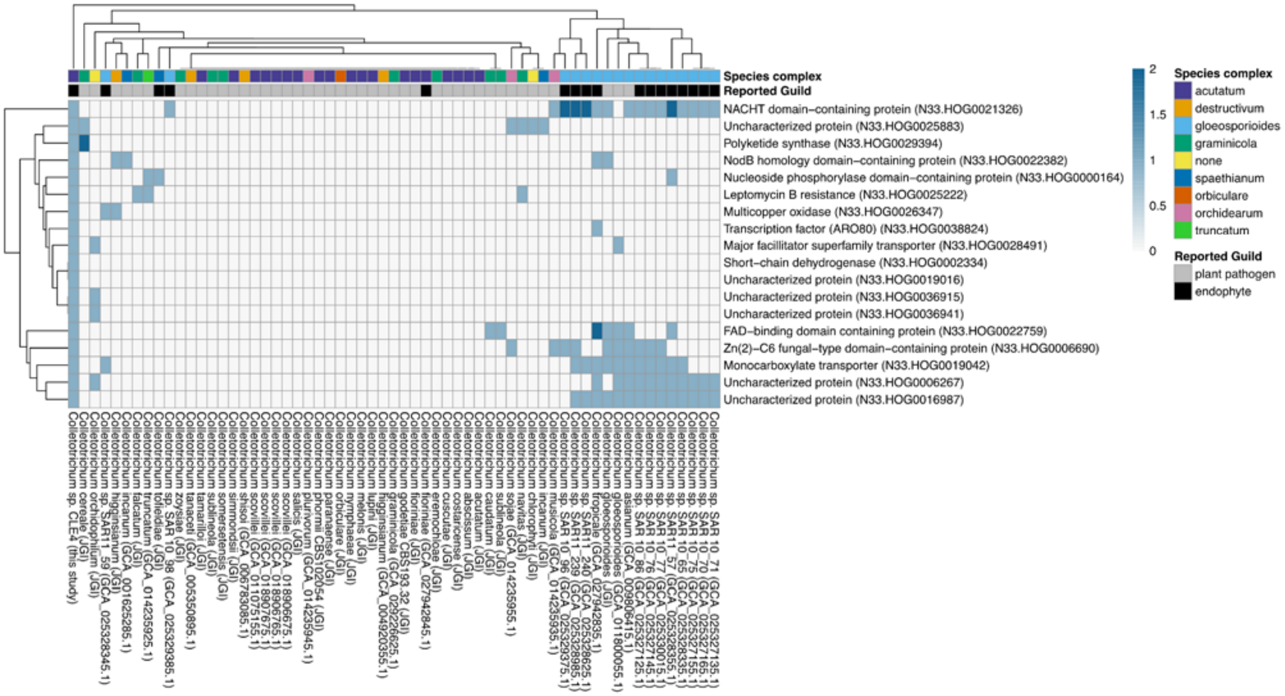
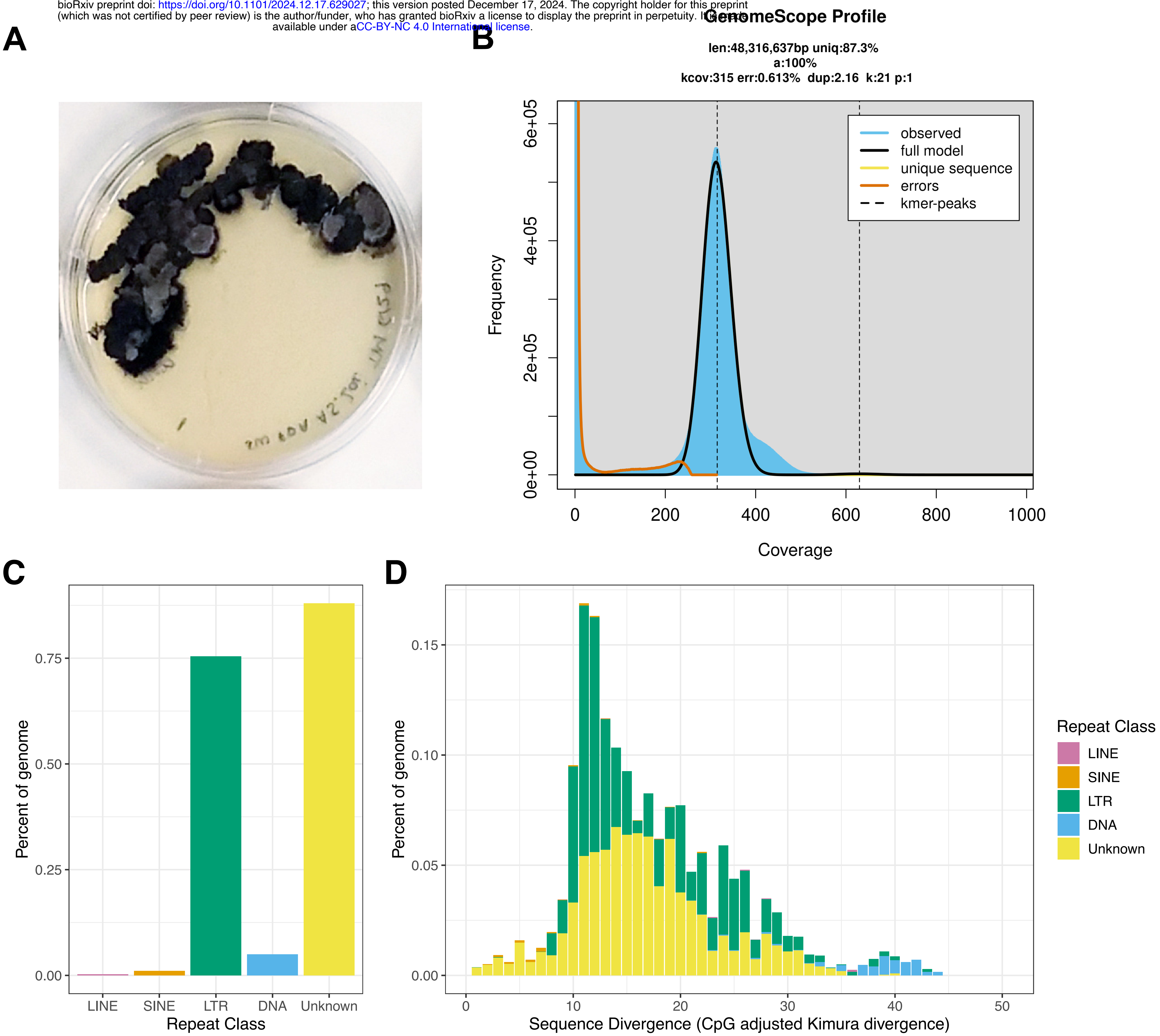


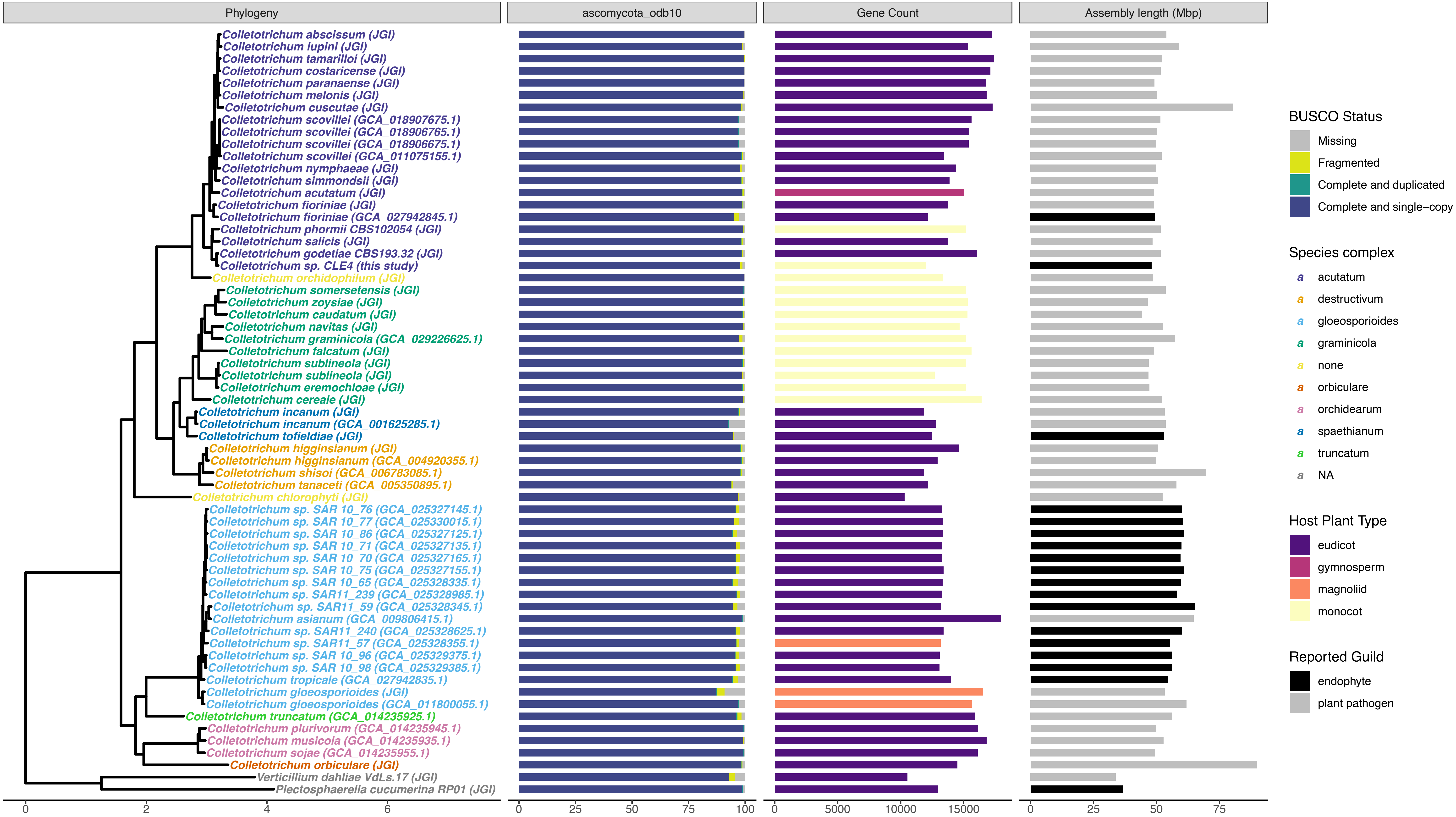
Table 1. Genome and annotation statistics for *Colletotrichum* sp. CLE4. Here we report various AAFTF assembly and Funannotate annotation statistics including the number of contigs in the assembly, the number of contigs of various lengths, the total assembly length, percent GC, N50, L50, number of gene models, percent of genes annotated with different databases, and the number of secreted genes with EffectorP predictions. We also report here the results of the BUSCO assessment using the fungi_odb10 gene set.

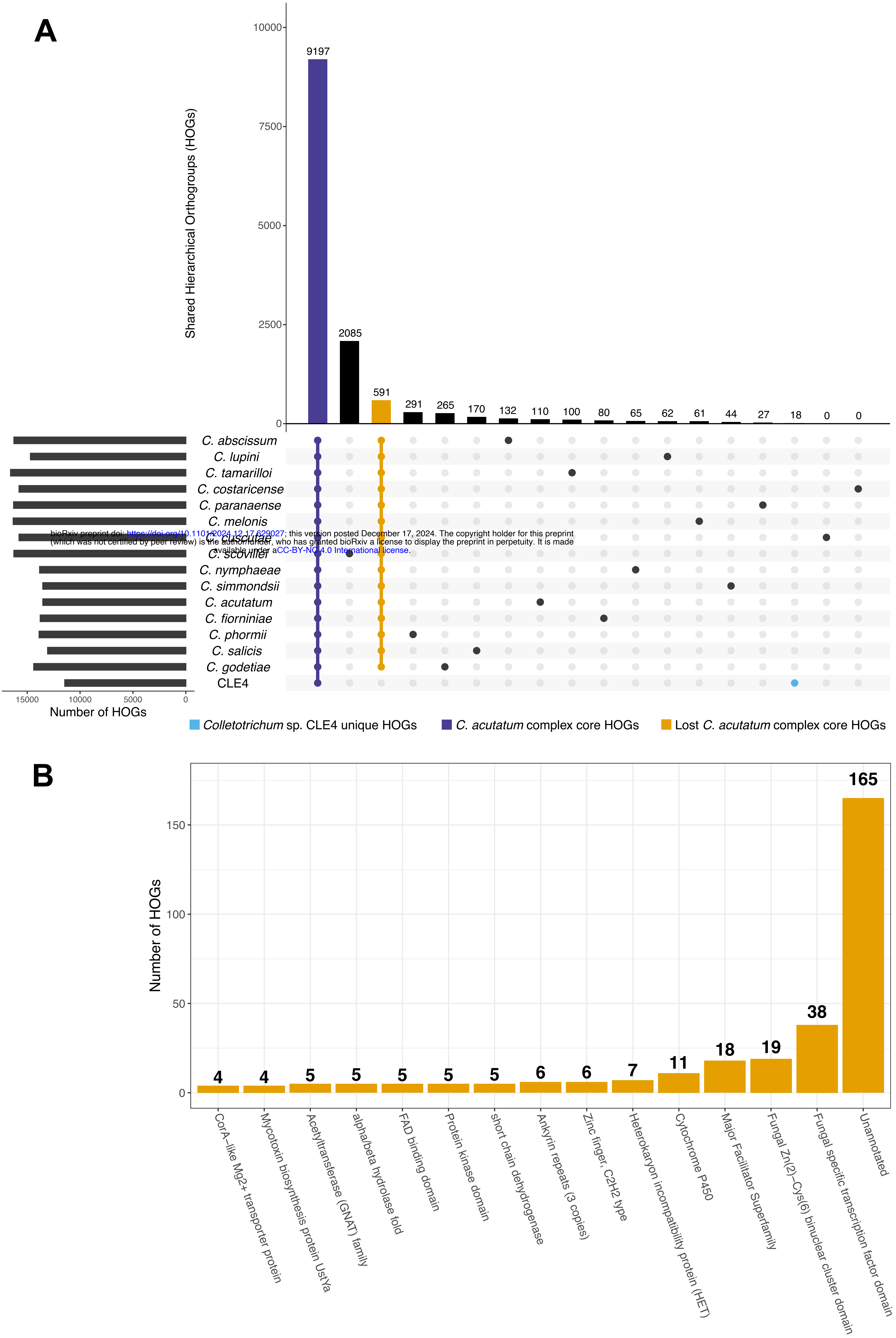
Software	**Statistic**	***Colletotrichum* sp. CLE4**
AAFTF	# contigs (>= 0 bp)	168
	# contigs (>= 50,000 bp)	127
	Total length (>= 0 bp)	48,026,924
	Total length (>= 50,000 bp)	47,411,477
	Largest contig (bp)	1,779,047
	Total length (bp)	48,026,924
	Repetitive regions (%)	2.87
	GC (%)	52.29
	N50	506,655
	L50	32
	# N's per 100 kbp	5.01
BUSCO (fungi_odb10)	Complete BUSCOs	98.8

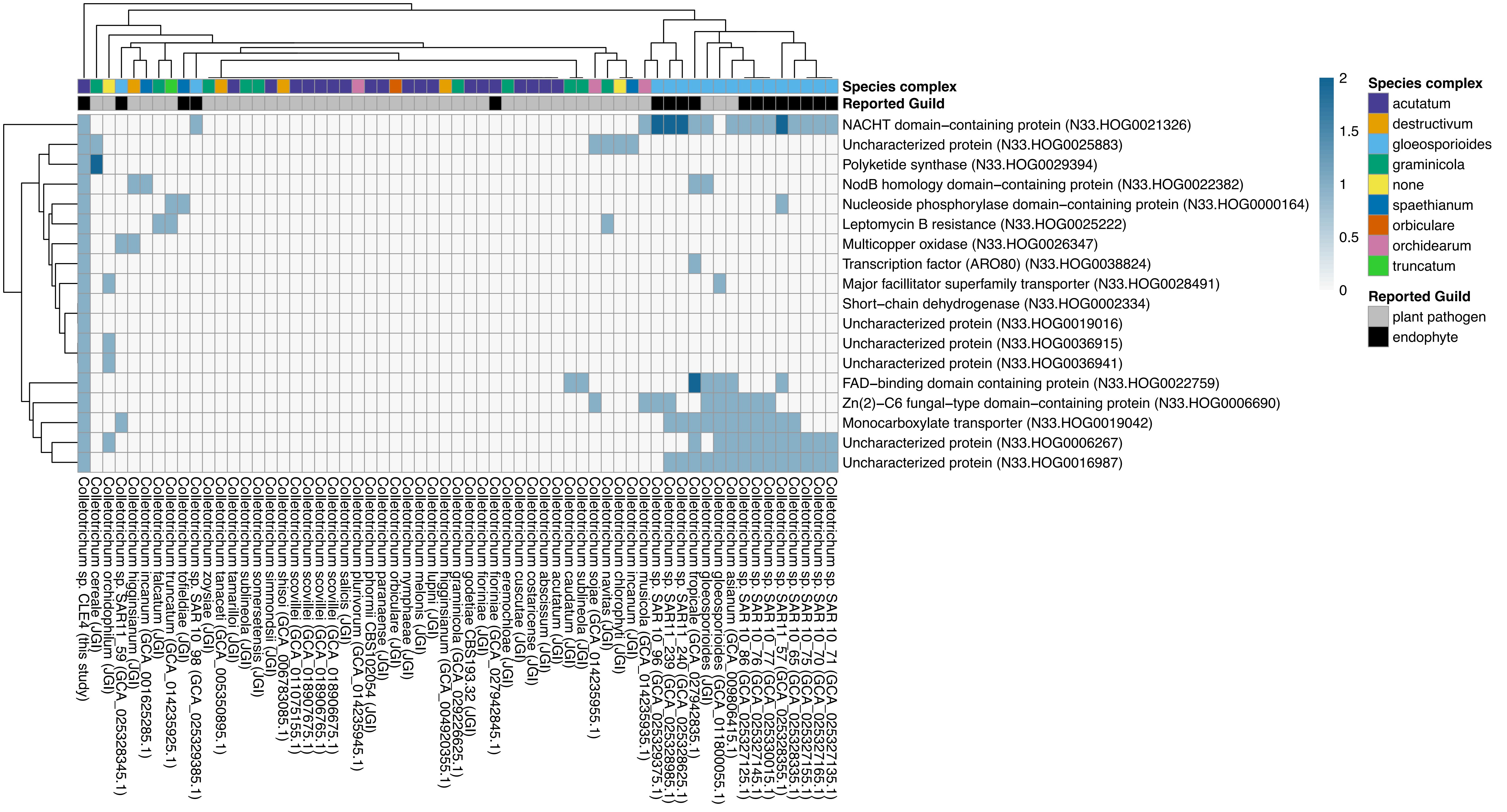
	Complete and single-copy BUSCOs	98.4
	Complete and duplicated BUSCOs	0.4
	Fragmented BUSCOs	0.3
	Missing BUSCOs	0.9
	Total BUSCO groups searched	758
Funannotate	Total number of gene models	12,015
	Number of mRNA genes	11,678
	Number of tRNA genes	337
	Genes with GO Term (%)	57.42
	Genes with InterProScan hit (%)	78
	Genes with EggNog hit (%)	95.53
	Genes with PFAM hit (%)	69.13
	Genes with CAZyme hits (%)	5.68
	Genes with MEROPS hits (%)	3.67
	Genes with secretion prediction (%)	11.97
EffectorP	Number of total predicted effectors	469
	Number of cytoplasmic effectors	141

	Number of apoplastic effectors	328


A
B
GenomeScope Profile
len:48,316,637bp uniq:87.3%
a:100%
kcov:315 err:0.613% dup:2.16 k:21 p:1
Frequency
0e+00
2e+05
4e+05
6e+05
0
200
400
600
800
1000
Coverage
observed
full model
unique sequence
errors
kmer-peaks
C
Percent of genome
0.00
0.25
0.50
0.75
LINE
SINE
LTR
DNA
Unknown
Repeat Class
D
Percent of genome
0.00
0.05
0.10
0.15
0
10
20
30
40
50
Sequence Divergence (CpG adjusted Kimura divergence)
Repeat Class
LINE
SINE
LTR
DNA
Unknown


Phylogeny
ascomycota_odb10
Gene Count
Assembly length (Mbp)
Colletotrichum abscissum (JGI)
Colletotrichum lupini (JGI)
Colletotrichum tamarilloi (JGI)
Colletotrichum costaricense (JGI)
Colletotrichum paranaense (JGI)
Colletotrichum melonis (JGI)
Colletotrichum cuscutae (JGI)
Colletotrichum scovillei (GCA_018907675.1)
Colletotrichum scovillei (GCA_018906765.1)
Colletotrichum scovillei (GCA_018906675.1)
Colletotrichum scovillei (GCA_011075155.1)
Colletotrichum nymphaeae (JGI)
Colletotrichum simmondsii (JGI)
Colletotrichum acutatum (JGI)
Colletotrichum fioriniae (JGI)
Colletotrichum fioriniae (GCA_027942845.1)
Colletotrichum phormii CBS102054 (JGI)
Colletotrichum salicis (JGI)
Colletotrichum godetiae CBS193.32 (JGI)
Colletotrichum sp. CLE4 (this study)
Colletotrichum orchidophilum (JGI)
Colletotrichum somersetensis (JGI)
Colletotrichum zoysiae (JGI)
Colletotrichum caudatum (JGI)
Colletotrichum navitas (JGI)
Colletotrichum graminicola (GCA_029226625.1)
Colletotrichum falcatum (JGI)
Colletotrichum sublineola (JGI)
Colletotrichum sublineola (JGI)
Colletotrichum eremochloae (JGI)
Colletotrichum cereale (JGI)
Colletotrichum incanum (JGI)
Colletotrichum incanum (GCA_001625285.1)
Colletotrichum tofieldiae (JGI)
Colletotrichum higginsianum (JGI)
Colletotrichum higginsianum (GCA_004920355.1)
Colletotrichum shisoi (GCA_006783085.1)
Colletotrichum tanaceti (GCA_005350895.1)
Colletotrichum chlorophyti (JGI)
Colletotrichum sp. SAR 10_76 (GCA_025327145.1)
Colletotrichum sp. SAR 10_77 (GCA_025330015.1)
Colletotrichum sp. SAR 10_86 (GCA_025327125.1)
Colletotrichum sp. SAR 10_71 (GCA_025327135.1)
Colletotrichum sp. SAR 10_70 (GCA_025327165.1)
Colletotrichum sp. SAR 10_75 (GCA_025327155.1)
Colletotrichum sp. SAR 10_65 (GCA_025328335.1)
Colletotrichum sp. SAR11_239 (GCA_025328985.1)
Colletotrichum sp. SAR11_59 (GCA_025328345.1)
Colletotrichum asianum (GCA_009806415.1)
Colletotrichum sp. SAR11_240 (GCA_025328625.1)
Colletotrichum sp. SAR11_57 (GCA_025328355.1)
Colletotrichum sp. SAR 10_96 (GCA_025329375.1)
Colletotrichum sp. SAR 10_98 (GCA_025329385.1)
Colletotrichum tropicale (GCA_027942835.1)
Colletotrichum gloeosporioides (JGI)
Colletotrichum gloeosporioides (GCA_011800055.1)
Colletotrichum truncatum (GCA_014235925.1)
Colletotrichum plurivorum (GCA_014235945.1)
Colletotrichum musicola (GCA_014235935.1)
Colletotrichum sojae (GCA_014235955.1)
Colletotrichum orbiculare (JGI)
Verticillium dahliae VdLs.17 (JGI)
Plectosphaerella cucumerina RP01 (JGI)
0
2
4
6
0
25
50
75
100
0
5000
10000
15000
0
25
50
75
BUSCO Status
Missing
Fragmented
Complete and duplicated
Complete and single−copy
Species complex
a acutatum
a destructivum
a gloeosporioides
a graminicola
a none
a orbiculare
a orchidearum
a spaethianum
a truncatum
a NA
Host Plant Type
eudicot
gymnosperm
magnoliid
monocot
Reported Guild
endophyte
plant pathogen


A
Shared Hierarchical Orthogroups (HOGs)
10000
7500
5000
2500
0
9197
2085
591
291
265
170
132
110
100
80
65
62
61
44
27
18
0
0
C. abscissum
C. lupini
C. tamarilloi
C. costaricense
C. paranaense
C. melonis
C. cuscutae
C. scovillei
C. nymphaeae
C. simmondsii
C. acutatum
C. fiorniniae
C. phormii
C. salicis
C. godetiae
CLE4
15000
10000
5000
0
Number of HOGs
Colletotrichum sp. CLE4 unique HOGs
C. acutatum complex core HOGs
Lost C. acutatum complex core HOGs
B
Number of HOGs
150
100
50
0
4
4
5
5
5
5
5
6
6
7
11
18
19
38
165
CorA–like Mg2+ transporter protein
Mycotoxin biosynthesis protein UstYa
Acetyltransferase (GNAT) family
alpha/beta hydrolase fold
FAD binding domain
Protein kinase domain
short chain dehydrogenase
Ankyrin repeats (3 copies)
Zinc finger, C2H2 type
Heterokaryon incompatibility protein (HET)
Cytochrome P450
Major Facilitator Superfamily
Fungal Zn(2)–Cys(6) binuclear cluster domain
Fungal specific transcription factor domain
Unannotated


Species complex
Reported Guild
NACHT domain–containing protein (N33.HOG0021326)
Uncharacterized protein (N33.HOG0025883)
Polyketide synthase (N33.HOG0029394)
NodB homology domain–containing protein (N33.HOG0022382)
Nucleoside phosphorylase domain–containing protein (N33.HOG0000164)
Leptomycin B resistance (N33.HOG0025222)
Multicopper oxidase (N33.HOG0026347)
Transcription factor (ARO80) (N33.HOG0038824)
Major facillitator superfamily transporter (N33.HOG0028491)
Short–chain dehydrogenase (N33.HOG0002334)
Uncharacterized protein (N33.HOG0019016)
Uncharacterized protein (N33.HOG0036915)
Uncharacterized protein (N33.HOG0036941)
FAD–binding domain containing protein (N33.HOG0022759)
Zn(2)–C6 fungal–type domain–containing protein (N33.HOG0006690)
Monocarboxylate transporter (N33.HOG0019042)
Uncharacterized protein (N33.HOG0006267)
Uncharacterized protein (N33.HOG0016987)
Colletotrichum sp. CLE4 (this study)
Colletotrichum cereale (JGI)
Colletotrichum orchidophilum (JGI)
Colletotrichum sp. SAR11_59 (GCA_025328345.1)
Colletotrichum higginsianum (JGI)
Colletotrichum incanum (GCA_001625285.1)
Colletotrichum falcatum (JGI)
Colletotrichum truncatum (GCA_014235925.1)
Colletotrichum tofieldiae (JGI)
Colletotrichum sp. SAR 10_98 (GCA_025329385.1)
Colletotrichum zoysiae (JGI)
Colletotrichum tanaceti (GCA_005350895.1)
Colletotrichum tamarilloi (JGI)
Colletotrichum sublineola (JGI)
Colletotrichum somersetensis (JGI)
Colletotrichum simmondsii (JGI)
Colletotrichum shisoi (GCA_006783085.1)
Colletotrichum scovillei (GCA_011075155.1)
Colletotrichum scovillei (GCA_018907675.1)
Colletotrichum scovillei (GCA_018906765.1)
Colletotrichum scovillei (GCA_018906675.1)
Colletotrichum salicis (JGI)
Colletotrichum plurivorum (GCA_014235945.1)
Colletotrichum phormii CBS102054 (JGI)
Colletotrichum paranaense (JGI)
Colletotrichum orbiculare (JGI)
Colletotrichum nymphaeae (JGI)
Colletotrichum melonis (JGI)
Colletotrichum lupini (JGI)
Colletotrichum higginsianum (GCA_004920355.1)
Colletotrichum graminicola (GCA_029226625.1)
Colletotrichum godetiae CBS193.32 (JGI)
Colletotrichum fioriniae (GCA_027942845.1)
Colletotrichum fioriniae (JGI)
Colletotrichum eremochloae (JGI)
Colletotrichum cuscutae (JGI)
Colletotrichum costaricense (JGI)
Colletotrichum abscissum (JGI)
Colletotrichum acutatum (JGI)
Colletotrichum caudatum (JGI)
Colletotrichum sublineola (JGI)
Colletotrichum sojae (GCA_014235955.1)
Colletotrichum navitas (JGI)
Colletotrichum chlorophyti (JGI)
Colletotrichum incanum (JGI)
Colletotrichum musicola (GCA_014235935.1)
Colletotrichum sp. SAR 10_96 (GCA_025329375.1)
Colletotrichum sp. SAR11_239 (GCA_025328985.1)
Colletotrichum sp. SAR11_240 (GCA_025328625.1)
Colletotrichum tropicale (GCA_027942835.1)
Colletotrichum gloeosporioides (JGI)
Colletotrichum gloeosporioides (GCA_011800055.1)
Colletotrichum asianum (GCA_009806415.1)
Colletotrichum sp. SAR 10_86 (GCA_025327125.1)
Colletotrichum sp. SAR 10_76 (GCA_025327145.1)
Colletotrichum sp. SAR 10_77 (GCA_025330015.1)
Colletotrichum sp. SAR11_57 (GCA_025328355.1)
Colletotrichum sp. SAR 10_65 (GCA_025328335.1)
Colletotrichum sp. SAR 10_75 (GCA_025327155.1)
Colletotrichum sp. SAR 10_70 (GCA_025327165.1)
Colletotrichum sp. SAR 10_71 (GCA_025327135.1)
2
1.5
1
0.5
0
Species complex
acutatum
destructivum
gloeosporioides
graminicola
none
spaethianum
orbiculare
orchidearum
truncatum
Reported Guild
plant pathogen
endophyte