



Comparative Genomics of Three *Colletotrichum scovillei* Strains and Genetic Analysis Revealed Genes Involved in Fungal Growth and Virulence on Chili Pepper

OPEN ACCESS

Edited by:

Martin Filion,

Saint-Jean-sur-Richelieu Research and Development Centre, Agriculture and Agri-Food Canada (AAFC), Canada

Reviewed by:

Riccardo Baroncelli, University of Salamanca, Spain
Rasappa Viswanathan, Indian Council of Agricultural Research (ICAR), India

*Correspondence:

Miin-Huey Lee
mhlee@nchu.edu.tw
Ming-Che Shih
mcsih@gate.sinica.edu.tw

† Present address:

Shu-Cheng Chuang, Department of Plant and Environmental Protection Sciences, University of Hawai'i at Mānoa, Honolulu, HI, United States

Specialty section:

This article was submitted to Microbe and Virus Interactions with Plants, a section of the journal Frontiers in Microbiology

Received: 19 November 2021

Accepted: 04 January 2022

Published: 27 January 2022

Citation:

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

Dai-Keng Hsieh^{1,2}, Shu-Cheng Chuang^{1†}, Chun-Yi Chen³, Ya-Ting Chao³, Mei-Yeh Jade Lu⁴, Miin-Huey Lee^{1,2,5*} and Ming-Che Shih^{1,3*}

¹ Ph.D. Program in Microbial Genomics, National Chung Hsing University and Academia Sinica, Taichung, Taiwan,

² Advanced Plant Biotechnology Center, National Chung Hsing University, Taichung, Taiwan, ³ Agricultural Biotechnology Research Center, Academia Sinica, Taipei, Taiwan, ⁴ Biodiversity Research Center, Academia Sinica, Taipei, Taiwan,

⁵ Department of Plant Pathology, National Chung Hsing University, Taichung, Taiwan

Colletotrichum scovillei causes anthracnose of chili pepper in many countries. Three strains of this pathogen, Coll-524, Coll-153, and Coll-365, show varied virulence on chili pepper. Among the three strains, Coll-365 showed significant defects in growth and virulence. To decipher the genetic variations among these strains and identify genes contributing to growth and virulence, comparative genomic analysis and gene transformation to show gene function were applied in this study. Compared to Coll-524, Coll-153, and Coll-365 had numerous gene losses including 32 candidate effector genes that are mainly exist in acutatum species complex. A cluster of 14 genes in a 34-kb genomic fragment was lost in Coll-365. Through gene transformation, three genes in the 34-kb fragment were identified to have functions in growth and/or virulence of *C. scovillei*. *CsPLAA* encoding a phospholipase A2-activating protein enhanced the growth of Coll-365. A combination of *CsPLAA* with one transcription factor *CsBZTF* and one C6 zinc finger domain-containing protein *CsCZCP* was found to enhance the pathogenicity of Coll-365. Introduction of *CsGIP*, which encodes a hypothetical protein, into Coll-365 caused a reduction in the germination rate of Coll-365. In conclusion, the highest virulent strain Coll-524 had more genes and encoded more pathogenicity related proteins and transposable elements than the other two strains, which may contribute to the high virulence of Coll-524. In addition, the absence of the 34-kb fragment plays a critical role in the defects of growth and virulence of strain Coll-365.

Keywords: *Colletotrichum scovillei*, chili pepper, gene transformation, virulence, comparative genomics

INTRODUCTION

Chili pepper (*Capsicum* spp.) is a globally significant spice crop. The cultivation of chili pepper is frequently threatened by the attacks of various pathogens, especially the anthracnose pathogens of the *Colletotrichum* species. *Colletotrichum* contains a large number of species members that have been classified into several species complexes (Cannon et al., 2012; Talhinas and Baroncelli, 2021). *Colletotrichum* pathogens associated with *Capsicum* plants come from diverse species

complexes including *C. acutatum*, *C. boninense*, *C. gloeosporioides*, *C. orchidearum*, *C. magnum*, *C. spaethianum*, and *C. truncatum* complexes (Damm et al., 2012a,b, 2019; Weir et al., 2012; De Silva et al., 2021; Talhinhos and Baroncelli, 2021). At least 28 species have been reported and most of them are from the acutatum, gloeosporioides, and truncatum complexes (Mongkolporn and Taylor, 2018; de Silva et al., 2019).

Colletotrichum scovillei is the most dominant species causing anthracnose of chili in Asia, and is found widely distributed in Indonesia, Malaysia, Thailand and Taiwan (De Silva et al., 2017, 2021; de Silva et al., 2019). It can infect many different species of *Capsicum*, especially species that are mainly grown for human consumption such as *C. annuum*, *C. frutescens*, *C. chinense*, and *C. baccatum* (Liao et al., 2012a; de Silva et al., 2019; De Silva et al., 2021). Two pathotypes of *C. scovillei*, CA1 and CA2, have been identified in Taiwan by the AVRDC–World Vegetable Center (Kumar et al., 2018; Sheu et al., 2019). CA2 is more virulent than CA1 in most tested cultivars of *Capsicum* species (Liao et al., 2012a; Kumar et al., 2018; Sheu et al., 2019). CA2 pathotype breaks down the resistance of *Capsicum chinense* PBC932-derived lines, which are resistant to CA1 pathotype (Liao et al., 2012a; Sheu et al., 2019). CA2 distribution was limited before the year 2000, but since then has become dominant and replaced CA1 in most of the chili pepper planting locations in Taiwan (Kumar et al., 2018; Sheu et al., 2019). According to amplified fragment length polymorphism (AFLP) analysis, CA2 members are homogenous and mostly clonal, whereas CA1 are genetically diverse (Sheu et al., 2019). Strains Coll-153 and Coll-365 are grouped in CA1 pathotype and strain Coll-524 is a member of the CA2 pathotype. Understanding the variation among the three strains may provide insight about the high virulence of the CA2 pathotype.

Whole genome sequencing has had a significant impact on understanding the biology, ecology, genetics, and evolution of various organisms. In fungi, comparative genomic study of genetic variations has revealed mobile pathogenicity chromosomes in *Fusarium* (Ma et al., 2010), evolutionary adaptations from a plant pathogenic to an animal pathogenic lifestyle in *Sporothrix* (Teixeira et al., 2014), an influx of transposable elements creating a genetically flexible landscape to respond to environmental changes in *Pyrenophora tritici-repentis* (Manning et al., 2013), potential orthologs in adaptation to specific hosts or ecological niches in *Botrytis*, *Colletotrichum*, *Fusarium*, *Parastagonospora*, and *Verticillium* (Klosterman et al., 2011; Gan et al., 2016; Buiate et al., 2017; Syme et al., 2018; van Dam et al., 2018; Valero-Jimenez et al., 2019), the evolution and diversity of putative pathogenicity genes in *Colletotrichum tanacetii* and other *Colletotrichum* species (Lelwala et al., 2019; Liu et al., 2020), core and specific genetic requirements for fungal endoparasitism of nematodes in *Drechmeria coniospora* (Lebrigand et al., 2016), host-specific and symptom-related virulence factors horizontally transferred from *Fusarium* to *Verticillium* (Chen et al., 2018; Zhang et al., 2019). Genomic comparisons with multiple races within a species have also been performed recently in *Fusarium* and *Verticillium* and revealed that secondary metabolites are crucial factors in *F. fujikuroi* for stunting symptom development and in *V. dahliae* for defoliation

symptom development (Niehaus et al., 2017; Zhang et al., 2019). In addition, a study on 18 isolates of *V. dahliae* identified highly variable regions with race specific signatures (Ingram et al., 2020). However, most of these studies were *in silico* analyses. Only few included genetic studies to demonstrate the functions of the potential factors selected by *in silico* analysis, and these studies were on *Fusarium* and *Verticillium* (Klosterman et al., 2011; Niehaus et al., 2017; Chen et al., 2018; Zhang et al., 2019).

In our previous study, we investigated the plant-pathogen interactions between the chili pepper fruit and the three *C. scovillei* strains Coll-153, Coll-365, and Coll-524 and revealed the infection process and several potential virulence factors determining the infection and colonization of host tissues (Liao et al., 2012a). On the original chili pepper host, Coll-524 shows highest virulence, while Coll-365 displays the lowest virulence. Coll-365 shows significantly slower growth on agar medium and has the weakest virulence on *Capsicum* spp. Coll-365 produces less spores in axenic culture and *in planta*. Coll-365 was found to accumulate less turgor pressure in appressorium but produces higher levels of cutinase activity than the other two strains. Examination of the infection process showed that the three strains can form primary hyphae in the epidermal cells at 72 h post-inoculation (hpi), indicating no delay in the penetration step for Coll-365 compared to Coll-524 and Coll-153. The three strains can grow into the cuticle layer of chili pepper fruit at 24 hpi and form a highly branched structure (HBPS) within the cuticle layer and penetrate epidermal cells at 72 hpi (Liao et al., 2012b). After infecting the epidermal cells of chili pepper fruits, Coll-365 appeared to have less ability to colonize host tissues, formed limited lesions on infected tissue and produced much fewer spores on the infected tissue. In addition, Coll-524 was more resistant to host defense compound capsaicin than Coll-153 and Coll-365. In this previous work, the genetic variations and genes contributing to the defects of growth and virulence remained undeciphered.

In this study, to decipher the genetic variations of the three *C. scovillei* strains, we sequenced the genomes of the three strains, investigated the differences in their genome compositions, identified potential genes involved in the phenotypes, and verified genes involved in the defects of growth and virulence of Coll-365. We setup a simple mathematical method to identify open reading frame (ORF) variations and successfully identified DNA fragments involved in the defects in growth and virulence of Coll-365. Moreover, by gene transformation we have demonstrated four genes that display function in germination, growth and/or virulence in the chili pepper pathogen *C. scovillei* strain Coll-365.

MATERIALS AND METHODS

Fungal Strains and Culture Conditions

Three *Colletotrichum* strains Coll-524, Coll-153, and Coll-365 were obtained from the AVRDC-The World Vegetable Center (Tainan, Taiwan) as mentioned in a previous study (Liao et al., 2012a). For sporulation, fungal strains were cultured on MS agar medium (0.1% yeast extract, 0.1% peptone, 1% sucrose, 0.25% MgSO₄•7H₂O, 0.27% KH₂PO₄, and 1.5% agar) for 6 days

at 25°C with a 12 h light/dark cycle. Spores were collected and used for general cultivation, assays of germination and appressorium formation, morphological examination, growth, and pathogenicity assay.

Genomic DNA Extraction and Library Preparation for Sequencing

Spores were inoculated into a 500 mL flask containing 100 mL MS liquid medium (1×10^5 spores/ml) at 25°C with shaking at 200 rpm for 1 day. The young hyphae were collected by filtering with a layer of miracloth. The DNA was extracted from the young hyphae using the CTAB extraction method (Chao et al., 2018). DNA purity and concentration were determined by Nanodrop 2000 (Thermo Fisher Scientific Inc., Waltham, MA, United States) and Qubit Fluorometric Quantification (Thermo Fisher Scientific Inc., Waltham, MA, United States) measurements. Libraries were generated from 2 to 5 µg of genomic DNA using the TruSeq 2 library preparation kit (Illumina, Inc., San Diego, CA, United States). Two types of libraries were prepared, short-fragment library (0.3 kb) for paired-end sequencing and long-fragment library (3 and 8 kb) for mate-pair sequencing. DNA fragmentation was made by shearing using an M220 Focused-Ultrasonicator (Covaris, Inc., Woburn, MA, United States) and the fragments with 296–300 bp, 3 kb or 8 kb were selected by gel-cutting. Without the 120 bp adaptor, the remaining 176–180 bp DNA fragments may have 20% overlap with 100 bp read length for paired-end sequencing. Long-fragment libraries, 3 and 8 kb, were prepared for Coll-524 only.

Genome Sequencing and Assembly

The Illumina HiSeq 2500 platform was used to sequence the genome of Coll-524 with 100 bp read length for paired-end and 3 kb mate pair libraries, and 125 bp read length for 8 kb mate pair library. The genomes of Coll-153 and Coll-365 were sequenced by Illumina Genome Analyzer with 150 bp read length. The raw reads were trimmed by Trimmomatic 0.39 to remove linkers and adaptors. Those refined reads of Coll-524 were further assembled by the ALLPATHS-LG *de novo* assembly program (Butler et al., 2008). The genome sequences of Coll-524 have been deposited in NCBI with accession numbers JAESDM010000001-JAESDM010000054.

Genome assembly of Coll-153 and Coll-365 was achieved by using the “Map Reads to contigs” function in CLC Genomic Workbench v.9.5.1. All trimmed reads were used to map to the Coll-524 genome under default settings. The consensus mapped sequences were extracted and joined to form a scaffold by the “Extract Consensus Sequence” function. Unmapped reads were collected and assembled by using the “*De Novo* Assembly” function in CLC software. The assembled contigs with sizes larger than 5 kb were collected and counted as the genome sequences of Coll-153 or Coll-365. The genome sequences of Coll-153 and Coll-365 have been deposited in NCBI with accession numbers JAESDN010000001- JAESDN010000059 for Coll-153 and JAESDO010000001- JAESDO010000059 for Coll-365. The genome completeness of the three strains were assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO

v.3.0) and the sordariomyceta_odb10 dataset (Simao et al., 2015; Waterhouse et al., 2018).

Multi-Locus Phylogenetic Analysis

The DNA sequences of five marker genes (ACT1, CHS1, GAPDH, ITS, and TUB2) were used for phylogenetic analyses (Cannon et al., 2012). The sequences of the five genes from the three strains, 170 other *Colletotrichum* strains and an outgroup species, *Monilochaetes infuscans*, were obtained and used in the assay. Among them, the five marker genes of 108 *Colletotrichum* species and *M. infuscans* were obtained from Q-bank database¹ according to the GenBank accession numbers listed in a publication of Cannon et al. (2012), while sequences from 62 *Colletotrichum* strains were obtained by downloading whole genome sequences from NCBI FTP and then identifying the five gene sequences by using CLC Genomic Workbench v.9.5.1. The information for all the genes sequences is provided in **Supplementary Table 1**. All collected sequences were aligned by MAFFT v.7 online version (Katoh et al., 2019) with default settings and then trimmed by trimAl 1.4.1 (Capella-Gutierrez et al., 2009) with automated1 setting. Five marker genes were concatenated in each strain. Bayesian inference of all the concatenated sequences was analyzed by MrBayes 3.2.7 (Ronquist et al., 2012). The Markov Chain Monte Carlo (MCMC) chains were set for 5,000,000 generations. Sample frequency was set for every 100 generations. Phylogenetic tree of the 174 strains were generated by FigTree v.1.4.3².

Genome Annotation

Gene annotation was conducted using MAKER (v.2.31.10) (Holt and Yandell, 2011) pipeline with default settings (Augustus, PJ7, RNAseq reads). The genome of Coll-524 was used for gene annotation. The assembled contigs from Coll-524 RNAseq data was provided for EST evidence. The RNAseq data of Coll-524 included 10 sets of RNAseq data, which were infected purified cuticle layer, 18 h hyphal of axenic culture, and 8 sets of infected chili pepper fruits at different infection time points (unpublished). The CDS and protein sequence of *Colletotrichum fioriniae* PJ7 (PRJNA244481) was provided to MAKER for the “EST evidence” and “Protein Homology Evidence” functions respectively. The gene annotation of Coll-153 and Coll-365 was conducted using the same method as that for Coll-524. The CDS region of each gene was extracted from the genome sequence according to the GFF file which was originated from MAKER by GffRead (Pertea and Pertea, 2020). Gene annotation results of Coll-524, Coll-153, and Coll-365 have been deposited in NCBI.

Gene Functional Analysis, Clustering and Orthology Analysis

The genes of Coll-524, Coll-153 and Coll-365 were searched with nr database by using BLASTx (BLAST 2.2.29+) to obtain the information about gene function. To find effector candidates, protein sequences of the three strains were analyzed by EffectorP 2.0 (Sperschneider et al., 2018). Carbohydrate active enzymes

¹<https://qbank.eppo.int/fungi/>

²<http://tree.bio.ed.ac.uk/software/figtree/>

(CAZymes) candidates were searched by dbCAN2 meta server (Zhang et al., 2018). Candidate gene selection followed the recommendation setting provided by dbCAN2 meta server. Genes related to secondary metabolism were predicted using the Secondary Metabolite Unique Regions Finder (SMURF) database (Khaldi et al., 2010). The Pathogen-Host Interactions database (PHI-base) v.4.10 was downloaded from PHI-base website. All genes of the three strains were used to blast against PHI-base using CLC Genomic Workbench (Urban et al., 2020). Genes with *E*-values lower than 10 were selected according to the default setting of the PHIB-BLAST web server. The protein sequences of all annotated genes of Coll-153, Coll-365, and Coll-524 were used for ortholog analysis with the default settings of OrthoFinder (Emms and Kelly, 2019). After the Orthofinder analysis, single-copy orthogroups were generated directly. Other analyses of orthologs, including multi-copy orthogroups with the same gene numbers, multi-copy orthogroups with different gene numbers, orthogroups in every two strains, strain-specific orthologs, and unassigned genes, were further classified by EXCEL. These unassigned genes were considered as “strain specific genes” together with genes in strain specific orthogroup. Genes of Coll-524 with extra copy number in comparison with the other two strains in the orthogroup, and genes analyzed to be Coll-524 strain specific genes were further considered as “Coll-524 extra genes.”

Identification of Microsatellites, Repeats and Transposon Elements

MISA 2.0 was used for microsatellite analysis (Beier et al., 2017) and transposon elements were analyzed by Transposon PSI 2.2.26³. The transposon element sequences were acquired according to the GFF files provided by Transposon PSI. These sequences were further extracted to create a Transposon element database (TED). RepeatModeler was used for *de novo* genome-wide repeat family analysis (Flynn et al., 2020). The sequences of TcMar-Fot1 were further blasted with BLASTn to the NCBI database and the functional domains were analyzed using InterPro (Blum et al., 2021) to determine if DDE was present (Braga et al., 2014).

Identification of Genome Sequence Variations and Open Reading Frame Variations

During the mapping of the Coll-153 or Coll-365 genome to the Coll-524 genome, data containing the variations between Coll-524 and the other two strains were generated with the “Extract Consensus Sequence” function. The data included four datasets, Misc. difference (miscellaneous differences), Insertion, Deletion and Removed, which are specified by CLC with default setting. Briefly, “Deletion” is for positions where a gap is called while the reference has a non-gap; “Insertion” is for positions where a non-gap is called while the reference has a gap; “Removal” is for positions where no reads are mapped to the reference; “Misc. difference” is for every position where the consensus is different

from the reference. The Misc. difference dataset consisted of single nucleotide polymorphism (SNP) and polymorphism with more than one nucleotide, which was named as multiple nucleotide polymorphism (MNP) in this study. The four datasets were combined to generate the genome sequence variations and used for genome comparisons to identify ORF variations between Coll-524 and the other two strains as described in **Figure 1**. To simplify the analysis of ORF variations, the Removed datasets of Coll-153 and Coll-365 were combined. Briefly, all the sequences which existed in Coll-524 but removed from Coll-153 or Coll-365 were extracted from Coll-524 genome to create a “removed sequence database.”

Open reading frame variation analysis was based on the calculations of physical position overlapping between each ORF and the four datasets of variations (misc. difference, Insertion, Deletion, and Removed). The GFF file generated by MAKER could provide the physical positions of mRNA, exon, 5'UTR, CDS, 3'UTR for each gene. A gene usually contains multiple CDSs in eukaryotes and all CDSs together within the gene is the complete coding sequence to translate a protein encoded by the gene. To analyze ORF variation of Coll-153 and Coll-365, the ORF position of each Coll-524 gene was generated by setting the first nucleotide of the first CDS as the beginning position and the last nucleotide of the last CDS as the end position.

To analyze the variation of each ORF between Coll-524 and the other two strains, firstly, all scaffolds of Coll-524 were added together one by one to form a single sequence. The physical positions of each ORF were modified to fit into the single sequence. The four datasets of variations (misc. difference, Insertion, Deletion, and Removed) generated by CLC were tagged with physical positions in each scaffold of Coll-524. Therefore, the second step was to edit the physical positions of each variation according to the single genome sequence of Coll-524. The position differences of each variation to the ORF were compared and calculated as indicated in **Figures 1C–G**. The physical position overlapping situation of each ORF and each variation was identified by the calculation formula listed in **Figure 1G**. An ORF overlapped with a variation indicated that this ORF varied between Coll-524 and the other strain carrying this variation. ORFs overlapped with variation were further called “ORF-V.”

Screening of Genes Potentially Involved in the Variations of the Three Strains for Functional Verification in Coll-365

To select genes for genetic analysis to verify their functions on the defects of growth and virulence of Coll-365, four selection criteria were set for gene selection. First were ORFs predicted to be ORF-V. Second were “Coll-524 extra genes” according to the result of OrthoFinder analysis. Third were genes analyzed to belonging to candidate effector, CAZyme, TF, SMURF, enzyme or PHI (*E*-value < 10e-50). Finally, were ORFs with <50% coverage to Coll-153 and/or 365 according to the blastn results. Genes with the four criteria were analyzed by Venn Diagram⁴.

³<http://transposonpsi.sourceforge.net/>

⁴<http://bioinformatics.psb.ugent.be/webtools/Venn/>

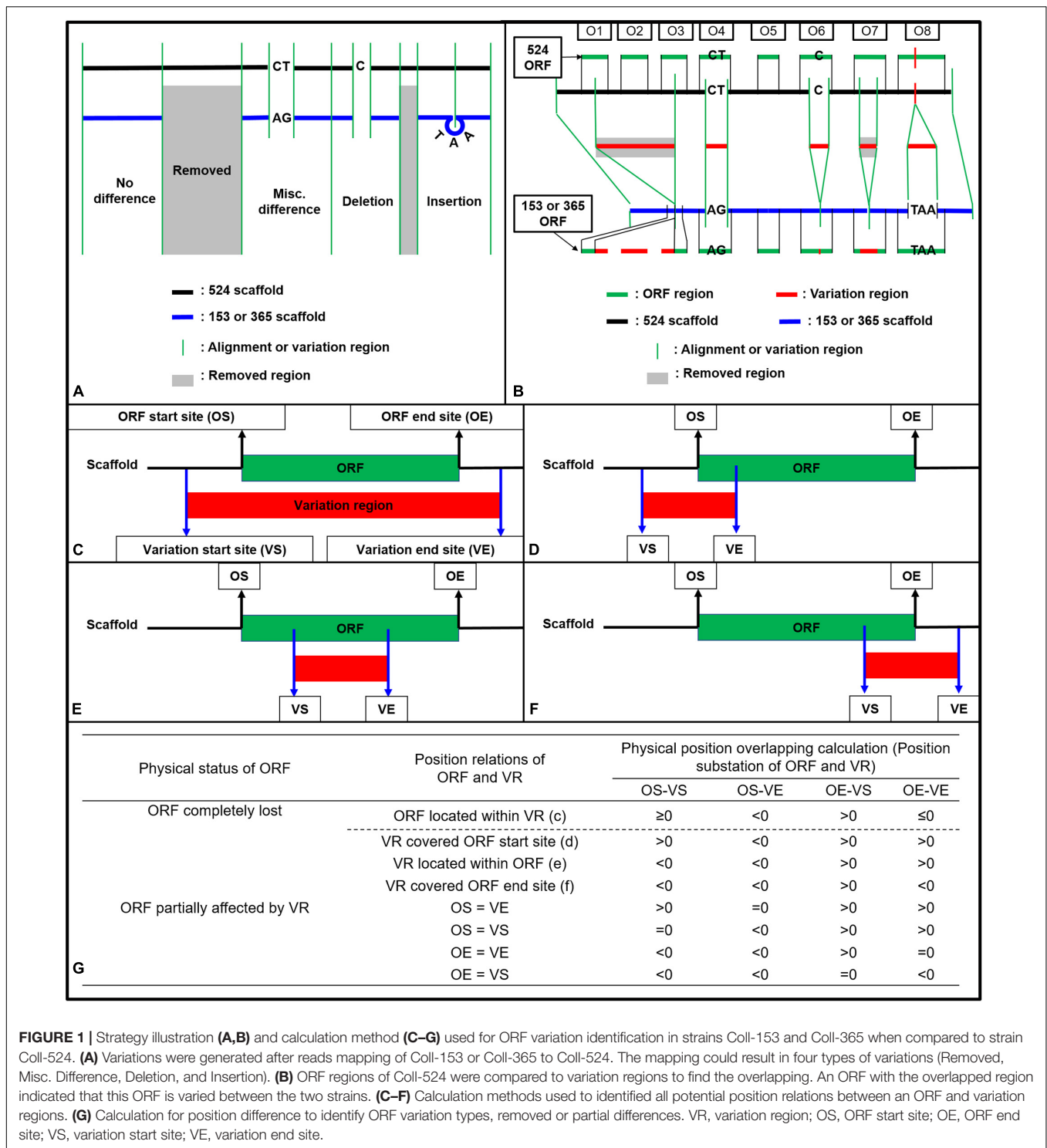


FIGURE 1 | Strategy illustration (A,B) and calculation method (C–G) used for ORF variation identification in strains Coll-153 and Coll-365 when compared to strain Coll-524. (A) Variations were generated after reads mapping of Coll-153 or Coll-365 to Coll-524. The mapping could result in four types of variations (Removed, Misc. Difference, Deletion, and Insertion). (B) ORF regions of Coll-524 were compared to variation regions to find the overlapping. An ORF with the overlapped region indicated that this ORF is varied between the two strains. (C–F) Calculation methods used to identified all potential position relations between an ORF and variation regions. (G) Calculation for position difference to identify ORF variation types, removed or partial differences. VR, variation region; OS, ORF start site; OE, ORF end site; VS, variation start site; VE, variation end site.

Pathogen–host interaction related genes were selected with an *E*-value less than 10e-50. Genes with an *E*-value less than 10e-50 but greater than 10e-100 were considered to have almost identical sequences to the database according to the manual for CLC Genomics Workbench 9.5.

Plasmid Construction

Plasmids BsHR and BsBR were constructed and used in this research to clone the selected target gene(s) for transformation into Coll-365. pBsHR was constructed by cloning hygromycin resistance cassette (HygR) of pBHt2 (Mullins et al., 2001) into

pBluescript SK(+) (Short et al., 1988) with PCR amplification, restriction enzyme digestion (*EcoRV* and *SwaI*, designed in the primers), and DNA ligation. Plasmid BsBR was modified from pBsHR by replacing (HygR) with bleomycin resistance gene (BleoR) of pAN8-1 (Mullaney et al., 1985). The BleoR cassette was PCR amplified with specific primers containing *EcoRV* and *SwaI* restriction sites and then cloned into *EcoRV*- and *SwaI*-digested pBsHR. To clone the target genes from Coll-524, a target gene containing an approximate 1-kb promoter and 0.5-kb terminator was amplified with a high-fidelity DNA polymerase KOD-plus-Neo (TOYOBO, Osaka, Japan) and ligated to *EcoRV* or *SwaI* digested and Shrimp Alkaline Phosphatase (rSAP) treated pBsHR. If double gene-transformation was needed, a second target gene was cloned in to pBsBR for protoplast transformation and phleomycin was used as the selection antibiotic (Pilgeram and Henson, 1990). For gene-transformation VII, the two genes were closely linked and were amplified together with one primer set by PCR and transferred to Coll-365 by single transformation. Transformations IX and X were completed with two steps of transformations by transferring additional gene(s) into the transgenic strain VII-1. Total 10 transformations (I-X) were conducted as listed in **Table 1**. The accession numbers of genes used for gene transformation were summarized in **Supplementary Table 2**.

Protoplast Transformation

Polyethylene glycol (PEG)/Ca²⁺-mediated protoplast transformation was used to transfer the target gene to Coll-365. Spores collected from a 6-day MS agar medium were inoculated into a 500-ml flask containing 400 ml MS liquid medium and a stirrer bar, and cultured for 16 h under 25°C at low speed to prevent the attachment of spores to the flask surface. The young hyphae were collected by filtering through a layer of miracloth and washed with 100 ml sterilized water and 100 ml wash buffer (1 M NaCl and 10 mM CaCl₂) under suction. The washed hyphae (250 mg) were then resuspended in a 50-ml flask containing 10 ml osmotic buffer (10 mM Na₂HPO₄, pH 5.8, 20 mM CaCl₂, and 1.2 M NaCl) with 90 mg lysing enzymes (Sigma, L1412), and incubated in an orbital shaker with 150 rpm at 30°C for 6 h. The undigested hyphae were removed by filtering through miracloth and the protoplasts were collected by centrifugation at 1500 g, 4°C for 10 min. The protoplasts were resuspended by adding 100 µl mixture of four parts of STC buffer (1.2 M sorbitol, 10 mM Tris-HCl, pH 7.5, and 10 mM CaCl₂) and one part of PEG [50% (w/v) polyethylene glycol 3350, 10 mM Tris-HCl, pH 7.5, and 10 mM CaCl₂]. For transformation, 20 µl of 20 µg plasmid DNA was added to a tube containing 100 µl protoplast suspension with gentle mixing and the mixture was placed in ice for 20 min. PEG was added into the DNA-protoplast solution four times with different volumes. The adding steps were performed by following a previous description (Liu and Friesen, 2012) with slight modifications. Briefly, 20, 80, 300, and 600 µl PEG were added into protoplast suspension step by step. After each addition, the mixture was gently mixed and left to stand for 3 min, and then left to stand a further 20 min after the last addition at room temperature. After the 20 min incubation, 3 ml regenerate buffer [4 mM Ca(NO₃)₂·4H₂O, 1.5 mM KH₂PO₄,

1 mM MgSO₄·7H₂O, 2.5 mM NaCl, 60 mM glucose, and 1 M sucrose] was added into the PEG-DNA-protoplast mixture and cultured at 25°C for 16 h with 100 rpm shaking. The protoplasts were collected with centrifugation at 25°C, 1800 g for 10 min. A total of 300 µl regenerated protoplasts were obtained and then evenly spread on three regeneration agar medium plates. After incubation for 6–8 days at 25°C, the transformants were isolated, single spore purification and PCR verification for the target gene insertion.

PCR Analysis for Gene Loss in Coll-153 and Coll-365 and Verification of Gene Transformation in Coll-365 Transformants

Eight genes were selected for functional assay in Coll-365 as listed in **Table 1**. PCR assays were used to confirm the loss of the eight genes in Coll-365 using specific primer sets as listed in **Supplementary Table 3**. Genomic DNA was extracted as described above and tubulin gene (Accession number: MW073123) was used as the control for PCR assay. To verify gene transformation of transgenic Coll-365 strains, regular PCR and RT-PCR assays were performed. Regular PCR using genomic DNA as template to amplify the transgenic gene was used for all transgenic strains. For transgenic strains carrying more than one transgene, RT-PCR assays were conducted in addition to the regular PCR. Primers used in these assays are listed in **Supplementary Table 3**. For RNA extraction, spores were inoculated into a 50 mL flask containing 20 mL MS liquid medium (1 × 10⁵ spores/ml) at 25°C with shaking at 150 rpm for 2 days. The hyphae were collected by filtering with a layer of miracloth. The RNA was extracted with Trizol reagent (Invitrogen). RNA purity and concentration were determined and cDNA was synthesized using 5 µg RNA and the MMLV reverse transcription kit (Invitrogen). PCR was performed using 20 µl reaction volume contained 1 × PCR Buffer, 0.2 mM dNTP, 0.4 µM primers, 1 U Blend Taq plus (TOYOBO), and 1 µl of template DNA. PCR reaction was performed as follows: 94°C for 2 min and 25 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 3 min. The PCR products were analyzed by agarose gel electrophoresis.

Fungal Growth, Spore Germination, and Appressorium Formation Assays

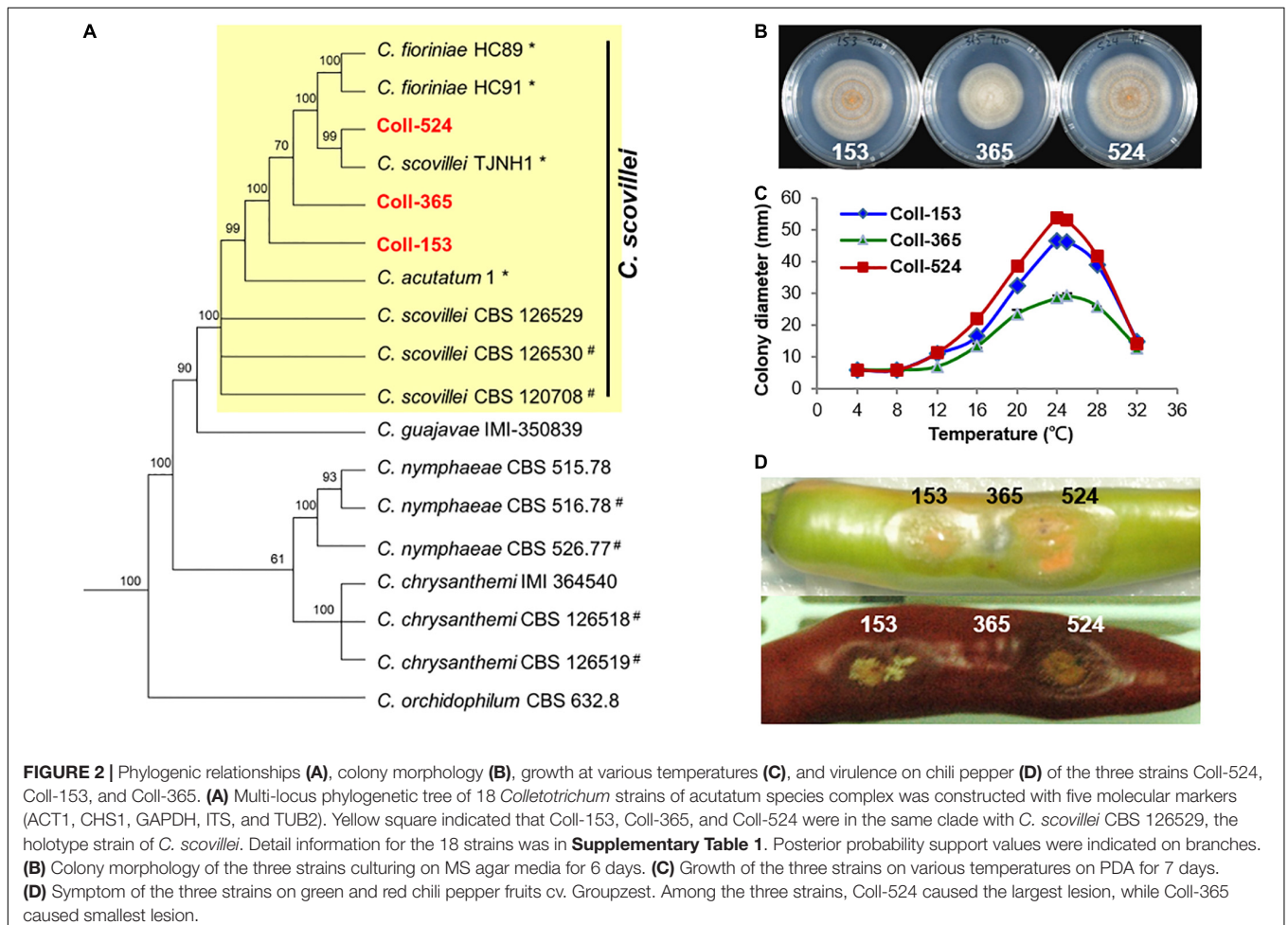
Fungal growth was assayed by inoculating spore suspension on the center of three different agar media. Spore germination and appressorium formation were detected in 96-well plates. Briefly, spores were collected from MS agar medium and the concentration was adjusted to 2 × 10⁵ spores/ml by sterilized water. For growth assay, a 5-µl, spore suspension containing 500 spores was dropped on the center of MS agar medium plate and incubated for 6 days at 25°C with a 12 h light/dark light cycle. Colony sizes were measured by using ImageJ 1.53a (Schneider et al., 2012). To detect germination and appressorium formation, 80 µl of spore suspension was added into each well of a 96-well plate and incubated at 25°C with 12 h light/dark light cycle for 8 h. To count the numbers of germinated spores

TABLE 1 | Genetic transformation and phenotyping in strain Coll-365.

Transformation	Gene(s) ^a	Predicted function	Transgenic strains used for phenotyping	Function identified in this study
I	CsEF1	Candidate effector	I-1, I-2	ND ^b
II	CsLAP	Laccase precursor	II-1, II-2	ND
III	CsBGN	Endo- β -1,3-glucanase	III-1, III-2	ND
IV	CsGIP	Non-specific hit to COG1196 containing hypothetical protein	IV-1, IV-2	Inhibited germination
V	CsBZTF	bZIP transcription factor	V-1, V-2	ND
VI	CsCZCP	C6 zinc finger domain-containing protein	VI-1, VI-2	ND
VII	CsBZTF CsCZCP		VI-1, VI-2	Enhanced virulence
VIII	CsPLAA	Phospholipase a-2-activating protein	VIII-1, VIII-2	Enhanced growth and virulence
IX	CsBZTF CsCZCP CsPLAA		IX-1, IX-2	Enhanced growth and virulence
X	CsBZTF CsCZCP CsPLAA CsWDCP	WD domain containing protein	X-1, X-2	Enhanced growth and virulence

^aGene(s) of Coll-524 used to be transferred into Coll-365 in each transformation experiment.

^bND, non-detected; UD, undetermined.



and appressorium, the 96-well plate was placed upside down and examined using a light microscope. All the experiments mentioned above were conducted at least two times with three replicates for each strain at each time. The statistical significance of the data was determined by One-way ANOVA at $P < 0.05$ using Statistical Package for the Social Sciences software, version 20 (IBM SPSS software, IBM Corp., Armonk, NY, United States).

Pathogenicity Assay

Pepper fruits were used for pathogenicity assay, including *Capsicum annuum* cv. Hero, Fushimi amanaga and Groupzest. The fresh harvested fruits were washed with water and surface sterilized with 0.5% bleach and left to dry overnight. Fungal spores were collected from MS agar medium and the concentration was adjusted to 2×10^5 spores/ml. Spore suspension was dropped (5 μ l/drop) on the fruit surface with dual inoculation, in which the wild type strain was inoculated on one side of the fruit and a transgenic strain was inoculated on the other site of the same fruit. The fruits were incubated in a growth chamber at 25°C with 12 h light/dark light cycle. At least three fruits having 3–5 inoculation sites were used for each strain in each inoculation. Lesion sizes were measured with Spot image analysis software (Spot RT Software version 3.5.1; Diagnostic Instruments, Inc., Sterling Heights, MI, United States) and the statistical analysis was performed by paired t -test at $P < 0.05$ using Statistical Package for the Social Sciences software, version 20 (IBM SPSS software, IBM Corp., Armonk, NY, United States).

RESULTS

Phylogenetic Analysis Revealed That Coll-524, Coll-153, and Coll-365 Are Closely Related to *Colletotrichum scovillei*

To understand the phylogenetic relationships among the three *Colletotrichum* strains, multi-locus phylogenetic analysis with 174 fungal strains was performed. The results indicated that Coll-153, Coll-365, and Coll-524 all belong to the acutatum species complex (Supplementary Figure 1). Further examination of the original hosts of members in acutatum species complex revealed that the three strains formed a small clade with other strains and most of the members were isolated from *Capsicum* plants (Supplementary Figure 2). However, the phylogenetic tree generated from the 173 *Colletotrichum* strains could not distinguish the phylogenetic relationship of the three strains with *C. scovillei* (CBS 126529, the holotype strain of *C. scovillei*; Damm et al., 2012a) and *C. guajavae* (IMI 350839, the holotype strain of *C. guajavae*; Damm et al., 2012a; Supplementary Figure 2). Another multi-locus phylogenetic analysis was conducted with 18 strains (Supplementary Table 1), including 12 strains from Supplementary Figure 2 and 6 strains (marked with #) that were closely related to *C. scovillei* CBS 126529 and *C. guajavae* IMI 350839 in the clade 2 of the *Colletotrichum acutatum* species complex (Damm et al., 2012a). The results indicated that the three strains were closer to *C. scovillei* CBS 126529

TABLE 2 | Genome features of *Colletotrichum scovillei* strains Coll-524, Coll-153, and Coll-365.

Genome features	Coll-524	Coll-153	Coll-365
Sequencing	Paired-end/mate pair	Paired-end	Paired-end
Assembly	<i>De novo</i>	Mapped to Coll-524	Mapped to Coll-524
Coverage	71.9	105	104
Fragment coverage	37.2	–	–
Jump coverage			
3K	24.4	–	–
8K	10.3	–	–
Number of scaffolds	54	59	59
Total scaffold length (Mb)	51.491	50.114	49.922
N50 scaffold size (Mb)	3.597	–	–
Max scaffold length (Mb)	8.040	7.995	7.989
Number of scaffolds (> 50 KB)	26	25	24
GC content (%)	51.98	52.03	52.05
BUSCO completeness (%)	99.80	99.80	99.80
Total genes	15,626	15,432	15,387

than *C. guajavae* IMI 350839 (Figure 2A). In addition, Coll-524 was closer to *C. scovillei* TJNH1 isolated from pepper in China and two *C. fioriniae* strains (HC89 and HC91) isolated from apple in the United States than Coll-365 and Coll-153 (Figure 2A). These results indicated that strains Coll-524, 153 and 365 all belong to the *C. scovillei* species. In addition, the three strains showed significant differences in growth and virulence but had a similar preference for temperature range (Figures 2B–D; Liao et al., 2012a).

Genomic Comparison Revealed DNA Deletion Patterns Among Different Scaffolds in the Three Strains

The genomic sequences of Coll-524 were generated from paired-end and mate-pair libraries and assembled and annotated using the ALLPATHS-LG (Butler et al., 2008) and MAKER (Holt and Yandell, 2011) pipelines. Genomic sequences of Coll-153 and Coll-365 were generated from paired-end libraries and mapped to the assembled Coll-524 genome (for details, see section “Materials and Methods”). The results of assembly and annotation are summarized in Table 2. The genome size of Coll-524 was 51. Four hundred and ninety one MB and consisted of 54 scaffolds with an N50 of ~ 3.6 Mb. The genome size of Coll-153 was 50.114 MB and consisted of 59 scaffolds. The genome size of Coll-365 was 49.922 MB and consisted of 59 scaffolds. A total of 15,626, 15,432, and 15,387 genes were annotated in the Coll-524, Coll-153, and Coll-365 genomes, respectively (Table 2). BUSCO (Waterhouse et al., 2018) analysis for genome completeness showed that 99.8% of the conserved proteins in sordariomycetes could be identified in all three strains.

Coll-153 and Coll-365 had their own special genome fragments compared to Coll-524, and there were six scaffolds

for Coll-153 and seven scaffolds for Coll-365. One (0.9-kb) and two (0.9-kb and 5.1-kb) of the 54 scaffolds in Coll-524 did not exist in Coll-153 and Coll-365, respectively, and no genes were encoded in the two scaffolds. A total of 29 and 30 genes were annotated from Coll-153 and Coll-365 specific genome sequences, respectively, and 23 of them are homologs in both Coll-153 and Coll-365. None of these genes had traits related to pathogenicity when blasting to the PHI database.

Genome mapping of Coll-153 and Coll-365 to Coll-524 was analyzed with CLC Genomic Workbench v.9.5.1 and provided the specified information of DNA insertion, deletion, removal, and Misc. difference in Coll-153 and Coll-365 (**Supplementary Table 4**). The results showed that Coll-153 and Coll-365 had similar patterns within the four types of variations, but Coll-365 had a slightly higher number of total variations than Coll-153. The removed sequence database showed that 991,585 and 1,179,624 bp were absent in Coll-153 and Coll-365, respectively, compared to Coll-524. The largest removed fragments were 28,776 bp in Coll-365 and 10,125 bp in Coll-153. Scaffolds 19, 20 and 22 in Coll-153 and Coll-365 showed high numbers of DNA sequence removal compared to the three scaffolds in Coll-524 (**Figure 3**). The total number of sites of DNA sequence removal in scaffolds 19, 20, and 22 were 377, 386, and 256 in Coll-153, and 167, 114, and 76 in Coll-365, respectively. A notable difference in DNA sequence removal patterns between Coll-153 and Coll-365 was that DNA sequence removal occurred significantly in scaffolds 17 in Coll-365 but not in Coll-153 (**Figure 3**). Large DNA fragment removals occurred in Coll-365 more frequently than Coll-153 (**Supplementary Table 5**). However, most of them were located at the short scaffolds 19, 20, and 22. Coll-365 had four removals with DNA fragments larger than 20 kb, but none of the large fragments was found in Coll-153. Regarding the removal of DNA fragments larger than 5 kb, a total of 66 removals appeared in Coll-365 but only 8 in Coll-153. Small DNA fragment removals (<1 kb) were found most often in Coll-153 (**Supplementary Table 5**).

Open Reading Frame Variation Analysis Identified Open Reading Frame Losses in Strains Coll-153 and Coll-365

DNA insertion, deletion, removal, and misc. differences (SNP and MNP) occurring in Coll-153 or Coll-365 may affect the open reading frames (ORFs) of their genes. To understand the influence on ORFs, we used a simple mathematical method to detect ORF variation (ORF-V) for all ORFs in Coll-153 and Coll-365 as illustrated in **Figure 1** and described in section “Materials and Methods.” The results showed that most of the ORF-Vs occurred with SNP in Coll-153 and 365, and Coll-365 appeared to have significantly greater numbers of full ORF losses than Coll-153 (**Table 3**). The total numbers of ORF-V of Coll-153 and Coll-365 compared to Coll-524 were 708 and 725, respectively, and 799 of them were singular. A total of 91 and 74 ORF-Vs were solely found in Coll-365 and Coll-153, respectively, but most of these ORF-Vs were SNP in Coll-153 or Coll-365, indicating that the two strains had similar genes with ORF-V to Coll-524. The major difference in the two strains with regard to ORF-V was

found in scaffold 17 and was caused by ORF deletion as described below. The 799 singular ORFs were used for further analysis.

Open reading frames affected by DNA removal (ORF-Vrem) in each scaffold were further analyzed and the results were displayed in **Figure 4**. ORF-Vrem mostly occurred in scaffolds 19, 20 and 22 in Coll-153 and 365. There were a total of 260 and 274 ORF-Vrem events in Coll-153 and Coll-365, respectively. All ORFs were ORF-Vrem in scaffolds 19, 20, 22, and 44; however, scaffold 44 was a short scaffold with only one ORF (**Figure 4**). The distribution of ORF-Vrem in Coll-153 and Coll-365 was very similar except that 14 ORFs clustered in a 34-kb genomic fragment were totally or partially removed in Coll-365 compared to 1 ORF with SNP only in Coll-153 in scaffold 17 (**Supplementary Figure 3**). The 14 ORFs encode 4 transcription regulation-related proteins, 1 GPI-anchored protein, 4 enzymes and enzyme-related proteins, and 5 hypothetical proteins. In addition, when compared with 62 genomes of other *Colletotrichum* strains that are available in the NCBI database, the 14 genes appeared in almost all members of the acutatum species complex, except Coll-365 (**Figure 5**). Moreover, *CsWDCP* and *CsPLAA* existed in nearly all assayed strains, indicating that the two genes are core genes of the *Colletotrichum* species. Five of these genes existed in almost all the strains in the acutatum species complex, suggesting the possibility of them being acutatum species complex-specific genes (**Figure 5**).

Gene Ortholog Analysis Showed That the Three Strains Carry Slightly Different Ortholog Compositions

To understand the difference in orthology among the three strains, proteins were analyzed using OrthoFinder (Emms and Kelly, 2019). Ortholog analysis showed a high similarity in gene compositions among the three strains as shown in **Supplementary Table 6**. More than 88% genes in the three strains belonged to single-copy orthogroups, indicating that the three strains all carried the 13,779 single-copy genes (**Supplementary Table 6**). A total of 502 multiple-copy orthogroups with the same gene numbers containing 1,167 genes were all identified in the three strains. A combination of single-copy orthogroups and multi-copy orthogroups with the same gene copy number showed that there were a total of 14,946 genes in the three strains and they were 95.7, 96.9, and 97.1% of the total genes in Coll-524, Coll-153, and Coll-365, respectively. This suggests that the three strains had high similarity in gene compositions (**Supplementary Table 6**). The ortholog differences within the three strains analyzed using the dataset of multi-copy orthogroups with different gene number showed Coll-524 carried considerably greater numbers of genes than the other two strains. The three strains all carried the 128 orthogroups but they had different gene numbers in each of the orthogroups. Among the 128 orthogroups, Coll-524 had 97 and 101 more genes than Coll-153 and Coll-365, respectively, and there were a total of 103 singular genes found from the 97 and 101 genes. Regarding the differences in orthogroups that only existed in two strains, 92, 33, and 94 orthogroups were identified in Coll-524 and Coll-153,

TABLE 3 | Open reading frame variation statistics of strains Coll-153 and Coll-365 detected and calculated based on the method shown in **Figure 1**.

ORF variations	Variation types	No. of ORF with variations		ORF-V or ORF-Vrem
		Coll-153	Coll-365	
ORF loss	Completely Removed	92	166	ORF-V/ORF-Vrem
ORF partially affected	Partially Removed	126	84	ORF-V/ORF-Vrem
	SNP	404	400	ORF-V
	MNP	4	4	ORF-V
	SNP, MNP, Deletion, Insertion and Removed	82	71	ORF-V
Sum		708	725	
Union of varying ORFs within Coll-153 and 365		799		

SNP, single nucleotide polymorphism; MNP, polymorphism with more than one nucleotide; ORF-Vrem, ORF affected by DNA removal.

Coll-524, and Coll-365, and Coll-153 and Coll-365, respectively (**Supplementary Table 6**). Further analysis of the differences between pairs of two strains showed that some orthogroups were only found in two strains, and Coll-524 and Coll-153 shared more orthogroups (92 orthogroups, 94 genes) than Coll-524 and Coll-365 (33 orthogroups, 33 genes). Based on the data mentioned above, the combinations of the 103 genes, the 94 and 33 genes, the 85 of strain-specific genes and the 196 unassigned genes of Coll-524 strain, 511 genes in total, were used together with other criteria for further gene selections used in gene functional analysis.

Comparison Analysis Revealed Variations in Pathogenicity-Related Categories in the Three Strains

To understand the variations in pathogenicity-related genes among the three strains, analyses of pathogenicity-related functional categories were performed. Six categories were used, candidate effectors, carbohydrate active enzymes (CAZymes),

secondary metabolism clusters, transcription factor (TF), and enzymes in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, and pathogen–host interaction (PHI) related genes. The results showed the three strains had very similar numbers of genes in various categories (**Supplementary Table 7**). When combining these data and the ortholog data in a further analysis, notable differences were observed (**Table 4** and **Supplementary Table 8**). With regard to strain-specific orthogroups and unassigned genes in ortholog analysis, Coll-524 had significantly greater gene numbers of candidate effectors and PHI than Coll-153 and Coll-365 (**Table 4**). With regard to multi-copy orthogroups with different gene numbers, Coll-524 had more genes belonging to the six functional categories than Coll-153 and Coll-365 (**Supplementary Table 8**). For orthologs only found in two strains, there were no notable differences between any set of two-strains with regard to gene numbers among the six functional categories.

The gene variations of the six functional categories in scaffolds 17, 19, 20, and 22 in the three strains are summarized

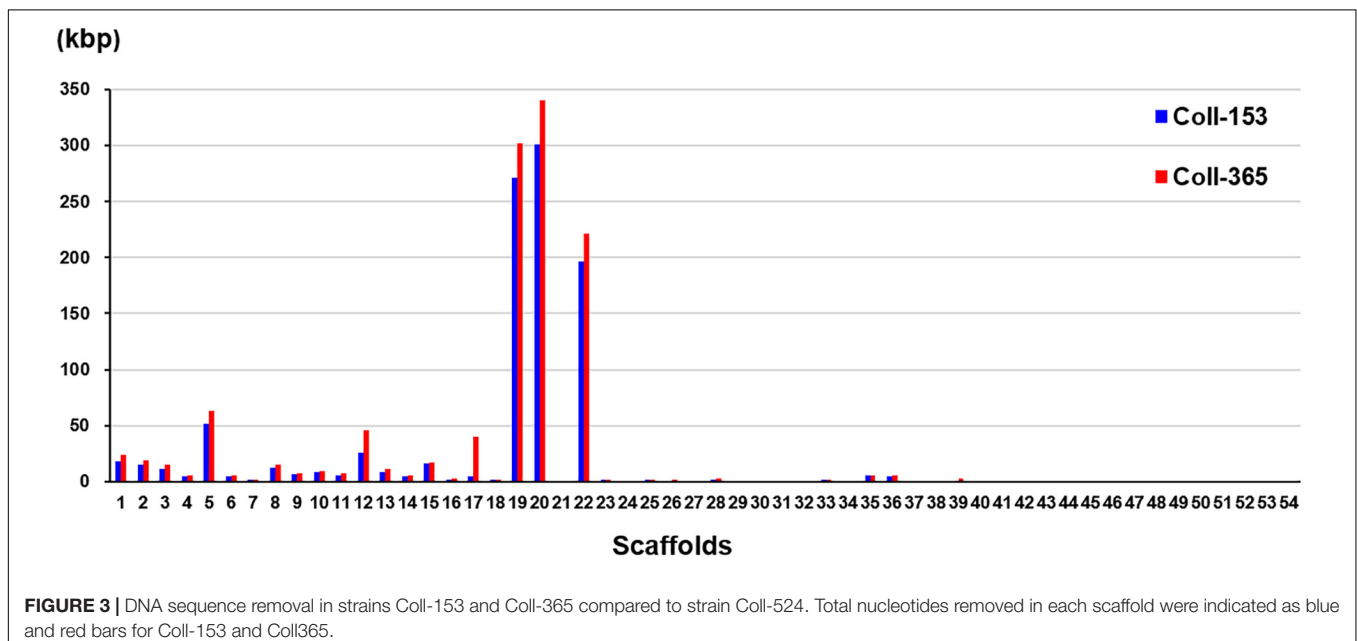
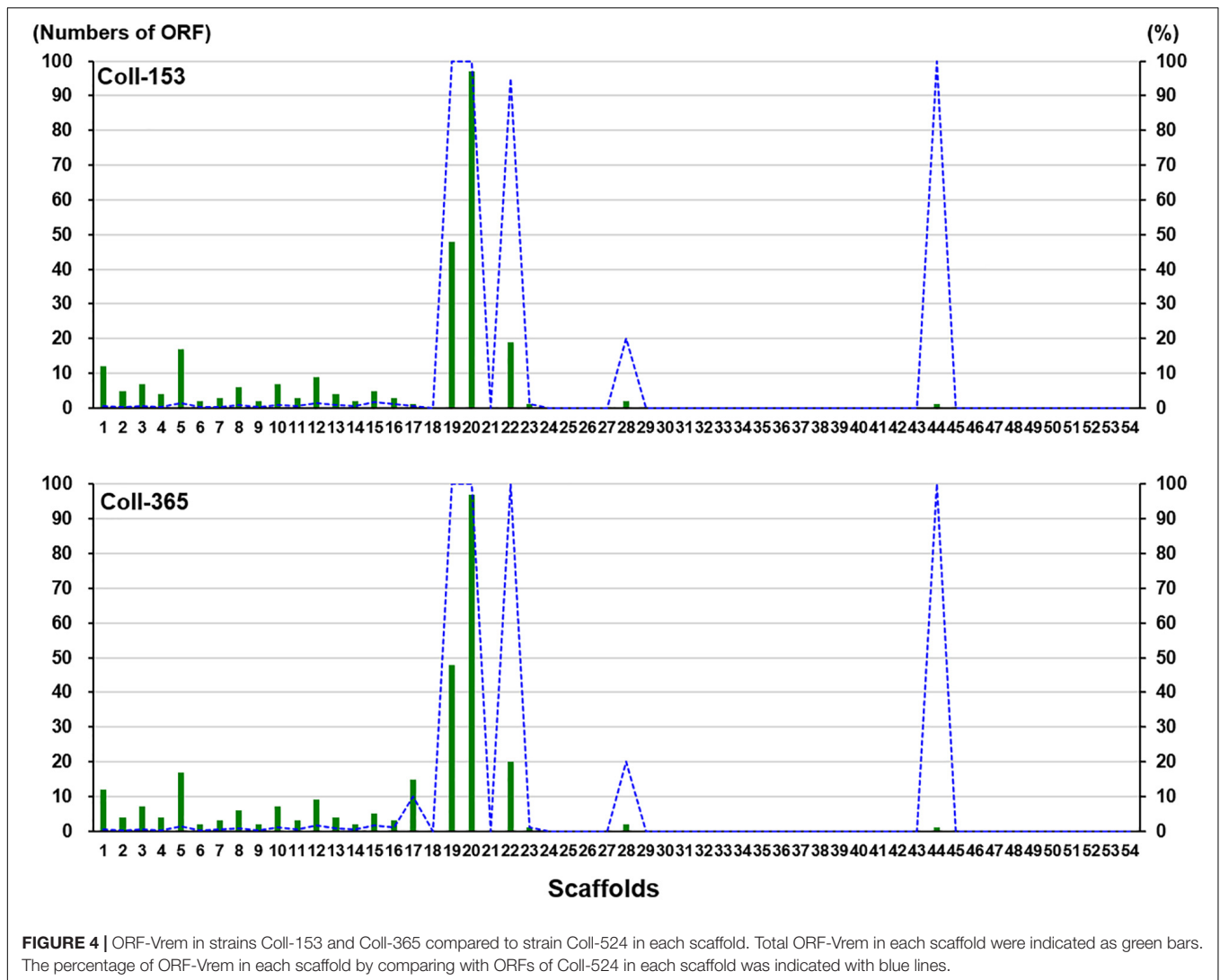


FIGURE 3 | DNA sequence removal in strains Coll-153 and Coll-365 compared to strain Coll-524. Total nucleotides removed in each scaffold were indicated as blue and red bars for Coll-153 and Coll365.



in **Supplementary Table 9**. Coll-153 and Coll-365 had remarkably fewer genes than Coll-524 in PHI (114 genes) and candidate effector (33 genes) categories (**Supplementary Table 9**). The ORFs of the 33 candidate effector genes were further analyzed using nucleotide BLAST (blastn) against 62 genomes of *Colletotrichum* strains with genomes available in the NCBI database. As shown in **Figure 6**, two strains of *C. acutatum* species complex (Coll-524 and *C. scovillei* TJNH1) carried all the 33 candidate effector genes, while one strain of this complex, *C. acutatum* 1, had 32 of the candidate effectors. However, two strains with phylogenetics closely related to Coll-524, *C. fiorinia* HC89 and HC91, only carried two of the 33 candidate effector genes. Among the 62 *Colletotrichum* strains, 27 did not carry any of the 33 candidate effectors and they were mainly in the species complexes *graminicola*, *spaethianum*, and *gloeosporioides*. Interestingly, none of the members of the *graminicola* species complex carried any of the 33 candidate effectors. Moreover, a total of 24 of the 33 candidate effectors

were only found in the *acutatum* complex, and not in the other complexes.

The variations of PHI genes among the three strains in scaffolds 17, 19, 20, and 22 were found to mainly occur in scaffolds 19, 20, and 22, especially genes required for fungal full virulence in the PHI database. A total of 79 genes with functions related to reducing virulence were found to be partially or fully removed in Coll-365 and 71 of them were not in Coll-153 (**Supplementary Table 9**).

The variations of the three strains in KEGG pathways on scaffolds 17, 19, 20, and 22 were found to mainly occur on scaffold 20, in which 100% of the KEGG genes were lost in Coll-153 and Coll-365. The investigation of KEGG combined with ortholog analysis showed that gene number variations within the three strains in scaffolds 17, 19, 20, and 22 were found in five metabolic pathways. These pathways were metabolisms of drugs, purine, thiamine, phenylpropanoid, and *N*-glycan biosynthesis. Coll-524 carried more genes than Coll-153 and/or Coll-365 in the five pathways (**Supplementary Figures 4–8**).

Species complex	Strain name	Host	CsWDCP	CsPLAA	KAG7038733.1	CsCZCP	KAG7038785.1	CsBZTF	CsBGN	KAG7038726.1	KAG7038750.1	KAG7038773.1	KAG7038789.1	KAG7038728.1	KAG7038757.1	KAG7038758.1
Orbiculare	<i>C. orbiculare</i> MAFF 240422	<i>Cucumis sativus</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. trifolii</i> 543-2	<i>Medicago sativa</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. sidae</i> CBS 518-97	<i>Sida spinosa</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. spinosum</i> CBS-515-97	<i>Xanthium spinosum</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. lindemuthianum</i> 83.501	<i>Phaseolus vulgaris</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. lindemuthianum</i> 89 A2	<i>Phaseolus vulgaris</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Acutatum	<i>C. godetiae</i> C184	<i>Olea europaea</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. acutatum</i> C71	<i>Olea europaea</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. fioriniae</i> PJ7	<i>Fragaria x ananassa</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. fioriniae</i> HC89	<i>Malus</i> sp.	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. fioriniae</i> HC91	<i>Malus</i> sp.	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. scovillei</i> TJNH1	<i>Capsicum annuum</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	Coll-524	<i>Capsicum annuum</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	Coll-153	<i>Capsicum annuum</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	Coll-365	<i>Capsicum annuum</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. acutatum</i> 1	Pepper	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. nymphææ</i> SA-01	<i>Fragaria x ananassa</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. orchidophilum</i> IMI 309357	<i>Phalaenopsis</i> sp.	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	Graminicola	<i>C. graminicola</i> M1.001	<i>Zea mays</i>	●	●	●	●	●	●	●	●	●	●	●	●	●
<i>C. graminicola</i> M5.001		<i>Zea mays</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
<i>C. sublineola</i> CgS11		<i>Sorghum bicolor</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
<i>C. sublineola</i> TX430BB		<i>Sorghum bicolor</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
<i>C. falcatum</i> Cf671		<i>Saccharum officinarum</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Spaethianum	<i>C. tofieldiae</i> 0861	<i>Arabidopsis thaliana</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. tofieldiae</i> CBS127615	<i>Agapanthus</i> sp.	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. tofieldiae</i> CBS130851	<i>Semele androgyna</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. tofieldiae</i> CBS168.49	<i>Lupinus polyphyllus</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. incanum</i> MAFF 238704	<i>Raphanus sativus</i> L. Daikon Group	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. incanum</i> MAFF 238712	<i>Raphanus sativus</i> var. <i>longipinnatus</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Destructivum	<i>C. higginsianum</i> MAFF 305635-RFP	<i>Brassica rapa</i> var. <i>perviridis</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. destructivum</i> YC1	Tobacco	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. shiso</i> PG-2018a	<i>Perilla frutescens</i> var. <i>crispa</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. lentis</i> CT-30	Lentil	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. tanacet</i> BRIP57314	<i>Tanacetum cinerariifolium</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Dematium	<i>C. coccodes</i> NJ-RT1	Pepper	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. coccodes</i> RP180a	Pepper	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Gloeosporioides	<i>C. chlorophyti</i> NTL11	<i>Solanum lycopersicum</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. sp.</i> JS-367	<i>Sansevieria trifasciata</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. gloeosporioides</i> Cg01	<i>Huperzia serrata</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. gloeosporioides</i> ES026	<i>Huperzia serrata</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. camelliae</i> CcLH18	<i>Camellia oleifera</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. gloeosporioides</i> CgLH19	<i>Camellia oleifera</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. sp.</i> COLG25	Citrus	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. gloeosporioides</i> SMCG1C	<i>Cunninghamia lanceolata</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. gloeosporioides</i> Lc1	<i>Liriodendron chinense</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. siamense</i> CsLH17	<i>Camellia oleifera</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. siamense</i> COLG-38	Citrus	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. siamense</i> COLG-44	Citrus	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. siamense</i> COLG-50	Citrus	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. siamense</i> COLG-90	Citrus	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. fruticola</i> CGMCC3-17371	<i>Fragaria x ananassa</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. fruticola</i> Nara gc5	Strawberry	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. fruticola</i> 1104-7	Apple	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. fruticola</i> 15060	Mango cv. Tainong	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. gloeosporioides</i> 30206	Apple	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. gloeosporioides</i> COLG-95	Citrus	●	●	●	●	●	●	●	●	●	●	●	●	●	●
<i>C. sp.</i> COLG31	Citrus	●	●	●	●	●	●	●	●	●	●	●	●	●	●	
<i>C. gloeosporioides</i> Cg 14	Avocado cv. Fuente	●	●	●	●	●	●	●	●	●	●	●	●	●	●	
<i>C. musae</i> GM20	<i>Musa</i> sp.	●	●	●	●	●	●	●	●	●	●	●	●	●	●	
Truncatum	<i>C. capsici</i> KLC-C	<i>Capsicum annuum</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. truncatum</i> MTCC-3414	<i>Capsicum annuum</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. gloeosporioides</i> TYU	<i>Taxus cuspidata</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. truncatum</i> KLC-C5	Pepper	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Boninense	<i>C. karsti</i> CKLH20	<i>Camellia oleifera</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<i>C. sansevieriae</i> Sa-1-2	<i>Sansevieria trifasciata</i>	●	●	●	●	●	●	●	●	●	●	●	●	●	●

FIGURE 5 | The distribution of the 14 genes of scaffold 17 in strains Coll-524, 153, and 365, and the 62 *Colletotrichum* strains with whole genome sequences deposited in NCBI. The 14 genes were analyzed using nucleotide BLAST against the 62 genomes of *Colletotrichum* strains. Blue squares indicate the presence of genes by the blasting with coverage > 90% and identity > 60%. The gene was ordered from left to right by distribution frequency in the 62 genomes.

Coll-153 and Coll-524 (Figure 7). Repeat sequence analysis with RepeatModeler showed that Coll-524, Coll-153, and Coll-365 had 53, 67, and 44 repeat families, respectively. The DDE_1 domain identified by TransposonPSI is from the TcMar-Fot1 family in the RepeatModeler database. The complete Coll-Fot1 sequence was identified by combining sequences provided by the TransposonPSI and RepeatModeler (Supplementary Figure 9).

Eight Genes Were Selected for Functional Analysis

Orthogroup and ORF-V analysis revealed two sets of variations in gene composition of the three strains. To obtain comprehensive analysis of variation, additional analysis was conducted by blastn using Coll-524 genes against the genome sequences of Coll-153 and 365. A total of 219 genes from Coll-153 and/or 365 with less than 50% coverage to Coll-524 were identified and used with genes selected from the other three groups, orthogroup analysis (511 genes), ORF-V analysis (799 genes) and pathogenicity-related categories analysis, for further analysis (Supplementary Figure 10).

The results are shown in Supplementary Figure 10A. A total of 59 genes were found from the four groups. Among the 59 genes, 47 genes were lost in both Coll-153 and Coll-365, and 12 genes were only absent in Coll-365, including 8 candidate effectors, one CAZyme, one TF, one KEGG pathway, and one PHI (Supplementary Figure 10B). Most of the 47 genes were located at scaffolds 19 and 20, while most of the 12 genes were located at the scaffolds 17 and 20 (Supplementary

Table 11). A total of 137 genes were clustered from three gene groups excluding the pathogenicity-related function categories (Supplementary Figure 10A). These 137 genes contained five cytochrome P450 genes, five FAD binding domain-containing protein genes and large numbers of unknown and hypothetical protein genes. Eight genes, five from the 59 gene group, and three from the 137 gene group, which were highly expressed during infection, based on RNAseq data, were then selected for functional analyses. Five genes from the 59 gene group are *CsEF1*, *CsBGN*, *CsLAP*, *CsBZTF*, and *CsPLAA* (Supplementary Figure 10B and Table 1), and the three genes from the 137 gene group are *CsWDCP*, *CsCZCP* and a hypothetical protein gene (*CsGIP*). *CsGIP* was shown to inhibit the germination of Coll-365 in this study (see below). Thus, it is named as germination inhibited protein (*GsGIP*). Genes encoding cytochrome P450 or FAD binding domain-containing protein were not selected because they belong to large gene families in Coll-524. The eight genes were analyzed by PCR to confirm their absence in the genome of Coll-153 and/or Coll-365 (Supplementary Figure 11A).

The Genes Lost in a 34-kb Fragment Have a Large Effect on Strain Coll-365 Morphology and Pathogenicity

The eight selected genes were driven by their native promoters and transferred into Coll-365. Five genes were located within a 34-kb fragment at scaffold 17 and two sets of two genes were closely linked (*CsCZCP* and *CsBZTF*; *CsWDCP* and *CsPLAA*). Therefore, for Coll-365 transgenic strains carrying two closely linked genes, three or four of the four genes were generated (Table 1). All transgenic strains were selected with PCR assays for the insertion fragment and/or further confirmed with RT-PCR for their expression in Coll-365 (Supplementary Figures 11B,C). The single spore-purified transgenic strains were used for assays on spore germination, appressorium formation, growth, and pathogenicity assay on chili pepper fruits. Two independent transgenic strains of every gene transformation were used in all assays. Coll-524, Coll-153, and Coll-365 showed similar ability with regard to spore germination and appressorium formation. Two independent transgenic strains of transformation IV (*CsGIP*; Table 1) showed a significantly lower germination rate than the wild-type Coll-365 in two independent experiments (Figure 8). There were no significant differences in appressorium formation between the gene transformation strains and wild-type strain Coll-365, indicating that the penetration ability of the transformants might not be affected. In the growth assay, Coll-365 had significantly slower growth than Coll-153 and Coll-524. Among the transgenic strains, two independent transgenic strains of transformation VIII (*CsPLAA*), IX (*CsBZTF*, *CsCZCP*, and *CsPLAA*), and X (*CsBZTF*, *CsCZCP*, *CsPLAA*, and *CsWDCP*) showed significantly enhanced growth on MS agar medium in comparison with their wild-type strain Coll-365, but still showed slightly slower growth than Coll-524 (Figure 9 and Supplementary Figure 12). In addition, transgenic strains VIII, IX, and X displayed puffy colony with abundant aerial hyphae and unclear orange rings that usually indicate sporulation

TABLE 4 | Functional category statistics of strain-specific orthogroups and unassigned genes of the three *Colletotrichum scovillei* genomes.

Categories	Strain-specific orthogroup			Unassigned genes		
	Coll-524	Coll-153	Coll-365	Coll-524	Coll-153	Coll-365
Effector	7	0	0	38	6	12
CAZyme	0	0	0	0	0	0
SMURF	3	3	2	0	0	0
TF	0	0	0	1	2	0
KEGG	6	6	6	3	2	2
PHI						
Increased virulence	0	0	0	8	1	1
Reduced virulence	7	0	0	51	6	15
Loss of pathogenicity	0	0	0	8	3	1
Lethal	2	0	0	3	0	1
Mixed results	39	36	39	8	3	1
Mixed functions with reduced virulence	2	2	2	1	0	0
Mixed functions without function in PHI	0	0	0	1	1	2
Unmatched	19	6	7	74	31	29
Sum	85	53	56	196	55	64

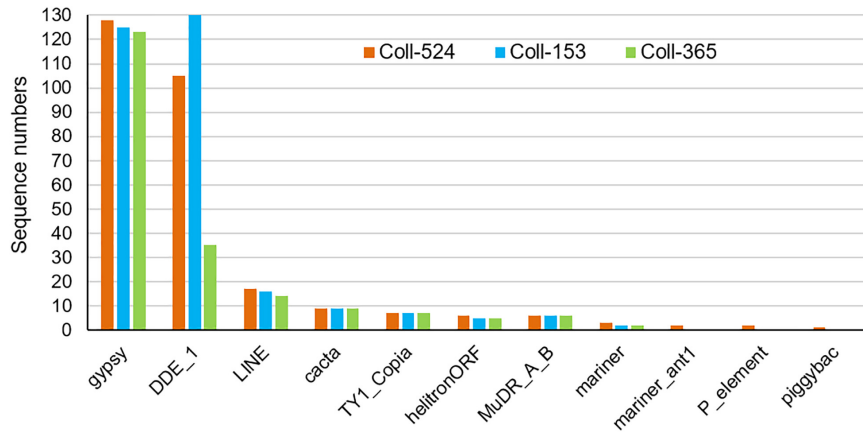


FIGURE 7 | Distribution of repeat sequences among *Colletotrichum scovillei* strains Coll-524, Coll-153, and Coll-365 when analyzed with TransposonPSI.

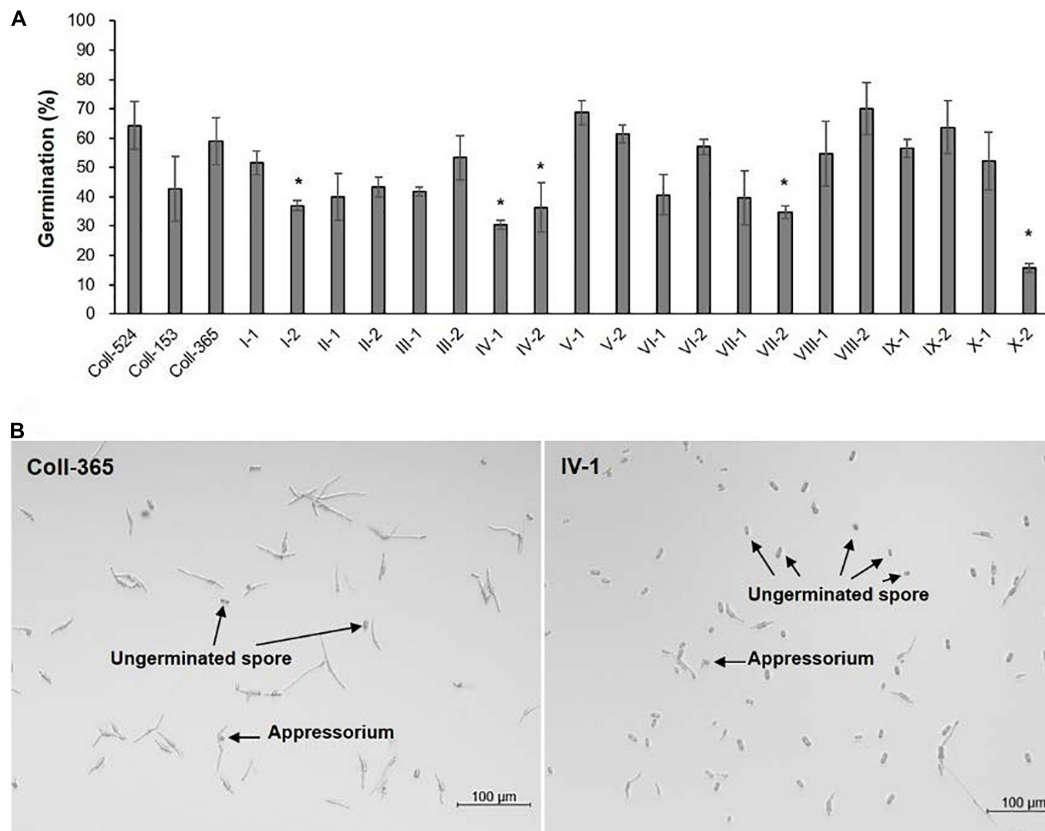


FIGURE 8 | Spore germination and appressorial formation of strains Coll-524, Coll-153, and Coll-365 and the transgenic strains of Coll-365. **(A)** Spore germination percentage of all assayed strains was presented. The significant difference between Coll-365 and other strains was indicated with a star when $P < 0.5$ appeared in two independent experiments after analysis with One-way ANOVA. **(B)** Spore germination and appressorial formation of strains Coll-365 and IV-1 in a plastic 96-well plate at 24 h post-incubation. Several ungerminated spores and appressoria are indicated.

(Figure 9B). However, no significant difference on sporulation was observed between these transgenic strains and their wild-type strain Coll-365. For pathogenicity, Coll-365 was compared with transgenic strains carrying gene *CsEF1*, *CsLAP*, *CsBGN* or *CsGIP*, and Coll-153 or Coll-524 on chili pepper *Capsicum*

annuum cv. Hero. Lesion size calculation showed that Coll-365 had no notable difference from the tested transgenic strains, and significantly lower virulence than Coll-153 (P -value = 0.023) and Coll-524 (P -value = 0.004). Inoculation assays conducted on *Capsicum annuum* cv. Fushimi-amanaga revealed that strains

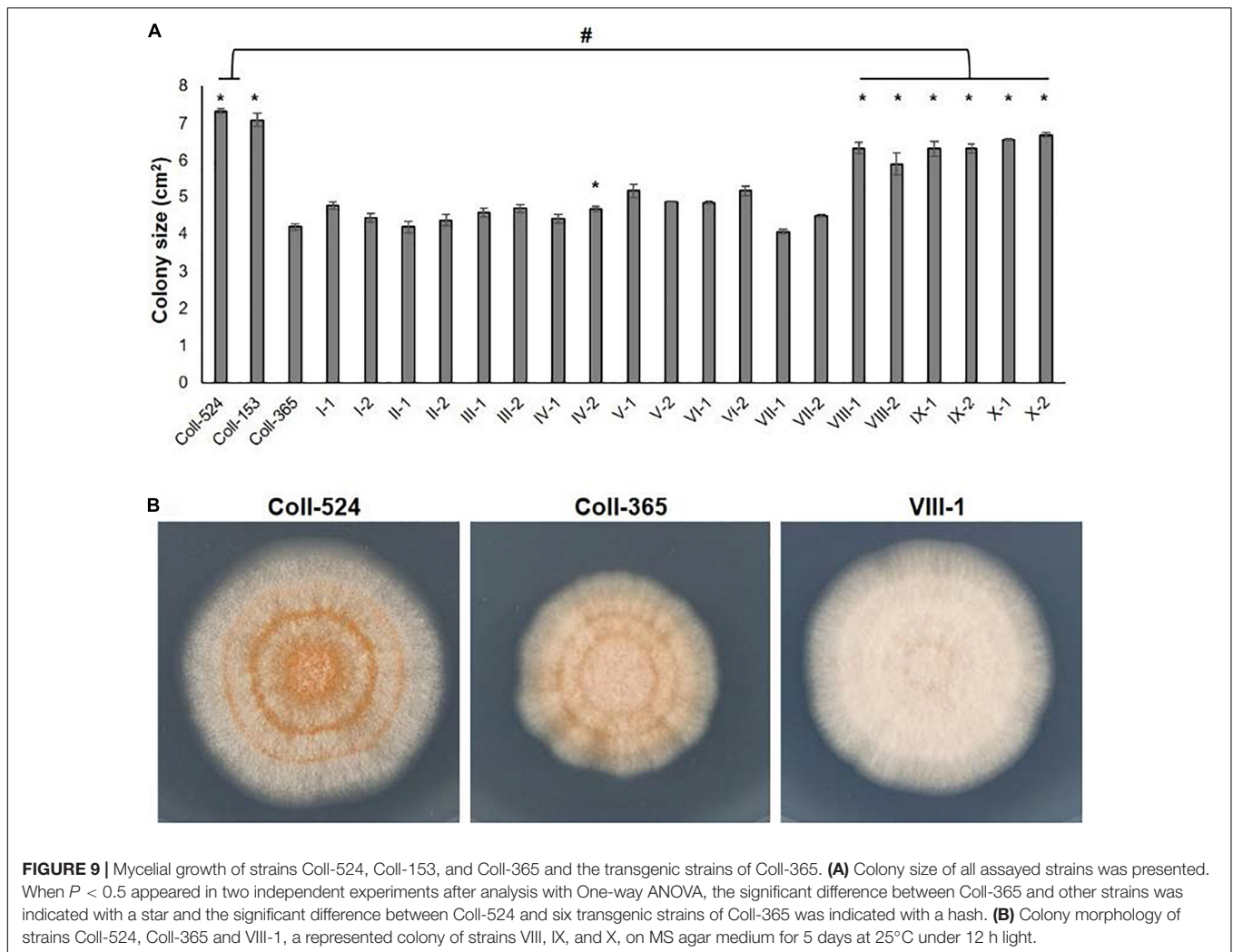


FIGURE 9 | Mycelial growth of strains Coll-524, Coll-153, and Coll-365 and the transgenic strains of Coll-365. **(A)** Colony size of all assayed strains was presented. When $P < 0.5$ appeared in two independent experiments after analysis with One-way ANOVA, the significant difference between Coll-365 and other strains was indicated with a star and the significant difference between Coll-524 and six transgenic strains of Coll-365 was indicated with a hash. **(B)** Colony morphology of strains Coll-524, Coll-365 and VIII-1, a represented colony of strains VIII, IX, and X, on MS agar medium for 5 days at 25°C under 12 h light.

generated from transformation VII - X produced larger lesion sizes than wild-type Coll-365 (Figure 10 and Table 5). The mean lesion sizes were increased between 49 and 400%. The lesion size increase level was similar for each transgenic strain in two independent experiments (Table 5). Transgenic strains of transformation VI and X were further inoculated on *Capsicum annum* cv. Groupzest, and transformation X caused significant lesion size increase (Supplementary Table 12).

DISCUSSION

Colletotrichum species can cause great economic loss to various crops. Among the *Colletotrichum*, more than 30 species have been documented to cause chili anthracnose disease, constituting a major limitation to chili pepper production in tropical and subtropical regions (Saxena et al., 2016). In Taiwan and other Asia countries, *C. scovillei* attacks chili fruits (De Silva et al., 2021), but its interactions with hosts at the molecular genetic level remain to be examined. In this study, we focused on genomic comparisons of the three *C. scovillei* strains and combined

these with genetic approaches to identify genes involved in fungal growth and virulence. We have provided the genome sequences of the three strains with gene functional annotation for *C. scovillei*. We have setup a simple mathematical method to search for ORF variations between Coll-153 or Coll-365 and Coll-524 and successfully identified DNA fragments containing genes involved in the defects in growth and virulence of Coll-365. Moreover, by genetic assay we have demonstrated four genes that have functions in germination, growth and/or virulence of *C. scovillei*.

Our data suggested that the three strains all belong to the acutatum complex and are members of *C. scovillei* because they were grouped in the same clade of *C. scovillei* CBS 126529, the holotype strain of *C. scovillei* (Damm et al., 2012a). The data were consistent with a previous study that showed that Coll-524 and Coll-153 are in the *C. scovillei* clade of acutatum species complex (De Silva et al., 2021).

For closely related species, *de novo* assembly by reference-guided or mapping through a well-sequenced species is efficient, and can often improve the completeness of the genome sequence (Lischer and Shimizu, 2017). Our data revealed the



FIGURE 10 | Anthracnose lesions caused by Coll-365 and its transgenic strain X-1 on *Capsicum annuum* cv. Fushimi-amanaga 5 days after paired inoculation. The photos were used to show the inoculation, symptom development and lesion size calculation in this study. The upper photograph shows the original lesions and the lower photograph shows the marked lesion sizes for Spot image calculation. The inoculation was performed by dropping 5 μ l spore suspensions of Coll-365 on the left side (marked with blue color) and a transgenic strain on the right side (marked with yellow color) of each fruit.

major variation of Coll-153 and Coll-365 was the result of sequence removal. Genetic variations can be caused by three different events, local nucleotide sequence changes, intragenomic rearrangement of DNA segments and the acquisition of a foreign DNA segment by horizontal gene transfer (Arber, 2010). The transposable element is one of the major factors leading to variations in fungi. In *C. higginsianum*, two closely related strains carry large-scale rearrangements and strain-specific regions that are frequently associated with transposable elements (Tsushima et al., 2019). In the two *C. higginsianum* strains, the gene-sparse regions are transposable element-dense regions that have more candidate effector genes, while gene-dense regions are transposable element-sparse regions harboring conserved genes (Tsushima et al., 2019). *C. higginsianum* and other eukaryotic plant pathogens, such as *Phytophthora infestans* and *Leptosphaeria maculans* have been referred to as “two-speed genomes” as these genomes have a compartmentalized genome structure to protect housekeeping genes from the deleterious effects of transposable elements and to provide

rapid evolution of effector genes (Grandaubert et al., 2014; Faino et al., 2016). In the genomes of Coll-524, Coll-153, and Coll-365, we did not find a relationship between the density of transposable elements and candidate effector genes or housekeeping genes. Coll-524 and *C. higginsianum* IMI 349063 have similar genome sizes but *C. higginsianum* IMI 349063 carries nearly three times the number of transposable element genes to Coll-524 (**Supplementary Table 13**). Moreover, compared with *C. higginsianum* IMI 349063, Coll-524 was found to be unlikely to have sequence structures like mini chromosomes which harbor high density transposable elements with over 40% sequences encoding transposable elements (Plaumann and Koch, 2020).

We designed a simple mathematical method to identify ORF-variations. Applying this method, a 34-kb fragment containing 14 genes that exist in the Coll-524 genome but are almost completely lost in Coll-365 were identified. The loss of the 14 genes in Coll-365 is likely caused by DNA deletion that might have resulted from DNA rearrangements. DNA rearrangement frequently occurs in fungi via the parasexual cycle that leads to recombination and chromosome gain or loss (Forche et al., 2008). It is possible that Coll-524 gained part or all of these genes by DNA rearrangement during the parasexual cycle in Coll-153 or Coll-365. Coll-365 and Coll-153 were collected 4 years earlier than Coll-524 from the fields by the ACRDC-the World Vegetable Center (Liao et al., 2012a). Coll-153 and Coll-365 are grouped in the CA1 pathotype and Coll-524 is a member of the CA2 pathotype. CA2 has higher virulence than CA1 and has replaced CA1 to become dominant in Taiwan (Kumar et al., 2018; Sheu et al., 2019). Thus, Coll-524 might have arisen evolutionarily from the CA1 pathotype members such as Coll-153 and Coll-365 by gaining genes horizontally or via gene arrangement. They might be evolutionarily developed from different branches of *C. scovillei*.

Gene family expansions and contractions are related to the changing of host range and virulence of plant pathogens (Baroncelli et al., 2016; Gan et al., 2016). A set of 33 candidate effectors was completely lost in Coll-365, but existed in two closely related chili pepper pathogens *C. acutatum* strain 1 and *C. scovillei* strain TJNH1. This suggests that the 33 candidate effectors might play a role in the virulence of Coll-524 to chili pepper. However, when one of the candidate effectors was transferred to Coll-365, it did not promote the virulence of the transgenic strains on chili pepper fruit, suggesting that multiple effectors may work together to affect fungal virulence (Arroyo-Velez et al., 2020).

CAZyme genes are potentially involved in fungal growth and colonization of host tissues. Coll-524, Coll-153 and Coll-365 have 661–663 CAZyme genes. The gene numbers of CAZyme and proteases vary in different *Colletotrichum* species complexes. Particularly, CAZymes and proteases were found to be associated with *Colletotrichum* spp. that have broad host range (Baroncelli et al., 2016). Research on Magnaporthaceae specific clusters showed CAZyme gene families may contribute to speciation (Okagaki et al., 2016). Transgenic strains carrying endo-beta-1,3-glucanase or laccase did not result in a gain of function of growth and virulence of chili pepper. Endo-beta-1,3-glucanase is expected to play a key role in cell wall softening in ascomycetes.

TABLE 5 | Pathogenicity assay of gene transformation strains (transgenic) and wild-type strain Coll-365 (WT) on fruits of *Capsicum annuum* cv. Fushimi-amanaga.

Experiment	Transgenic strains	N ^a	Mean of lesion size (mm ²)		P-value	Lesion size increased (%)	
			WT	Transgenic		Individual experiment	Means of experiments 1 and 2
1	V-1	5	16.50	19.43	0.2467	18	2
	V-2	6	29.73	23.07	0.1429	-22	-27
	VII-1	4	14.18	26.00	0.0102	83	284
	VII-2	4	18.60	33.90	0.0916	82	71
	VIII-1	7	12.90	60.48	0.0000	369	274
	VIII-2	6	21.95	37.42	0.0814	70	49
	IX-1	8	13.26	49.43	0.0000	273	221
	IX-2	9	24.88	31.04	0.0364	25	54
	X-1	9	10.11	58.87	0.0000	482	400
2	X-2	8	23.97	49.00	0.0000	104	123
	V-1	7	22.72	19.67	0.5799	-13	2
	V-2	6	33.73	23.20	0.1201	-31	-27
	VII-1	5	5.63	32.88	0.0001	484	284
	VII-2	5	19.00	30.40	0.0665	60	71
	VIII-1	7	23.60	65.87	0.0001	179	274
	VIII-2	7	24.80	31.53	0.2146	27	49
	IX-1	8	13.04	35.13	0.0000	169	221
	IX-2	6	14.64	26.86	0.0269	83	54
	X-1	8	11.00	45.89	0.0000	317	400
	X-2	7	12.91	31.26	0.0015	142	123

^aN indicates the numbers of fruits used in each inoculation.

The endo β -1,3-glucanase in *Schizosaccharomyces pombe*, *Candida albicans*, and *Saccharomyces cerevisiae* has been showed to be required for dissolution of the primary septum to allow cell separation (Baladron et al., 2002; Martin-Cuadrado et al., 2003; Esteban et al., 2005). Fungal laccase can detoxify phenolic pollutants and has a role in detoxifying host defense phenolic compounds, such as capsaicin in chili pepper (Strong and Claus, 2011). Coll-524 has five copies of laccase-related genes and two endo-beta-1,3-glucanase and one endo-beta-1,3(4)-glucanase. A single laccase or endo-beta-1,3-glucanase gene may not be able to contribute the function of growth or virulence of Coll-365.

Eight genes were selected for gene transformation in Coll-365. Three genes *CsEF1*, *CsLAP* and *CsBGN* did not show a role in germination, growth and virulence of Coll-365 after transforming into Coll-365. The candidate effector gene *CsEF1*, which encodes a hypothetical protein, was expressed strongly and specifically at the early infection stage, while laccase gene *CsLAP* was expressed highly at all infection stages but not in axenic cultures. Endo-beta-1,3-glucanase gene *CsBGN* was highly expressed in axenic cultures and at the late infection stage. *CsGIP* was expressed at extremely high levels in axenic culture and at all infection stages. The transformation of *CsGIP* to Coll-365 showed no significant influence on fungal growth and virulence but inhibited spore germination ability. *CsGIP* has a non-specific hit (2.06e-05) in the Conserved Domains Database (CCD) of NCBI to COG1196. COG1196 is classified as a model that may span more than one domain and is the only member of the superfamily cl34174 in the CDD. The best hit of gene *CsGIP* in CDD is condensin subunit SMC4 (NP_013187.1) of *Saccharomyces cerevisiae* S288C.

Condensin is a subunit of SMC (Structural Maintenance of Chromosomes) protein complexes that exist in all eukaryotes and play important roles in chromosome organization and dynamics (Losada and Hirano, 2005; D'Ambrosio et al., 2008). Condensin was shown as a major determinant that changes the chromatin landscape as cells prepare their genomes for cell division in fission yeast (Kakui et al., 2017). It is possible that the transgenic Coll-365 carrying *CsGIP* might in some way lead to the changes of chromosome organization and dynamics and then reduce spore germination.

Four genes (*CsBZTF*, *CsCZCP*, *CsPLAA*, and *CsWDCP*) located within the 34-kb fragment at scaffold 17 were transferred into Coll-365 as individual genes or gene sets since they are closely linked. *CsBZTF* had a high expression level in axenic cultures and all infection stages. The other three genes were expressed in axenic cultures and the late infection stage, but genes *CsPLAA* and *CsCZCP* also had relatively high expression in the cuticle infection stage. These four genes did not influence the germination and appressorium formation of Coll-365. However, transgenic strains carrying one *CsPLAA*, three genes *CsBZTF*, *CsCZCP*, and *CsPLAA*, or four genes *CsBZTF*, *CsCZCP*, *CsPLAA*, and *CsWDCP* have similar colony sizes but are larger than the wild-type Coll-365, indicating that *CsPLAA* contributes to the growth enhancement of Coll-365. *CsBZTF* is a bZIP transcription factor and might be highly related to pathogenicity (Yu et al., 2017). The opposite direction *CsCZCP*, 1,313 bp away from *CsBZTF* encodes a C6 zinc finger domain-containing protein. The C6 zinc finger proteins are strictly fungal proteins and involved in fungal-host interactions (MacPherson et al., 2006).

CsBZTF or CsCZCP alone did not affect the virulence of Coll-524. However, transgenic strains carrying the two genes can enhance the virulence of Coll-365, indicating that the co-existence of the two genes together might contribute to the virulence enhancement of Coll-365.

CsPLAA encodes a protein related to phospholipase A2-activation, while gene *CsWDCP* encodes a WD domain-containing protein. Transgenic strains carrying one *CsPLAA*, three genes *CsBZTF*, *CsCZCP*, and *CsPLAA*, or four genes *CsBZTF*, *CsCZCP*, *CsPLAA*, and *CsWDCP* enhance the virulence of Coll-365, indicating *CsPLAA* might be the major contributor to the virulence enhancement of Coll-365. Moreover, transgenic strains carrying genes *CsBZTF*, *CsCZCP*, and *CsPLAA*, or genes *CsBZTF*, *CsCZCP*, *CsPLAA*, and *CsWDCP* have a smaller *P* value than transgenic strains carrying *CsPLAA* only (Table 5), which is consistent with the virulence contribution by *CsBZTF* and *CsCZCP* as mentioned above. *CsPLAA* encodes a protein related to phospholipase A2-activation. Phospholipase A2-activating protein has been relatively intensively studied in humans and is involved in apoptosis and tumor regression; however, there is only a single study related to its functional characterization in fungi (Zhang et al., 2009; Liu et al., 2011). The *Magnaporthe royzae Moplaa* gene encodes a phospholipase A2-activating protein and the deletion of this gene results in the reduction of fungal growth and pathogenicity (Liu et al., 2011).

CONCLUSION

In this study, four genes were identified which function in germination, growth and/or virulence of the chili pepper pathogen *C. scovillei*. *CsGIP* reduces the germination of Coll-365. Genes *CsCZCP* and *CsBZTF* together enhance the virulence of Coll-365. *CsPLAA* enhances the growth and virulence of Coll-365. In addition, the 34-kb fragment in scaffold 17 contributes a lot to the defects in growth and virulence in Coll-365 because three genes *CsPLAA*, *CsBZTF*, and *CsCZCP* are located on this fragment. Interestingly, we also identified 33 candidate effectors lacking in Coll-153 and Coll-365 which may be involved in the full virulence of Coll-524 on chili pepper. Future study will focus on the 33 candidate effectors and other candidate effectors that may have led to Coll-524 becoming a dominant and virulent pathogen.

REFERENCES

- Arber, W. (2010). Genetic engineering compared to natural genetic variations. *N. Biotechnol.* 27, 517–521. doi: 10.1016/j.nbt.2010.05.007
- Arroyo-Velez, N., Gonzalez-Fuente, M., Peeters, N., Lauber, E., and Noel, L. D. (2020). From effectors to effectomes: are functional studies of individual effectors enough to decipher plant pathogen infectious strategies? *PLoS Pathog.* 16:e1009059. doi: 10.1371/journal.ppat.1009059
- Baladron, V., Ufano, S., Duenas, E., Martin-Cuadrado, A. B., del Rey, F., and Vazquez de Aldana, C. R. (2002). Eng1p, an endo-1,3-beta-glucanase localized at the daughter side of the septum, is involved in cell separation in *Saccharomyces cerevisiae*. *Eukaryot. Cell* 1, 774–786. doi: 10.1128/EC.1.5.774-786.2002
- 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* 17:555. doi: 10.1186/s12864-016-2917-6
- Beier, S., Thiel, T., Munch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198
- Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, D344–D354. doi: 10.1093/nar/gkaa977

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI BioProject – PRJNA692809.

AUTHOR CONTRIBUTIONS

D-KH, S-CC, M-HL, and M-CS contributed to the design of the experiments. S-CC designed and prepared the DNA and RNA materials for sequencing. D-KH performed all bioinformatic analyses and gene transformation as well as functional characterizations. Y-TC and C-YC assisted in sequence assembly and annotation. M-YL performed NGS sequencing. M-HL and M-CS supervised the experiments. D-KH, M-HL, and M-CS wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Center for Sustainability Science of Academia Sinica in Taiwan (grant number AS-106-SS-A03), the Ministry of Science and Technology in Taiwan (grant numbers MOST 104-2313-B-005-025-MY3 and MOST 110-2313-B-005-013-MY3), and the Advanced Plant Biotechnology Center from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan.

ACKNOWLEDGMENTS

The authors would like to thank Academia Sinica High throughput Genomics Core for NGS sequencing services, and Miranda Loney for English editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.818291/full#supplementary-material>

- Braga, R. M., Santana, M. F., Veras da Costa, R., Brommonschenkel, S. H., de Araujo, E. F., and de Queiroz, M. V. (2014). Transposable elements belonging to the Tc1-Mariner superfamily are heavily mutated in *Colletotrichum graminicola*. *Mycologia* 106, 629–641. doi: 10.3852/13-262
- Buiate, E. A. S., Xavier, K. V., Moore, N., Torres, M. F., Farman, M. L., Schardl, C. L., et al. (2017). A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species *Colletotrichum graminicola* and *Colletotrichum sublineola*. *BMC Genomics* 18:67. doi: 10.1186/s12864-016-3457-9
- Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I. A., Belmonte, M. K., Lander, E. S., et al. (2008). ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res.* 18, 810–820. doi: 10.1101/gr.7337908
- 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
- Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chao, Y. T., Chen, W. C., Chen, C. Y., Ho, H. Y., Yeh, C. H., Kuo, Y. T., et al. (2018). Chromosome-level assembly, genetic and physical mapping of *Phalaenopsis aphrodite* genome provides new insights into species adaptation and resources for orchid breeding. *Plant Biotechnol. J.* 16, 2027–2041. doi: 10.1111/pbi.12936
- Chen, J. Y., Liu, C., Gui, Y. J., Si, K. W., Zhang, D. D., Wang, J., et al. (2018). Comparative genomics reveals cotton-specific virulence factors in flexible genomic regions in *Verticillium dahliae* and evidence of horizontal gene transfer from *Fusarium*. *New Phytol.* 217, 756–770. doi: 10.1111/nph.14861
- D'Ambrosio, C., Schmidt, C. K., Katou, Y., Kelly, G., Itoh, T., Shirahige, K., et al. (2008). Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. *Genes Dev.* 22, 2215–2227. doi: 10.1101/gad.1675708
- Damm, U., Cannon, P. F., Woudenberg, J. H., and Crous, P. W. (2012a). The *Colletotrichum acutatum* species complex. *Stud. Mycol.* 73, 37–113. doi: 10.3114/sim0010
- Damm, U., Cannon, P. F., Woudenberg, J. H., Johnston, P. R., Weir, B. S., Tan, Y. P., et al. (2012b). The *Colletotrichum boninense* species complex. *Stud. Mycol.* 73, 1–36. doi: 10.3114/sim0002
- Damm, U., Sato, T., Alizadeh, A., Groenewald, J. Z., 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., Ades, P. K., and Taylor, P. W. J. (2021). Pathogenicity of *Colletotrichum* species causing anthracnose of Capsicum in Asia. *Plant Pathol.* 70, 875–884. doi: 10.1111/ppa.13351
- De Silva, D. D., Crous, P. W., Ades, P. K., Hyde, K. D., and Taylor, P. W. J. (2017). Life styles of *Colletotrichum* species and implications for plant biosecurity. *Fungal Biol. Rev.* 31, 155–168. doi: 10.1016/j.fbr.2017.05.001
- de Silva, D. D., Groenewald, J. Z., Crous, P. W., Ades, P. K., Nasruddin, A., Mongkolporn, O., et al. (2019). Identification, prevalence and pathogenicity of *Colletotrichum* species causing anthracnose of *Capsicum annuum* in Asia. *IMA Fungus* 10:8. doi: 10.1186/s43008-019-0001-y
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y
- Esteban, P. F., Rios, I., Garcia, R., Duenas, E., Pla, J., Sanchez, M., et al. (2005). Characterization of the CaENG1 gene encoding an endo-1,3-beta-glucanase involved in cell separation in *Candida albicans*. *Curr. Microbiol.* 51, 385–392. doi: 10.1007/s00284-005-0066-2
- Faino, L., Seidl, M. F., Shi-Kunne, X., Pauper, M., van den Berg, G. C., Wittenberg, A. H., et al. (2016). Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res.* 26, 1091–1100. doi: 10.1101/gr.204974.116
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* 117, 9451–9457. doi: 10.1073/pnas.1921046117
- Forche, A., Alby, K., Schaefer, D., Johnson, A. D., Berman, J., and Bennett, R. J. (2008). The parasexual cycle in *Candida albicans* provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS Biol.* 6:e110. doi: 10.1371/journal.pbio.0060110
- Gan, P., Narusaka, M., Kumakura, N., Tsushima, A., Takano, Y., Narusaka, Y., et al. (2016). Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome Biol. Evol.* 8, 1467–1481. doi: 10.1093/gbe/evw089
- Grandaubert, J., Lowe, R. G. T., Soyer, J. L., Schoch, C. L., Van de Wouw, A. P., Fudal, I., et al. (2014). Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans*-*Leptosphaeria biglobosa* species complex of fungal pathogens. *BMC Genomics* 15:891. doi: 10.1186/1471-2164-15-891
- Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. doi: 10.1186/1471-2105-12-491
- Ingram, T. W., Oh, Y., Adhikari, T. B., Louws, F. J., and Dean, R. A. (2020). Comparative genome analyses of 18 *Verticillium dahliae* tomato isolates reveals phylogenetic and race specific signatures. *Front. Microbiol.* 11:573755. doi: 10.3389/fmicb.2020.573755
- Kakui, Y., Rabinowitz, A., Barry, D. J., and Uhlmann, F. (2017). Condensin-mediated remodeling of the mitotic chromatin landscape in fission yeast. *Nat. Genet.* 49, 1553–1557. doi: 10.1038/ng.3938
- Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108
- Khalidi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., et al. (2010). SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.* 47, 736–741. doi: 10.1016/j.fgb.2010.06.003
- 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
- Kumar, S., Shieh, H. C., Lin, S. W., Schafleitner, R., Kenyon, L., Srinivasan, R., et al. (2018). *Peppers (Capsicum spp.): Domestication and Breeding for Global Use*. Oakville, ON: Apple Academic Press, Inc.
- Lebrigand, K., He, L. D., Thakur, N., Arguel, M. J., Polanowska, J., Henrissat, B., et al. (2016). Comparative genomic analysis of *drechmeria coniospora* reveals core and specific genetic requirements for fungal endoparasitism of nematodes. *PLoS Genet.* 12:e1006017. doi: 10.1371/journal.pgen.1006017
- Lelwala, R. V., Korhonen, P. K., Young, N. D., Scott, J. B., Ades, P. K., Gasser, R. B., et al. (2019). Comparative genome analysis indicates high evolutionary potential of pathogenicity genes in *Colletotrichum tanacetii*. *PLoS One* 14:e0212248. doi: 10.1371/journal.pone.0212248
- Liao, C. Y., Chen, M. Y., Chen, Y. K., Kuo, K. C., Chung, K. R., and Lee, M. H. (2012b). Formation of highly branched hyphae by *Colletotrichum acutatum* within the fruit cuticles of *Capsicum* spp. *Plant Pathol.* 61, 262–270. doi: 10.1111/j.1365-3059.2011.02523.x
- Liao, C.-Y., Chen, M.-Y., Chen, Y.-K., Wang, T.-C., Sheu, Z.-M., Kuo, K.-C., et al. (2012a). Characterization of three *Colletotrichum acutatum* isolates from *Capsicum* spp. *Eur. J. Plant Pathol.* 133, 599–608. doi: 10.1007/s10658-011-9935-7
- Lischer, H. E. L., and Shimizu, K. K. (2017). Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics* 18:474. doi: 10.1186/s12859-017-1911-6
- Liu, X. H., Zhuang, F. L., Lu, J. P., and Lin, F. C. (2011). Identification and molecular cloning *Moplaa* gene, a homologue of *Homo sapiens* PLAA, in *Magnaporthe oryzae*. *Microbiol. Res.* 167, 8–13. doi: 10.1016/j.micres.2011.02.003
- Liu, X., Li, B., Yang, Y., Cai, J., Shi, T., Zheng, X., et al. (2020). Pathogenic adaptations revealed by comparative genome analyses of two *Colletotrichum* spp., the causal agent of anthracnose in rubber tree. *Front. Microbiol.* 11:1484. doi: 10.3389/fmicb.2020.01484
- Liu, Z., and Friesen, T. L. (2012). Polyethylene glycol (PEG)-mediated transformation in filamentous fungal pathogens. *Methods Mol. Biol.* 835, 365–375. doi: 10.1007/978-1-61779-501-5_21
- Losada, A., and Hirano, T. (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes Dev.* 19, 1269–1287. doi: 10.1101/gad.1320505
- Ma, L. J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M. J., Di Pietro, A., et al. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464, 367–373. doi: 10.1038/nature08850
- MacPherson, S., Laroche, M., and Turcotte, B. (2006). A fungal family of transcriptional regulators: the zinc cluster proteins.

- Microbiol. Mol. Biol. Rev.* 70, 583–604. doi: 10.1128/MMBR.0015-06
- Manning, V. A., Pandelova, I., Dhillon, B., Wilhelm, L. J., Goodwin, S. B., Berlin, A. M., et al. (2013). Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. *G3 (Bethesda)* 3, 41–63. doi: 10.1534/g3.112.004044
- Martin-Cuadrado, A. B., Duenas, E., Sipiczki, M., Vazquez de Aldana, C. R., and del Rey, F. (2003). The endo-beta-1,3-glucanase eng1p is required for dissolution of the primary septum during cell separation in *Schizosaccharomyces pombe*. *J. Cell Sci.* 116(Pt 9), 1689–1698. doi: 10.1242/jcs.00377
- Mongkolporn, O., and Taylor, P. W. J. (2018). Chili anthracnose: *Colletotrichum* taxonomy and pathogenicity. *Plant Pathol.* 67, 1255–1263. doi: 10.1111/ppa.12850
- Mullaney, E. J., Hamer, J. E., Roberti, K. A., Yelton, M. M., and Timberlake, W. E. (1985). Primary structure of the trpC gene from *Aspergillus nidulans*. *Mol. Gen. Genet. MGG* 199, 37–45. doi: 10.1007/BF00327506
- Mullins, E. D., Chen, X., Romaine, P., Raina, R., Geiser, D. M., and Kang, S. (2001). Agrobacterium-mediated transformation of *Fusarium oxysporum*: an efficient tool for insertional mutagenesis and gene transfer. *Phytopathology* 91, 173–180. doi: 10.1094/phyto.2001.91.2.173
- Niehaus, E. M., Kim, H. K., Munsterkotter, M., Janevska, S., Arndt, B., Kalinina, S. A., et al. (2017). Comparative genomics of geographically distant *Fusarium fujikuroi* isolates revealed two distinct pathotypes correlating with secondary metabolite profiles. *PLoS Pathog.* 13:e1006670. doi: 10.1371/journal.ppat.1006670
- Okagaki, L. H., Sailsbery, J. K., Eyre, A. W., and Dean, R. A. (2016). Comparative genome analysis and genome evolution of members of the magnaportheaceae family of fungi. *BMC Genomics* 17:135. doi: 10.1186/s12864-016-2491-y
- Perteu, G., and Perteu, M. (2020). GFF utilities: GffRead and GffCompare. *F1000Res* 9:ISC Comm J-304. doi: 10.12688/f1000research.23297.2
- Pilgeram, A. L., and Henson, J. M. (1990). Transformation and cotransformation of *Gaeumannomyces graminis* to phleomycin resistance. *Phytopathology* 80, 1124–1129. doi: 10.1094/phyto-80-1124
- Plumann, P. L., and Koch, C. (2020). The many questions about mini chromosomes in *Colletotrichum* spp. *Plants (Basel)* 9:641. doi: 10.3390/plants9050641
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Saxena, A., Raghuvanshi, R., Gupta, V. K., and Singh, H. B. (2016). Chili anthracnose: the epidemiology and management. *Front. Microbiol.* 7:1527. doi: 10.3389/fmicb.2016.01527
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089
- Sheu, Z.-M., Chiu, M.-H., Chang, R.-J., Wang, J.-F., and Kenyon, L. (2019). “Pathotype dynamics and genetic variability of *Colletotrichum scovillei* causing pepper anthracnose in Taiwan,” in *Poster at the Asian Mycological Congress 2019*. (Taiwan: World Vegetable Center), 3.
- Short, J. M., Fernandez, J. M., Sorge, J. A., and Huse, W. D. (1988). Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res.* 16, 7583–7600. doi: 10.1093/nar/16.15.7583
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., and Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol. Plant Pathol.* 19, 2094–2110. doi: 10.1111/mpp.12682
- Strong, P. J., and Claus, H. (2011). Laccase: a review of its past and its future in bioremediation. *Crit. Rev. Environ. Sci. Technol.* 41, 373–434. doi: 10.1080/10643380902945706
- Syme, R. A., Tan, K. C., Rybak, K., Friesen, T. L., McDonald, B. A., Oliver, R. P., et al. (2018). Pan-parastagonospora comparative genome analysis-effector prediction and genome evolution. *Genome Biol. Evol.* 10, 2443–2457. doi: 10.1093/gbe/evy192
- Talhinhas, P., and Barancelli, 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
- Teixeira, M. M., de Almeida, L. G., Kubitschek-Barreira, P., Alves, F. L., Kioshima, E. S., Abadio, A. K., et al. (2014). Comparative genomics of the major fungal agents of human and animal Sporotrichosis: *Sporothrix schenckii* and *Sporothrix brasiliensis*. *BMC Genomics* 15:943. doi: 10.1186/1471-2164-15-943
- Tsushima, A., Gan, P., Kumakura, N., Narusaka, M., Takano, Y., Narusaka, Y., et al. (2019). Genomic plasticity mediated by transposable elements in the plant pathogenic fungus *Colletotrichum higginsianum*. *Genome Biol. Evol.* 11, 1487–1500. doi: 10.1093/gbe/evz087
- Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S. Y., et al. (2020). PHI-base: the pathogen-host interactions database. *Nucleic Acids Res.* 48, D613–D620. doi: 10.1093/nar/gkz904
- Valero-Jimenez, C. A., Veloso, J., Staats, M., and van Kan, J. A. L. (2019). Comparative genomics of plant pathogenic *Botrytis* species with distinct host specificity. *BMC Genomics* 20:203. doi: 10.1186/s12864-019-5580-x
- van Dam, P., de Sain, M., Ter Horst, A., van der Gragt, M., and Rep, M. (2018). Use of comparative genomics-based markers for discrimination of host specificity in *Fusarium oxysporum*. *Appl. Environ. Microbiol.* 84:e01868-17. doi: 10.1128/AEM.01868-17
- Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Kliuchnikov, G., et al. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319
- Weir, B. S., Johnston, P. R., and Damm, U. (2012). The *Colletotrichum gloeosporioides* species complex. *Stud. Mycol.* 73, 115–180. doi: 10.3114/sim0011
- Yu, P. L., Wang, C. L., Chen, P. Y., and Lee, M. H. (2017). YAP1 homologue-mediated redox sensing is crucial for a successful infection by *Monilinia fructicola*. *Mol. Plant Pathol.* 18, 783–797. doi: 10.1111/mpp.12438
- Zhang, D. D., Wang, J., Wang, D., Kong, Z. Q., Zhou, L., Zhang, G. Y., et al. (2019). Population genomics demystifies the defoliation phenotype in the plant pathogen *Verticillium dahliae*. *New Phytol.* 222, 1012–1029. doi: 10.1111/nph.15672
- Zhang, F., Suarez, G., Sha, J., Sierra, J. C., Peterson, J. W., and Chopra, A. K. (2009). Phospholipase A2-activating protein (PLAA) enhances cisplatin-induced apoptosis in HeLa cells. *Cell Signal.* 21, 1085–1099. doi: 10.1016/j.celsig.2009.02.018
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi: 10.1093/nar/gky418

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Hsieh, Chuang, Chen, Chao, Lu, Lee and Shih. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.