



## OPEN ACCESS

## EDITED BY

Riccardo Baroncelli,  
University of Bologna, Italy

## REVIEWED BY

Joanna Katharine Bowen,  
The New Zealand Institute for Plant and Food  
Research Ltd, New Zealand  
Andrei S. Steindorff,  
Berkeley Lab (DOE), United States

## \*CORRESPONDENCE

Marisa Vieira de Queiroz

✉ mvqueiroz@ufv.br

RECEIVED 03 July 2024

ACCEPTED 18 September 2024

PUBLISHED 15 October 2024

## CITATION

da Silva LL, Correia HLN, Gonçalves OS,  
Vidigal PMP, Rosa RO, Santana MF and  
de Queiroz MV (2024) What lies  
behind the large genome of  
*Colletotrichum lindemuthianum*.  
*Front. Fungal Biol.* 5:1459229.  
doi: 10.3389/ffunb.2024.1459229

## COPYRIGHT

© 2024 da Silva, Correia, Gonçalves, Vidigal,  
Rosa, Santana and de Queiroz. This is an open-  
access article distributed under the terms of  
the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# What lies behind the large genome of *Colletotrichum lindemuthianum*

Leandro Lopes da Silva<sup>1,2</sup>, Hilbert Lucas Nunes Correia<sup>1</sup>,  
Osiel Silva Gonçalves<sup>1</sup>, Pedro Marcus Pereira Vidigal<sup>3</sup>,  
Rafael Oliveira Rosa<sup>1</sup>, Mateus Ferreira Santana<sup>1</sup>  
and Marisa Vieira de Queiroz<sup>1\*</sup>

<sup>1</sup>LGMM, Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, Viçosa, Brazil, <sup>2</sup>Crop Production and Pest Control Research Unit, Agricultural Research Service, United States Department of Agriculture (USDA), West Lafayette, IN, United States, <sup>3</sup>NuBioMol, Centro de Ciências Biológicas, Universidade Federal de Viçosa, Viçosa, Brazil

*Colletotrichum lindemuthianum* is the etiological agent of anthracnose disease in common bean (*Phaseolus vulgaris* L.), noted for its ability to cause serious damage and significant pathogenic variability. This study reveals the features of the high-quality genome of *C. lindemuthianum*. Analysis showed improvements over the first assembly, with the refined genome having 119 scaffolds, ten times fewer than the first, and a 19% increase in gene number. The effector candidates increased by nearly 1.5 times. More than 40% of the amino acid sequences with homologs in the Pathogen-Host Interactions (PHI-base) database are linked to pathogenicity. Of 18 putative proteins identified as Chitinase-like Protein, six have a mutation in the enzyme catalytic motif, and three showed gene expression in the biotrophic phase, indicating they can act as effectors. Comparative genomic analyses with 30 other fungal species revealed that *C. lindemuthianum* is among the top three fungi encoding transport proteins. Seven Necrosis and Ethylene-Inducing Peptide 1 (Nep1)-Like Proteins (NLPs) are present in the *C. lindemuthianum* genome, but none had complete identity with the GHRHDWE conserved motif of NLPs; two were grouped with proteins that induce necrosis and may retain the capability to induce host necrosis. *Colletotrichum* species show a high number of secondary metabolite (SM) clusters, with *C. lindemuthianum* having 47 SM clusters. Approximately 60% of the *C. lindemuthianum* genome is composed of repetitive elements, a significantly higher proportion than in other fungi. These differences in transposable element (TE) numbers may explain why *C. lindemuthianum* has one of the largest genomes among the fungi analyzed. A significant portion of its genome comprises retroelements, particularly the *Ty1/Copia* superfamily, which accounts for 22% of the genome and represents 40% of the repetitive elements. The genomic profile features a remarkably high RIP-affected genomic proportion of 54.77%, indicating substantial RIP activity within this

species. This high-quality genome of *C. lindemuthianum*, a significant pathogen in common bean cultivation, will support future research into this pathosystem, fostering a deeper understanding of the interaction between the fungus and its host.

#### KEYWORDS

common bean anthracnose, comparative genomics, repetitive elements, repeat-induced point (RIP) mutations, long read DNA sequencing

## 1 Introduction

Fungi from the *Colletotrichum* genus are responsible for causing anthracnose in crops worldwide, being a relevant pathogen in both pre-harvest and post-harvest, mainly in tropical and subtropical regions (Cai et al., 2011). The genus is among the ten most relevant phytopathogenic fungi when considering its biological characteristics and economic impact (Dean et al., 2012). It is also recognized as a great model for studies on host-pathogen interaction (Perfect et al., 1999).

Although Brazil is among the top three bean (*Phaseolus vulgaris* L.) producers in the world (FAOSTAT, 2022), several factors can contribute to decreasing its productivity, among them, diseases caused by fungi, such as anthracnose (Perseguiuni et al., 2016). The common bean anthracnose caused by *Colletotrichum lindemuthianum* (Saccardo & Magnus) Briosi & Cavara is one of the major diseases of this crop in Brazil, which can cause significant losses in production (Barcelos et al., 2014). With favorable conditions of humidity and temperature, anthracnose results in premature defoliation, premature flowering, and pods falling, and it can lead to plant death in extreme cases, decreasing productivity and causing grain depreciation (Sharma et al., 2008).

The fungus *C. lindemuthianum* has high genetic variability, with at least 182 physiological races identified in different world regions (Padder et al., 2017). Thus, over the years, several studies investigated the genetic and pathogenic variability of *C. lindemuthianum* (Mahuku and Riascos, 2004; Sharma et al., 2007; Ishikawa et al., 2010; Pinto et al., 2012). This fungus and its association with common bean has been studied with research focusing on the cytology and physiology of common bean infection (O'Connell et al., 1985, 1996; O'Connell, 1987; O'Connell and Ride, 1990); conidia anastomosis and parasexuality (Roca et al., 2003; Ishikawa et al., 2010, 2013; Rosada et al., 2010); sexual reproduction (Souza et al., 2010); differential gene expression during pathogenesis (Fontenelle et al., 2017; Romero et al., 2024); identification of genes required for pathogenicity (Parisot et al., 2002; Pellier et al., 2003; Hoi et al., 2007; Cnossen-Fassoni et al., 2013; Soares et al., 2014; Nogueira et al., 2019); analysis of mtDNA and genes encoding

effector proteins and pectinases (Queiroz et al., 2018, 2019; Silva et al., 2021, 2022), among others.

With the expanding knowledge of fungal genomes, comparative genomics can be applied to analyze the organization and composition of the *Colletotrichum* genome, facilitating a better understanding of its relationship to lifestyle. Several gene families are more abundant in *Colletotrichum* genomes than observed in other phytopathogenic fungi. Comparative genomics studies make it possible to analyze the gene repertoire linked to secondary metabolism, the enzymes related to plant cell wall degradation, pathogen virulence, and effector candidates to understand the pathogenic process (O'Connell et al., 2012; Gan et al., 2016; Rao and Nandineni, 2017; Bhadauria et al., 2019; Lelwala et al., 2019). In addition, *Colletotrichum* species have a high potential for secondary metabolite production, surpassing several other phytopathogenic fungi (Gan et al., 2012; O'Connell et al., 2012; Baroncelli et al., 2016; Liang et al., 2018).

The orbiculare complex has hemibiotrophic species, which cause diseases in herbaceous plants. Among these species, *Colletotrichum orbiculare*, *C. lindemuthianum*, *Colletotrichum trifolii*, *Colletotrichum spinosum*, and *Colletotrichum sidae* have already had their genomes sequenced (Queiroz et al., 2017; Gan et al., 2019). These species have distinct characteristics from other *Colletotrichum* species, such as an expansion in genome size not accompanied by gene number increase and G+C content (%) reduction (Gan et al., 2012, 2019; Queiroz et al., 2017).

Most *Colletotrichum* genomes have been sequenced using the Illumina platform, including *C. lindemuthianum* (Queiroz et al., 2017), but these genomes have gaps due to small and repetitive reads which are difficult to assemble using this technology (English et al., 2012). The *C. lindemuthianum* genomes sequenced are among the largest ever reported in the *Colletotrichum* genus (Queiroz et al., 2017), surpassed only by the *C. trifolii* genome (Gan et al., 2019), which is a drawback when trying to obtain genomes with better quality. Alternative sequencing strategies, such as long read sequencing, facilitate the assembly of larger genomes (English et al., 2012). Due to the importance of *C. lindemuthianum* as a phytopathogen, the objective of the present work was to present a high-quality genome to reveal its remarkable genomic characteristics and to identify virulence

factors involved in common bean infection through comparative genomics with other fungi.

## 2 Materials and methods

### 2.1 Microorganism and growth conditions

The *C. lindemuthianum* isolate A<sub>2</sub> 2-3 physiological race 89 analyzed in this study is from the collection of the Laboratory of Molecular Genetics of Microorganisms/Bioagro from Universidade Federal de Viçosa, Viçosa, Brazil. For culture maintenance and sporulation, fungal cultivation was on potato dextrose agar (PDA) (Kasvi) or YMC (malt extract 10 g.L<sup>-1</sup>; yeast extract 2 g.L<sup>-1</sup>; agar 15 g.L<sup>-1</sup>) medium at 22°C for six to eight days. For growth in liquid medium, approximately 10<sup>5</sup> conidia.mL<sup>-1</sup> were inoculated in GPYECH medium (20 g.L<sup>-1</sup> glucose; 5 g.L<sup>-1</sup> peptone; 1.0 g.L<sup>-1</sup> yeast extract; 1.0 g.L<sup>-1</sup> hydrolyzed casein) and cultured with shaking at 150-200 rpm for four days at 22°C.

### 2.2 Nucleic acid extraction and sequencing

Total DNA of the *C. lindemuthianum* isolate was extracted from fresh mycelia after four days of growth in GPYECH medium with the NucleoBond™ Buffer Set IV in combination with NucleoBond™ AXG 100 columns (Macherey-Nagel Bioanalysis™) following the manufacturer's instructions. The *C. lindemuthianum* isolate genomic DNA was sequenced using the PacBio Sequel platform (CD Genomics, New York, United States). The RNA of bean leaves inoculated with the *C. lindemuthianum* at different times post inoculation (48 h and 120 h) and in GPYECH medium were obtained as previously described (Queiroz et al., 2019). The RNA was sequenced using the Illumina HiSeq platform with three biological replicates per time point (GenOne Soluções em Biotecnologia, Rio de Janeiro, Brazil).

### 2.3 Genome assembly

The genome assembly used a hybrid approach combining the previously short paired-read data generated using Illumina HiSeq 2500 (Queiroz et al., 2017) and long read data obtained using PacBio Sequel subreads. The quality assessment of the short reads used the reports of FastQC v. 0.11.9 (Andrews, 2010). Trimgalore v. 0.6.7 (Krueger et al., 2015) removed the library adapters using the “auto-detection” setting, and Trimmomatic v. 0.39 (Bolger et al., 2014) trimmed and filtered the reads.

The MaSuRCA v. 4.0.3 (Zimin et al., 2013) executed the hybrid assembly using trimmed short and long reads data, combining SOAPdenovo2 (Luo et al., 2012) v. 2.04 to assembly the short reads and Flye v. 2.5 (Kolmogorov et al., 2019) to assembly the long reads. The scaffolds were corrected by mapping the paired reads using the BWA-MEM algorithm of BWA v. 0.7.17 (Li and Durbin, 2009), sorting the mapping files using the Picard toolkit v. 2.26.2 (<https://github.com/broadinstitute/picard>), and polishing the sequences

using the Pilon v. 1.24 (Walker et al., 2014), with default parameters.

### 2.4 Structural annotation

The genome annotation followed the pipeline of Funannotate v. 1.8.7 (Palmer and Stajich, 2020) and combined *ab initio* and evidence-based methods, using transcriptome data and homology with known proteins for gene prediction. First, the Repeat-Modeler v. 2.0.1 (Flynn et al., 2020) executed an *ab-initio* prediction of transposable elements (TEs) and repetitive elements to produce a library from the scaffolds. Then, RepeatMasker v. 4.1.2-p1 (Tarailo-Graovac and Chen, 2009) executed a soft-masking of these elements in the scaffold sequences using the predicted library. Before the gene prediction, the tool “train” of Funannotate aligned RNA-seq data in the scaffold sequences using the option “jaccard-clip” since high gene density in fungal genomes is expected. After training, the tool “predict” of Funannotate predicted the genes using protein sequences of the predicted reference proteomes of the *Colletotrichum* genus (Taxon ID 5455) available in the UniProt database (<https://www.uniprot.org/proteomes>). Then, the tool “update” of Funannotate added UTR data to the predictions and fixed the gene models. BUSCO (Benchmarking Universal Single-Copy Orthologs) (Simão et al., 2015) evaluated the genome and gene prediction completeness using the proteome mode and the orthologous dataset of the Fungi kingdom (fungi\_odb10) and Glomerellales order (glomerellales\_odb10).

### 2.5 Functional annotation

The pipeline for secretome prediction executed sequential filtering steps with the following software: SignalP v. 4.1 (Petersen et al., 2011), Target-P v. 1.1 (Emanuelsson et al., 2000), and WoLF PSORT (Horton et al., 2007) to identify signal peptides and predict the subcellular location; TMHMM v. 2.0 (Emanuelsson et al., 2007) and Phobius v. 1.1 (Käll et al., 2007) to exclude sequences with transmembrane regions; and PredGPI (Eisenhaber et al., 2004) to exclude amino acid sequences that have GPI anchors.

A search against the MEROPS database (Rawlings and Bateman, 2021) identified putative proteases and protease inhibitors, excluding sequences with incomplete domains and mutated active sites. dbCAN (Yin et al., 2012) identified the carbohydrate active enzymes (CAZymes). Within the CAZyme analysis, chitinase enzymes were presumed to belong to glycoside hydrolase family 18 (GH18). We used the first predicted proteome of *C. lindemuthianum* (Queiroz et al., 2017) and the newly predicted proteome to conduct an improved effector analysis using Predector v1.2.7 (Jones et al., 2021). The pathogenesis interactions database (PHI-base) (Winnenburg, 2006) allowed the identification of candidate pathogenicity genes, and cytochrome P450 (CYP450) enzymes were screened using the cytochrome P450 database (Nelson, 2009). A local BLAST search using the CDS and amino acid sequences of MAT1-1 from *Fusarium graminearum*

(AF318048.1) and MAT1-2-1 from *C. lindemuthianum* isolate MU 03 (EU236950.1) was performed against the *C. lindemuthianum* isolate A<sub>2</sub> 2-3 genome assembly to determine its mating type.

BLASTp was used to identify proteins similar to Necrosis and Ethylene-Inducing Peptide 1 (Nep1)-Like Proteins (NLPs) to identify putative homologs of known NLP proteins PsojNIP (type 1) from *Phytophthora sojae* (Qutob et al., 2002), NLPpcc (type 2) from *Pectobacterium carotovorum* subsp. *carotovorum* (Mattinen et al., 2004), and Afu5g02100 (type 3) from *Aspergillus fumigatus* (Upadhyay et al., 2012). Sequence alignments using Clustal Omega (Sievers et al., 2011) allowed the identification of the conserved consensus motif GHRHDWE in the amino acid sequence of *C. lindemuthianum* NLPs. To confirm the homology of the *C. lindemuthianum* NLP candidates, a Maximum Likelihood (ML) analysis was conducted in MEGA v. 11 (Tamura et al., 2021). The Clustal tool aligned the amino acid sequences of NLP proteins of *C. lindemuthianum* and other fungal species, and the ML analysis was performed using the JTT matrix-based substitution model with bootstraps of 10,000 replicates to infer the confidence of the phylogenetic tree.

Transporter prediction used TMHMM v. 2.0 (Emanuelsson et al., 2007) to identify proteins with transmembrane regions, and similarity searches using the BLASTp tool to search for putative homologs in the Transporter Classification Database (TCDB) (Saier et al., 2021). The Secondary Metabolite Analysis SHell (antiSMASH fungal version) (Medema et al., 2011) was used to scan the genome to identify potential secondary metabolite (SM) gene clusters, and RIPper (Van Wyk et al., 2019) was used to analyze the genome to identify Repeat-Induced Point (RIP) mutations.

To provide a broader overview of functional annotation, protein sequences annotated in genomes from 29 fungal species available in the NCBI Genome database (<https://www.ncbi.nlm.nih.gov/genome/>) (Supplementary Table S1) were used for comparative analysis. In addition, the RNA-seq data of bean leaves inoculated with the *C. lindemuthianum* isolate A<sub>2</sub> 2-3 at different times post inoculation (48 h and 120 h) allowed the investigation of the expression patterns of genes involved in pathogenicity. The gene expression analysis followed the methodology previously described by Silva et al. (2022), whereas the genome of *C. lindemuthianum* was indexed using Bowtie2 v software. 2.2.8 (Langmead et al., 2009) in default settings and mapped against each transcriptome library

using Tophat v.2.1.1 (Trapnell et al., 2009). Transcriptome assembly and gene expression quantification were performed using Cufflinks v.2.2.1 (Trapnell et al., 2010). The fragments per kilobase of transcript per million mapped reads (FPKM) values of the gene candidates were used to analyze gene expression levels.

## 3 Results

### 3.1 *Colletotrichum lindemuthianum* genome assembly and gene prediction

To obtain a more refined genome, a hybrid assembly was performed using the previous data generated by Illumina sequencing (Queiroz et al., 2017) and the new data from the PacBio platform. The new genome assembly has only 119 scaffolds, ten times less than the first assembly (Table 1). The scaffold length in the new assembly is greater than the previous one. Gene prediction was performed using Funannotate integrating specific RNA-Seq evidence from *C. lindemuthianum*, protein homology from other *Colletotrichum* species, and *ab initio* gene predictions to obtain consensus gene models. A total of 13,821 genes were predicted from the new assembly, an increment of almost 19% in the number of previously predicted genes. A total of 14,126 putative proteins were predicted from the analyzed genome. The greater putative protein number than the predicted gene number is related to alternative splicing variants. The new assembly completeness was accessed through BUSCO, which showed more than 95% completeness when compared with genes from both Fungi and Glomerellales. This new and more refined assembly makes it possible to better study and understand the *C. lindemuthianum* genome.

### 3.2 Predicted secretome

The prediction of putatively secreted proteins in *C. lindemuthianum* was performed by a sequence of analyses. Initially, signal peptide prediction was performed for the predicted *C. lindemuthianum* proteins. Thus, 1,346 putative proteins showed a secretion signal and did not have the mitochondria and other sub-cellular regions predicted to be their final destination. Putative proteins that presented transmembrane helices and that had glycosylphosphatidylinositol anchor motifs (GPI-anchor) were also excluded. Thus, a total of 1,054 putative protein sequences were predicted to be secreted, which corresponds to approximately 7.5% of all *C. lindemuthianum* putative proteins.

### 3.3 CAZymes

Seven hundred and ten putative proteins were identified as CAZymes in the *C. lindemuthianum* predicted proteome (Figure 1). Of these, 300 putative proteins classified into 70 CAZyme families were predicted to be secreted, corresponding to 42% of the CAZymes present in the proteome. The CAZyme classes glycosyl hydrolase (GH), polysaccharide lyase (PL), auxiliary activities (AA), and carbohydrate

TABLE 1 Comparing the *C. lindemuthianum* genome assembly using different approaches.

	Illumina	Pacbio + Illumina
Scaffold number	1,276	119
N50	158 kbp	2.1 Mbp
Smallest scaffold	1.03 kbp	3.6 Kbp
Biggest scaffold	1.11 Mbp	9.3 Mbp
Genome Size	99.16 Mbp	100.48 Mbp
GC Content	37.3%	37.25%
Predicted gene number	11,627	13,821

esterase (CE) had the greatest number of secreted protein members. The three most abundant AA families with the greatest number of members were those comprised of putative oxidases, monooxygenases, and carbohydrate dehydrogenases. The enzymes present in the three PL families with the most members had activities related to pectin degradation. The enzymes present in the four most abundant GH families with the most members had activities related to the degradation of cellulose, hemicellulose, and pectin. The enzymes present in the three CE families with the most members had activities related to xylan acetyltransferase. Of those secreted CAZymes involved in plant cell wall degradation, there was a greater abundance of enzymes related to pectin degradation than the other constituents.

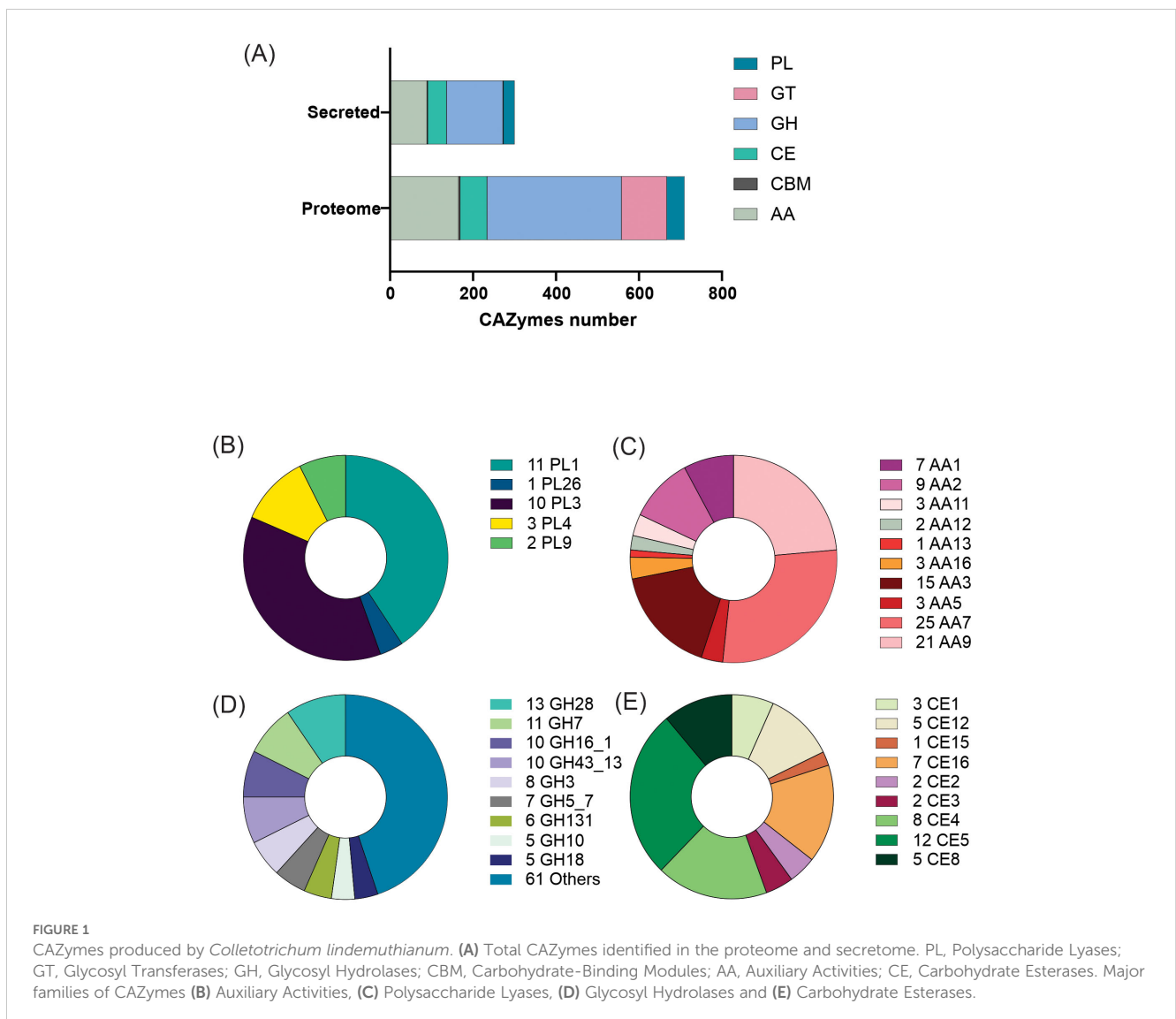
### 3.4 Proteases

Using BLASTp against the MEROPS database, it was possible to identify 107 putative proteases in the *C. lindemuthianum* secretome

(Figure 2A). Serine-proteases are the largest group with 63 proteins, followed by metallopeptidases with 30 proteins, which correspond to almost 87% of the total secreted proteases. In addition, 19 putative proteins that showed similarity with non-peptidases and two proteins considered protease inhibitors were identified in the secretome.

### 3.5 Cytochrome P450s

A total of 157 putative proteins with similar proteins in the cytochrome P450 database were identified (Figure 2B). These proteins were distributed among 74 CYP450 families. The CYP65 family has the most putative protein members, with 21 putative proteins, followed by the CYP68 family with 9 members, corresponding to 19% of putative CYP450 proteins. CYP65 and CYP68 are related to secondary metabolism. Most families (45 families) comprise one putative protein.



### 3.6 Sequences encoding proteins related to pathogenicity

In *C. lindemuthianum* 4,343 amino acid sequences with similarity to proteins in the PHI database were identified (Figure 2C), with most of the similar proteins having demonstrated effects on pathogenicity in other pathogens. While almost 40% of the identified putative proteins showed similarity with other proteins with no effect on pathogenicity. Of the 4,343 putative proteins, 1,749 were associated with reduced virulence, 208 with loss of pathogenicity, 76 with hypervirulence, and 42 with effector functions in the related pathogens.

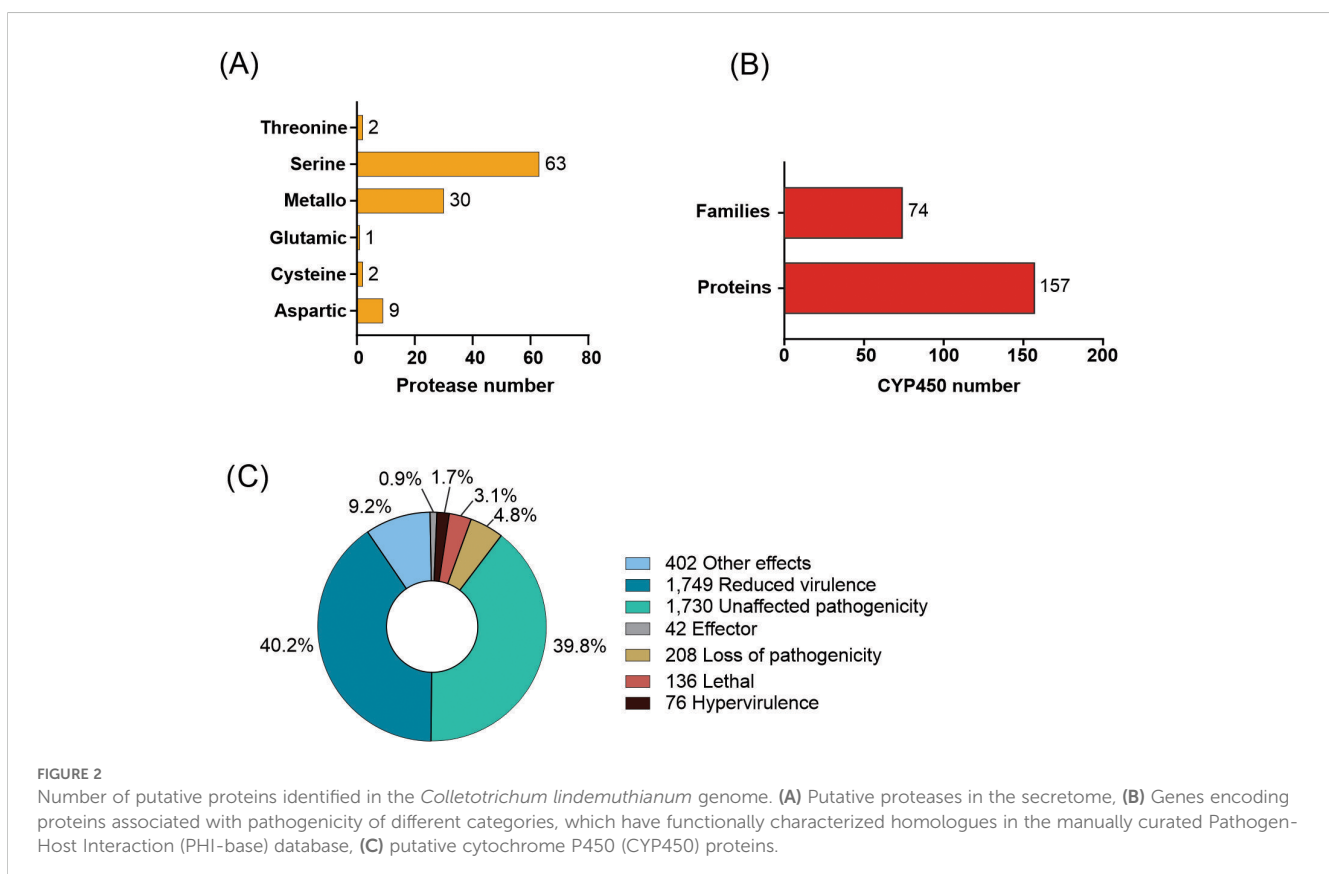
### 3.7 Effector candidates

Previously, 349 candidate effector genes were found in the genome of *C. lindemuthianum* (Queiroz et al., 2019). The analysis was redone using Predector, an automated and combinative method for the predictive ranking of candidate effector proteins (Jones et al., 2021), and compared to the previous analysis (Supplementary Figure S1). This new prediction revealed a slight variation in the number of effector candidates; 373 were found originally, while the new assembly increased the number of effector candidates by nearly 1.4X (525 effectors). Alignment comparison showed that 262 of the candidate effector proteins from Queiroz et al. (2019) were similar to the newly predicted effectors, with 183

proteins sharing 100% similarity (Supplementary Table S2). In addition, all the effector candidates in the current assembly were found to be putatively secreted, being 45% (n= 238) cytoplasmic effectors, 50% (n= 264) apoplasmic effectors, and 28% (n= 148) dual-localized effectors (apoplasmic/cytoplasmic). Most effector candidates had no similarities to proteins in the PHI-base database and were predicted to be subcellularly localized to the nucleus (n= 78), chloroplast (n= 43), and mitochondria (n= 16). Information on the effector candidates may be found in Supplementary Tables S3 and S4.

### 3.8 Chitinase-like proteins

Some chitinase-like proteins can act as effectors, when harboring mutations in their catalytic motifs along with high expression of their encoding genes during the initial phases of a plant-pathogen interaction (Fiorin et al., 2018). Based on this information, a search was carried out to identify chitinase-like proteins in *C. lindemuthianum*. As a result, 18 putative proteins similar to chitinases were identified (Figure 3). Of these putative proteins, three were mutated in one amino acid present in the enzyme catalytic motif (xDxxDxDxE), with the replacement of aspartic acid (D) by asparagine (N), glutamine (Q), or valine (V). Three other putative proteins had a methionine (M) substitution for glutamine (Q) at a conserved residue. These mutated enzymes may have lost chitinolytic activity. Eight genes that encode chitinase-like



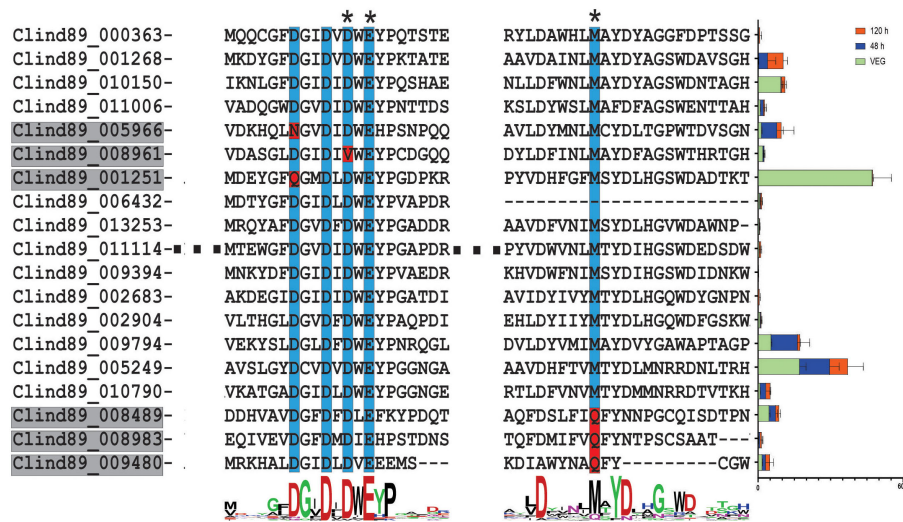


FIGURE 3

The *Colletotrichum lindemuthianum* chitinase-like proteins. The catalytic site of chitinase is DxExE. Conserved amino acids are indicated by the blue color. Amino acid substitution in conserved regions is highlighted in red. Asterisk (\*) indicates that amino acid substitution at the position results in loss of catalytic activity of the enzyme (Fiorin et al., 2018). At the end of each amino acid sequence, the expression of genes encoding chitinases (FPKM) at two stages of the *C. lindemuthianum*-*Phaseolus vulgaris* interaction (48 h and 120 h) and grown in GPYECH culture medium (VEG) is represented.

proteins were expressed in the biotrophic phase of the host-pathogen interaction. Of the six mutated genes, three were expressed in the biotrophic phase.

### 3.9 Comparative genomic analysis

While certain gene groups related to pathogenicity have been analyzed in *C. lindemuthianum*, specific groups have been selected for comparative analysis, offering a broader perspective on the

representation of these genes in comparison to other phytopathogenic fungi. Overall, comprehending the diversity and functionality of transporters, NLPs, SM gene clusters, TEs, and RIP mutations through comparative genomic analysis can yield insights into genome dynamics, evolution, and adaptation in phytopathogenic fungi.

#### 3.9.1 Transporters

*C. lindemuthianum* possesses a considerable number of transporters, totaling 1,323 putative proteins. Among the fungi studied, *Fusarium oxysporum* exhibited the highest number (n =

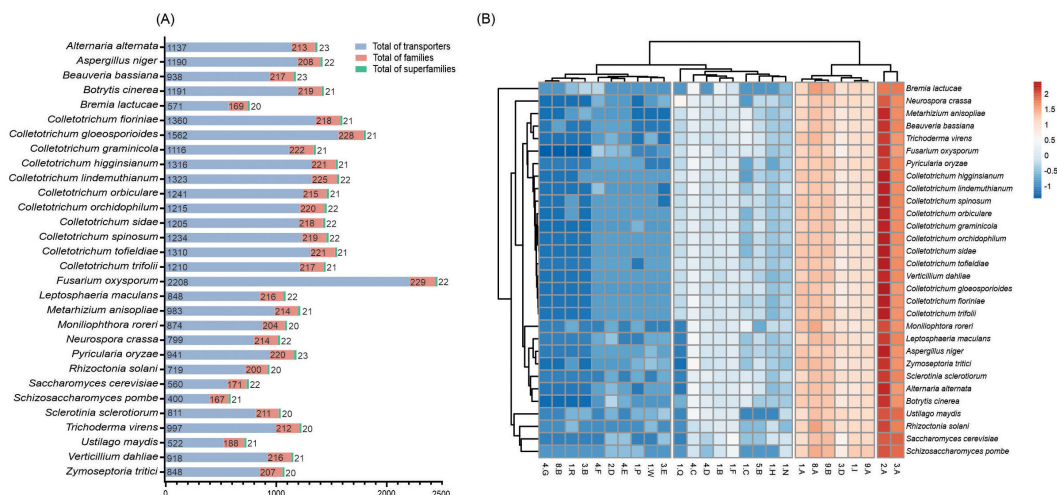


FIGURE 4

Repertoires of transporter proteins in the fungal genomes. (A) The comparative analysis of transporters, families, and superfamilies transporters in different fungal predicted proteomes. (B) Distribution of superfamilies transporters in different fungal predicted proteomes. Original values are  $\ln(x)$ -transformed. Rows are centered and clustered using correlation distance and average linkage, while columns are clustered using Euclidean distance and average linkage.

2,208), while *Ustilago maydis* encoded the fewest transporters ( $n = 522$ ). These transporters were categorized into an average of 233 families and 23 superfamilies (Figure 4). Across *Colletotrichum* species, a consistent range of transporter numbers was noted, varying from 1,116 in *C. graminicola* to 1,562 in *Colletotrichum gloeosporioides*. Remarkably, fungal genomes encode for three major superfamilies, the major facilitator superfamily (MFS) (2.A.1), the ATP-binding Cassette (ABC) Superfamily (3.A.1), and the Membrane Fusion Protein (MFP) Family (8.A.1). These transporters are key classes that include uniporters, symporters, and antiporters involved in nutrient uptake and ion movement (Tanford, 1983; Forrest et al., 2011). They represent about 46, 15, and 7% of the transporter repertoire in *C. lindemuthianum*, respectively. In general, the Mitochondrial Carrier (MC) Family and the Amino Acid-Polyamine-Organocation (APC) Family were those with the most members in the MFS superfamily, followed by the  $H^+$  or  $Na^+$ -translocating NADH Dehydrogenase (NDH) family in the ABC superfamily and the Voltage-gated  $K^+$  Channel  $\hat{I}^2$ -subunit ( $Kv\hat{I}^2$ ) family in the MFP superfamily.

### 3.9.2 Necrosis and ethylene-inducing peptide 1 - like proteins

NLPs are present in microorganisms from different taxa that adopt diverse lifestyles (Figure 5). The number of putative NLPs in the fungi analyzed ranged from zero to 14. Seven putative NLPs

were identified in *C. lindemuthianum*. The number of putative NLPs in other *Colletotrichum* species ranged from five to 14. Thirteen of the 30 analyzed fungi showed from zero to three NLP. The fungi with few NLPs were the basidiomycetes *Rhizoctonia solani* and *U. maydis*, and the ascomycetes yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* with no NLPs. Those fungi with the greatest number of NLPs were the ascomycetes *Verticillium dahliae*, *Colletotrichum fioriniae*, and the oomycete *Bremia lactucae* with 10, 14, and 13 NLPs, respectively.

The seven putative NLPs in *C. lindemuthianum*, all had a mutated GHRHDWE conserved motif (Figure 6). The motif was 85% conserved in two putative proteins (Clind89-006653-T1 and Clind89-005275-T1). All *C. lindemuthianum* putative NLPs had the conserved Histidine (H) at the same position as the consensus motif. Amino acid substitution at some position in the GHRHDWE motif may cause abolition or reduction in necrosis activity (Ottmann et al., 2009). Only the Clind89-005275-T1 putative protein had no mutation in any of the three amino acid positions that when mutated may cause necrosis activity abolition. The gene encoding Clind89-006653-T1 had the highest genic expression at 120 h post inoculation during the *C. lindemuthianum*-*Phaseolus vulgaris* interaction, corresponding to the necrotrophic phase of the disease.

A phylogenetic reconstruction of three NLP types from fungi related to *C. lindemuthianum* (*Colletotrichum higginsianum*, *C.*



FIGURE 5

The comparative analysis of necrosis and ethylene-inducing peptide 1. (Nep1)-like proteins (NLPs) in different fungal predicted proteomes.



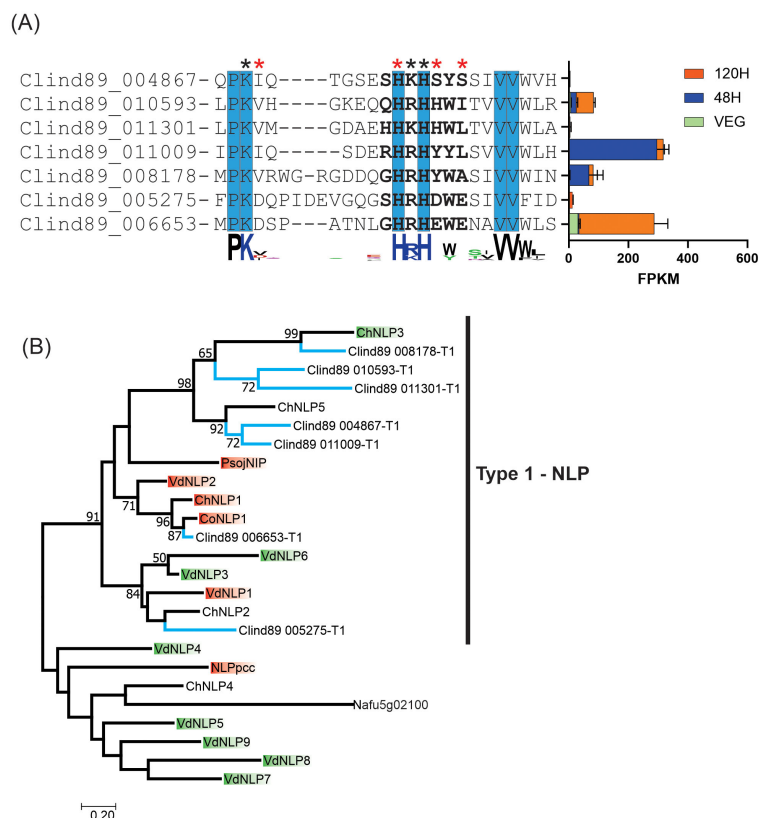


FIGURE 6

The *Colletotrichum lindemuthianum* necrosis and ethylene-inducing peptide 1 (Nep1)-like proteins (NLPs) and phylogenetic reconstruction of NLPs. (A) The conserved characteristic motif of NLPs is shown in bold (GHRHDWE). Invariant amino acids near or in the GHRHDWE motif are highlighted in blue. Red asterisks (\*) indicate that substitution of amino acids at the position causes abolition in necrosis activity and black asterisks indicate reduction (Ottmann et al., 2009). At the end of each amino acid sequence, the expression of genes encoding NLPs (FPKM) in two stages of the *C. lindemuthianum*-*Phaseolus vulgaris* interaction (48 h and 120 h) and grown in GPYEC culture medium (VEG) is represented. (B) Blue branches indicate the NLPs of *C. lindemuthianum* ChNLP1. Red boxes indicate necrosis activity while green boxes indicate absence. ChNLPs are from *Colletotrichum higginsianum* (Kleemann et al., 2012), CoNLP1 from *Colletotrichum orbiculare* (Azmi et al., 2018), VdNEPs from *Verticillium dahliae* (Zhou et al., 2012), PsojNIP (type 1) from *Phytophthora sojae* (Qutob et al., 2002), NLPpcc (type 2) from *Pectobacterium carotovorum* subsp. *carotovorum* (Mattinen et al., 2004) and Afu5g02100 (type 3) from *Aspergillus fumigatus* (Upadhyay et al., 2012). Maximum Likelihood phylogenetic analysis was performed with JTT matrix-based substitution model and 10,000 replicates of bootstrap.

*orbiculare*, and *V. dahliae*) showed a clade including all NLPs from *C. lindemuthianum* together with type 1 NLP (PsojNIP) from *P. sojae* (Figure 6). Some of the NLPs used for phylogenetic reconstruction were analyzed for their capability to cause necrosis (Qutob et al., 2002; Kleemann et al., 2012; Zhou et al., 2012; Azmi et al., 2018). The putative proteins Clind89 005275-T1 and Clind89-006653-T1 were grouped with one and three proteins with a proven ability to induce necrosis, respectively. These putative proteins had the highest motif identity, and the gene encoding one of them had the highest genic expression.

### 3.9.3 Secondary metabolite gene clusters

A total of 47 SM clusters were predicted in the *C. lindemuthianum* genome (Figure 7). Most SM clusters were Polyketide synthetases (PKSs) and Nonribosomal Peptide Synthetases (NRPSs). SM clusters for terpene and indole production were also identified. Out of the 10 putatively identified NRPSs, two related to chrysogine and dimethylcoprogene production displayed 100% similarity with described clusters, whereas those related to aspirochlorine and

apicidin production showed less than 30% (Supplementary Table S5). Among the 18 T1-PKSs, only the cluster related to alternapyrone production demonstrated 100% similarity, while those related to naphthalene, cercosporin, and monacolin K production exhibited less than 35%. Among the 5 NRPS-Like, only the cluster related to EQ-4 recorded 100% similarity, while the cluster related to biotin showed 66%. In the case of the 6 terpenes, only a cluster related to squalastatin S1 production was found, with 40% similarity. No known biosynthetic clusters were found for the fungal-RiPP, hybrid, and indole production.

The number of SM clusters in the genomes of the *Colletotrichum* species ranged from 43 for *Colletotrichum orchidophilum* to 75 for *C. gloeosporioides*. Among the species belonging to the orbiculare complex (*C. lindemuthianum*, *C. orbiculare*, *C. sidae*, *C. spinosum* and *C. trifolii*), *C. lindemuthianum* showed the fewest number of SM clusters. The fungi from the *Colletotrichum* genus presented a greater number of SM clusters compared to all 30 microorganisms analyzed. The greatest number of SM clusters was identified in *C. gloeosporioides* (n= 75), and the lowest in *B. lactucae* (n= 0). The same type of SM clusters were

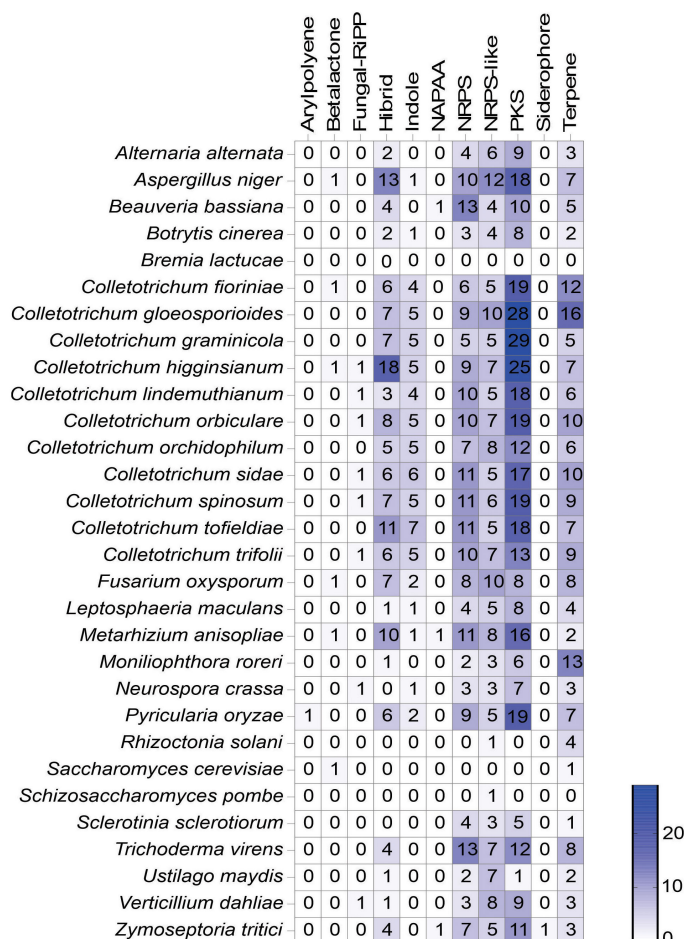


FIGURE 7

Distribution of secondary metabolite (SM) gene clusters predicted in different fungi using antiSMASH (fungal version). The numbers inside the boxes represent the number of secondary metabolite clusters found of each type, and the color intensity is related to the number. NRPS, Non-ribosomal peptide synthetases; PKS, Polyketide synthases; RiPP, Ribosomally synthesized and post-translationally modified peptide; NAPAA, Non-alpha poly-amino group acids.

observed in the genome of most fungi analyzed, mainly PKS, NRPS, indole and terpene clusters. Most fungi presented some hybrid SM clusters (Supplementary Figure S2). *Colletotrichum lindemuthianum* had the fewest hybrid SM clusters among the fungi from *Colletotrichum* species, and the *C. higginsianum* had the greatest number among all fungi analyzed, 3 and 18 respectively. The most common hybrid SM clusters were a combination of NRPS and PKS type 1.

### 3.9.4 Transposable elements and repeat-induced point mutations

The diversity and abundance of repetitive elements in the *C. lindemuthianum* genome was investigated. Comparisons with metrics obtained from other fungal genomes that had been sequenced with a long-read sequencing strategy similar to that used in the present study were made (Figure 8). Our analysis revealed that 60% of the *C. lindemuthianum* genome comprises repetitive elements, a significantly higher proportion compared to other fungi (Figure 8A). Specifically, this exceeds the repetitive element content of *A. alternata* by over 30 times and that of its close

relative, *C. higginsianum*, by nine times. The comparative genomic analysis of seven *Colletotrichum* species revealed a significant variation in total transposable element (TE) content, ranging from 4.3% to 44.8% of the genome in *Colletotrichum scovillei* to *C. orbiculare*, respectively (Rao et al., 2018). Hence, the observed differences in TE numbers may shed light on why *C. lindemuthianum* possesses one of the largest genomes within its genus.

Upon closer examination of the classes of repetitive elements present, we found that a significant portion of *C. lindemuthianum*'s genome comprises retroelements, particularly the *Ty1/Copia* superfamily, which accounts for 22% of the genome and represents 40% of the repetitive elements. Additionally, the *Gypsy/DIRS1* superfamily is notable by its presence. Furthermore, DNA transposons, specifically *Tc1-IS630-Pogo*, constitute approximately 1% of the genome, with 9% of these elements remaining unclassified (Figure 8B).

*Colletotrichum lindemuthianum* stands out for having a remarkably high proportion of its genome affected by RIP (54.77%), the highest among all genomes evaluated, indicating

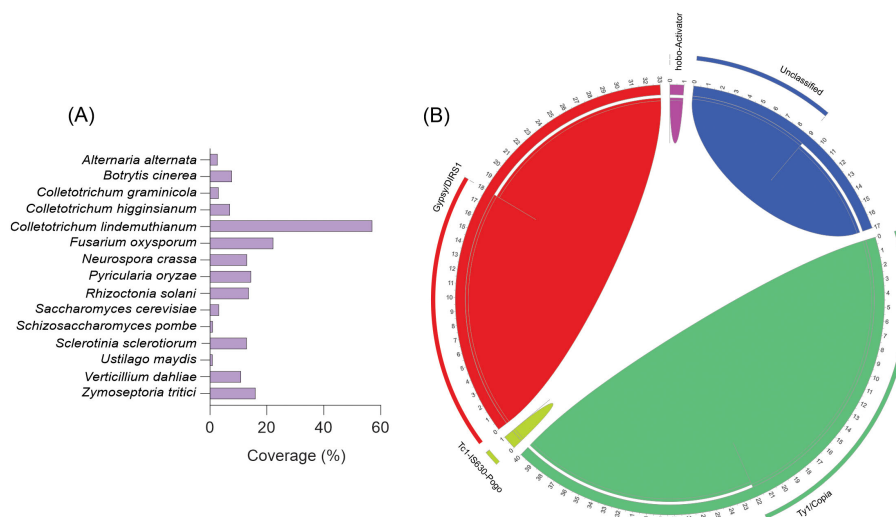


FIGURE 8

Repetitive elements among different fungal species. (A) The comparative analysis of repetitive elements in different fungi. (B) Repetitive elements in *Colletotrichum lindemuthianum*. The chord diagram illustrates the distribution and relationships of various transposable element families. The outer chord around the circle represents the percentage of each element family. The inner chords depict the proportion of these elements compared to the total elements found in the genome.

substantial RIP activity within this species (Table 2). Although the other members of the orbiculare complex analyzed (*C. orbiculare*, *C. sidae*, *C. spinosum*, *C. trifolii*) also have large genomes, they were not as affected by RIP as *C. lindemuthianum*. Apparently, RIP contributes significantly to the low GC content in the *C. lindemuthianum* genome. The RIP index values were similar to those observed in other species of the genus *Colletotrichum* and even *Neurospora crassa* (Supplementary Table S6).

The homologous sequence of MAT1-2-1 from *C. lindemuthianum* isolate MU 03 (García-Serrano et al., 2008) was identified in the genome of *C. lindemuthianum* A<sub>2</sub> 2-3, with over 99% coverage and identity, but no homologous sequences were found for the MAT1-1 sequences from *F. graminearum*, as has already been observed in other genomes of *Colletotrichum* spp. analyzed. Although RIP is associated with the sexual cycle, this cycle is not frequently observed in *C. lindemuthianum* and is not subject to the control of a traditional mating-type system, since both idiomorphs were not found in all isolates analyzed to date.

## 4 Discussion

The genome of *C. lindemuthianum* isolate A<sub>2</sub> 2-3 physiological race 89 presented in this study is one of the largest genomes reported for the *Colletotrichum* genus. Considering the genome complexity of this fungus due to its high number of repetitive elements and larger size than most fungi of the same genus, it was possible to obtain a high-quality 100 Mpb genome containing 119 scaffolds.

In eukaryotic organisms with small genomes, the genome size and gene number are commonly associated (Raffaele and Kamoun, 2012). The *C. lindemuthianum* genome presented here increased the number of predicted genes compared to the previous assembly

(Queiroz et al., 2017). However, there is no correlation between the genome size and gene number for fungal genomes of the *Colletotrichum* genus. In addition, the *C. lindemuthianum* genome comprises more than 60% of A+T-rich regions. This low G+C content (%) and gene density are indicators that *C. lindemuthianum* genome size expansion may have followed the “mutation-hazard” model (Kelkar and Ochman, 2012), where genome expansion would be driven by genetic drift, allowing mobile element accumulation, which could contribute to the existence of genomic regions with low G+C content (%).

Carbohydrate-active enzymes (CAZymes) have central roles for degradation of polysaccharides and oligosaccharides in plants, and there were many CAZymes in *C. lindemuthianum*, with more than 40% of the CAZymes predicted to be secreted. Carbohydrate degradation is one of the main strategies of phytopathogenic fungi to enable them to establish themselves in host plants and cause disease (Schmitz et al., 2019). In *Colletotrichum*, CAZymes are mainly xylanases, pectinases, and cellulases (Gan et al., 2012). The enzymes involved in *C. lindemuthianum* plant cell wall degradation include a higher proportion of pectinases compared to cellulases and xylanases, with many of these enzymes directly contributing to the fungus’s pathogenicity in plants. In *C. lindemuthianum*, the expression of many genes encoding pectinases during the fungus-host interaction has already been shown (Silva et al., 2022), and the deletion of some of these genes may cause reduction in pathogen aggressiveness (Cnossen-Fassoni et al., 2013). The pectinases produced by *C. lindemuthianum* are necessary for the maceration of common bean tissue in the necrotrophic phase, but they also play a crucial role in a stage preceding the biotrophic phase, as has been demonstrated for the endopolygalacturonase PLPG2, which associates with the cell wall, possibly reducing severe cellular damage that could trigger defense responses (Herbert et al., 2004). Therefore, it is evident that

TABLE 2 RIP profile in different fungi.

Species	Genome Size	GC content (%)	RIP affected genomic (%)
<i>Alternaria alternata</i>	33 Mb	51,4	0,31
<i>Aspergillus niger</i>	34 Mb	50,35	0,07
<i>Beauveria bassiana</i>	34 Mb	51,51	2,88
<i>Botrytis cinerea</i>	43 Mb	42	0,49
<i>Colletotrichum fioriniae</i>	50 Mb	51,91	2,79
<i>Colletotrichum gloeosporioides</i>	62 Mb	50,83	7,75
<i>Colletotrichum graminicola</i>	52 Mb	49,11	12,69
<i>Colletotrichum higginsianum</i>	51 Mb	54,41	3,34
<i>Colletotrichum lindemuthianum</i>	100 Mb	37,25	54,77
<i>Colletotrichum orbiculare</i>	90 Mb	37,57	22,62
<i>Colletotrichum orchidophilum</i>	49 Mb	51,07	5,26
<i>Colletotrichum sidae</i>	87 Mb	37,96	19,25
<i>Colletotrichum spinosum</i>	83 Mb	38,71	20,46
<i>Colletotrichum tofieldiae</i>	53 Mb	53,28	3,51
<i>Colletotrichum trifolii</i>	110 Mb	36,62	25,27
<i>Fusarium oxysporum</i>	48 Mb	47,63	2,61
<i>Leptosphaeria maculans</i>	45 Mb	45,19	11,84
<i>Metarhizium anisopliae</i>	43 Mb	50,9	1,37
<i>Moniliophthora roreri</i>	57 Mb	46,51	0,22
<i>Neurospora crassa</i>	41 Mb	48,23	15,26
<i>Pyricularia oryzae</i>	41 Mb	51,61	0,36
<i>Rhizoctonia solani</i>	41 Mb	47,63	0,52
<i>Saccharomyces cerevisiae</i>	13 Mb	36,05	0,07
<i>Schizosaccharomyces pombe</i>	12 Mb	38,15	0,16
<i>Sclerotinia sclerotiorum</i>	38 Mb	41,8	0,45
<i>Trichoderma virens</i>	39 Mb	49,25	1,16
<i>Ustilago maydis</i>	20 Mb	54,03	0,01
<i>Verticillium dahliae</i>	34 Mb	55,79	1,76
<i>Zyoseptoria tritici</i>	40 Mb	52,14	15,3

CAZymes can play a central role in fungal pathogenicity caused by *Colletotrichum* species.

Proteases are enzymes that degrade protein into amino acids and peptides, being implicated in the pathogenicity of many phytopathogenic fungi such as *Colletotrichum* species (Redman and Rodriguez, 2002). The predicted secretome of *C. lindemuthianum* contains 107 proteases, mainly serine and metalloproteases, such as observed for other *Colletotrichum* species (Rao and Nandineni, 2017). There are more proteases encoded by the genomes of *Colletotrichum* species than by other fungal pathogens (Gan et al., 2012), and a study with *Colletotrichum coccodes* demonstrated a relationship between proteases and plant mortality and symptoms induction, suggesting that these enzymes play a fundamental role in pathogenicity and virulence (Redman and Rodriguez, 2002). Although some studies have investigated *C. lindemuthianum* proteases, little is known about the role of these enzymes in this fungus's pathogenicity (Wijesundera et al., 1989). However, the data revealed about these enzymes after sequencing may serve as a guide for future studies. The characterization of both CAZymes and proteases of *C. lindemuthianum*, presented here, may help to elucidate the mechanisms by which the fungus invades its host and establishes pathogenesis.

In the *C. lindemuthianum* genome, 157 genes encoding for cytochrome P450s (CYP450s) were found and distributed in 74 families, with CYP65 and CYP68 being the families with the most putative proteins. CYP450s are oxidative enzymes widely distributed in living organisms, including fungi (Moktali et al., 2012). In phytopathogenic fungi, CYP450s play critical roles in the detoxification of toxic compounds produced by the host and in secondary metabolite synthesis, activities which are essential for pathogenicity and virulence (Črešnar and Petrič, 2011). In the fungus *Colletotrichum siamense*, a cytochrome P450-encoding gene was proven to be related to fungal sensitivity for the toxic compound fludioxonil by genetic manipulation (Guan et al., 2022). Many CYP450s genes in the *C. lindemuthianum* genome are related to secondary metabolite production (Moktali et al., 2012). PHI-base is a manually curated database that gathers information on validated genes related to the pathogenicity of fungi and other pathogens (Urban et al., 2015). Putative proteins from *C. lindemuthianum* were similar to proteins in PHI-base that affected pathogenicity, that when mutated resulted in virulence reduction, pathogenicity loss, and hypervirulence, and had effector functionality. The information from the PHI-base analysis could be used to determine candidates for further research and could help identify possible targets for developing more specific and effective fungicides.

In this study, the new genome assembly of *C. lindemuthianum* revealed 525 effector candidates, a slight increase from the previous count of 349 (Queiroz et al., 2019). Alignment comparison showed 262 candidate effector proteins from the prior study were similar to the new predictions, with 183 sharing full similarity. The effector candidates were predominantly putatively secreted, including cytoplasmic, apoplasmic, and dual-localized effectors. Most candidates lacked similar proteins in the PHI-base database, and various subcellular localizations were predicted, such as the nucleus, chloroplast, and mitochondria. Overall, this study indicates a

refinement in the identification of effector candidates in *C. lindemuthianum* and provides insights into their potential roles, secretion, and subcellular localization. The increase in effector candidate counts and the findings about their predicted localization and interactions with host organisms could have implications for understanding the pathogenic mechanisms and developing strategies for disease management.

In *C. lindemuthianum*, 18 sequences encoding proteins are similar to chitinases and six present mutations in conserved regions. Chitinase-like proteins differ from classical chitinases because they do not hydrolyze chitin due to mutations in some conserved amino acids (Fiorin et al., 2018). When analyzing the interaction of another *C. lindemuthianum* isolate, transcripts of chitin-like protein genes were observed at 24, 48, and 72 hours post-inoculation, corresponding to the biotrophic phase of the pathogen (Romero et al., 2024). These *C. lindemuthianum* putative proteins may also have lost their chitinolytic activity and remain capable of only binding to chitin, thus acquiring a new function as effector proteins, preventing the elicitation of plant defense responses. A similar situation was observed in *Monilophthora* spp., where chitinase-like proteins presented mutations in conserved sites and lost their chitinolytic activity but the encoding genes were highly expressed during the biotrophic phase (Fiorin et al., 2018). It was suggested that these proteins function as effector proteins, sequestering free fragments of chitin from the pathogen and preventing recognition by the host. These proteins play an interesting role in the pathogenicity of these organisms, by manipulating the host plant's immune system.

There were a greater number of transporters in the genome of *C. lindemuthianum* compared to the other fungi analyzed. The main transporters are members of the MFS and ABC superfamilies. The transition to necrotrophy in pathogenic fungi leads to a significant shift in the expression of genes encoding hydrolases, proteases, and transporters, which collectively facilitate the degradation of plant tissues and the assimilation of intracellular nutrients (Gan et al., 2012; O'Connell et al., 2012; Alkan et al., 2015; Zhang et al., 2018). Nutrients liberated from the infected plant tissue are subsequently absorbed through transmembrane transporters, including oligopeptide transporters, ABC transporters, and members of the MFS. Some of these transporters have been implicated in the pathogenicity of *Colletotrichum* spp. with documented roles in sugar transport, oxidative stress resistance, and pathogenesis (Chagué et al., 2009; Shnaiderman et al., 2013; Liu et al., 2017). Similarly, comparative genome analyses across the *Colletotrichum* genus unveiled tailored transporter profiles, often associated with the MFS type transporters, encompassing myo-inositol transporters, general glucose transporters, and multiple members of the anion:cation symporter (ACS) family (Gan et al., 2016). In addition, within *C. lindemuthianum*, our study revealed a prominent representation of the MFS, ABC, and MFP superfamilies, collectively constituting a substantial portion of the fungal transporter repertoire. The results obtained with the isolation of mutants ( $\Delta abcCl1$ ) demonstrate that the AbcCl1 transporter protects *C. lindemuthianum* from toxic compounds produced by

the common bean, probably in the biotrophic phase of the fungus (Oliveira et al., 2022). In addition, transporters should also be important for the export of toxins during the necrotrophic phase. Notably, the deletion of specific genes encoding these transporters has also been associated with the role of hexose transporters in the capture of carbon sources during the necrotrophic phase of the plant-pathogen interaction (Pereira et al., 2013).

NLPs are conserved proteins presenting a seven amino acid motif (GHRHDWE), they induce ethylene production and necrosis in leaves of dicot plant species (Oome and Van Den Ackerveken, 2014). In this study, the majority of microorganism genomes analyzed contained at least one NLP. It described three types of NLP in different eukaryotic and prokaryotic microorganisms, pathogenic or not for plants (Seidl and Van den Ackerveken, 2019). *Colletotrichum lindemuthianum* has a similar number of NLPs to most other *Colletotrichum* species, and these NLPs seem to be all from type 1. Type 1 NLP is mainly found in microorganisms that interact with plants (Oome and Van Den Ackerveken, 2014). In previous studies genomes from the *Colletotrichum acutatum* species complex (*Colletotrichum nymphaeae*, *Colletotrichum simmondsii*, *Colletotrichum fiorinae*, and *Colletotrichum salicis*) have a greater number of NLPs, almost double that of the others *Colletotrichum* species (Baroncelli et al., 2016), and some *Colletotrichum* species may have different types of NLP (Liu et al., 2020). In the obligate biotrophic oomycete *Plasmopara viticola* genome eight NLPs were identified, and their encoding genes were expressed at different stages of the pathogen-host interaction, but none of them were able to cause necrosis in *Nicotiana benthamiana* when transiently expressed (Schumacher et al., 2020; Askani et al., 2021).

In *C. lindemuthianum*, the seven putative NLP identified had variations in the heptamer consensus motif and encoding genes had different expression profiles. Previously mutant NLP proteins with amino acid alterations in the conserved consensus motif "GHRHDWE" had their activity reduced or lost (Ottmann et al., 2009). In *C. higginsianum*, similar results were observed (Kleemann et al., 2012). The ChNLP1 NLP was expressed at the switch to necrotrophy in the interaction of *C. higginsianum* with *N. benthamiana* and ChNLP3 was expressed in appressoria before penetration. When these proteins are transiently expressed in *N. benthamiana*, ChNLP1 was able to induce cell death, whereas ChNLP3 was not. The ChNLP3 had a mutation at critical amino acids that resulted in loss of function whereas, ChNLP1 had not. We observed that the *C. lindemuthianum* Clind89-006653-T1 putative protein is very similar (92% of identity) to the *C. orbiculare* CoNLP1 protein, and both had the same amino acid substitution in the motif GHRHDWE. Despite the mutation, the *C. orbiculare* CoNLP1 protein can induce necrosis in *N. benthamiana* when transiently expressed via agroinfiltration (Azmi et al., 2018). In *C. orbiculare* an amino acid substitution in the critical motif of NLP1H127A reduced host specificity resulting in ability to cause necrosis only in melon, while the protein with a complete motif was capable of causing necrosis in melon and *N. benthamiana* (Chen et al., 2021). Based on sequence analyses of the conserved NLP motif, gene expression, and phylogenetic reconstruction, we hypothesize that only a few *C. lindemuthianum* NLPs could induce necrosis in the

host, with Clind89-006653-T1 being prime candidate. This can be confirmed by isolation and analysis of mutants in future research.

Control of NLP gene expression is critical since recognition of NLPs may trigger host defenses (Azmi et al., 2018). An isolate of *C. orbiculare* with constitutive NLP expression had a reduced host range when measured against five typical hosts. In contrast, pathogen inoculation in *N. benthamiana* was able to cause disease. NLPs may be expressed during different phases of the pathogen-host interaction, but the gene is commonly expressed in the necrotrophic phase, where the NLP protein is considered a virulence factor (Seidl and Van den Ackerveken, 2019). Because of the importance of NLPs for several phytopathogenic microorganisms in the interactions with their hosts, the NLPs may be a target for developing novel agents and strategies for plant pathogen control (Pirc et al., 2021).

The genome of *C. lindemuthianum* contains a diverse range of secondary metabolite (SM) clusters, including NRPS, PKS, and those responsible for synthesis of indoles and terpenes. The number of SM clusters in *C. lindemuthianum* is fewer than in the majority of other *Colletotrichum* species, but greater than in most other analyzed genera. Ascomycetes exhibit in general a higher number of SM clusters compared to other fungal taxa (Collemare et al., 2008). The variable number of SM clusters found in different species, both within the *Colletotrichum* genus and across analyzed species, may be influenced by the fungi's lifestyle or their interaction with the host. Also, this variation indicates a vast diversity of produced compounds, presuming a specific metabolite for each predicted cluster (Collemare et al., 2008; O'Connell et al., 2012). Fungal ecology may drive adaptations that result in the gain or loss of specific metabolite clusters, influenced by various environmental pressures (Wisecaver and Rokas, 2015). In phytopathogenic fungi, these metabolites are of great importance in plant interactions, facilitating host infection and colonization, and directly influencing disease development and fungal survival (Pusztahelyi et al., 2015; Moraga et al., 2019). Many of these metabolites are phytotoxic, such as the colletotrichins found in *Colletotrichum* spp., capable of inducing symptoms similar to the disease itself and damaging the host's plasma membrane (Gohbara et al., 1978; Duke et al., 1992).

From the 47 SM clusters found in *C. lindemuthianum*, only a few were similar with known biosynthetic SM clusters. Among their products, chrysogine (NRP) is recognized as a yellow pigment produced by various fungi, including *Penicillium chrysogenum*. Although it does not exhibit antimicrobial properties, there are indications that it may play a role in the protection against ultraviolet (UV) radiation (Viggiano et al., 2018). Dimethylcopropane (NRP) is classified as a siderophore, found in various fungi, and may be involved in iron uptake in *C. lindemuthianum*, as described in other phytopathogens such as *C. graminicola*, *A. alternata*, *Cochliobolus heterostrophus*, and *F. gramineum* (Lee et al., 2005; Oide et al., 2006; Chen et al., 2013; Albarouki et al., 2014). Studies also suggest that siderophores play a relevant role in virulence and protection against oxidative stress responses (Lee et al., 2005; Oide et al., 2006; Chen et al., 2013; Albarouki et al., 2014).

In *C. lindemuthianum*, the alternapyrone biosynthesis cluster was identified, which is also present in *Parastagonospora nodorum*, a wheat pathogen. During the interaction between *P. nodorum* and wheat, an increase in expression of the genes in the cluster was observed during plant development, suggesting positive regulation. Beyond that, when alternapyrones were applied in wheat seed germination assays a phytotoxic effect was observed (Ipcho et al., 2012; Li et al., 2018). *Colletotrichum lindemuthianum* possesses the microperforanone biosynthesis cluster (EQ-4, NRP) with 100% similarity to that in the *A. nidulans* genome (MIBig accession: BGC0001668). This molecule has antimicrobial activity against various microorganisms, such as *Edwardsiella ictarda*, *C. gloeosporioides*, *Vibrio harveyi*, and others (Lü et al., 2020).

As noted, the majority of SM clusters in *C. lindemuthianum* show no similarity to previously described SM clusters. This indicates that the functions of several products derived from these SM clusters remain unknown, underscoring the need for further investigations. These unidentified products may contribute to the pathogenicity of *C. lindemuthianum*, as demonstrated in *C. gloeosporioides*, where secondary metabolites extracted from different isolates exhibited varying phytotoxic activities in yam (*Dioscorea* spp.) (Abang et al., 2009).

Different metabolites with varied functions are produced by *Colletotrichum* spp. during host plant infection (Moraga et al., 2019), and most of these metabolites are unknown. Considering that these metabolites may have functions such as helping to combat competitors (Arivudainambi et al., 2011) or contributing to plant tissue necrosis (Reveglia et al., 2023), the phase of the life cycle in which they are produced is related to the environment where the pathogen is found. Determining the ideal conditions for the expression of a given SM cluster remains a challenge, as is evaluating the metabolites produced during infection.

The new genome assembly of *C. lindemuthianum* comprises a high percentage of repetitive sequences. This represents the highest percentage of repetitive sequences reported for fungi within the *Colletotrichum* genus, surpassing figures for other species like *Colletotrichum lentis* (4.56%), *Colletotrichum truncatum* (6.8%), and *C. orbiculare* (8.3%) (Rao et al., 2018; Bhadauria et al., 2019). Among the repetitive sequences identified in the genome of *C. lindemuthianum*, Class I transposable elements (TEs) are the most abundant. Elements of the LTR order, belonging to the *Gypsy* and *Copia* superfamilies, are the most represented among these. These results corroborate findings in other species such as *C. truncatum*, *C. higginsianum*, *C. graminicola*, *C. scovillei*, *Colletotrichum chlorophyti*, and *C. orchidophilum* (Rao et al., 2018). Despite their phylogenetic proximity and having several similar genomic characteristics, the species *C. lindemuthianum* and *C. orbiculare* seem to exhibit differences in the distribution of their TEs. While the *Ty1/Copia* superfamily is one of the most abundant in the genomes of *C. lindemuthianum*, Rao et al. (2018) reported that elements from the *Copia* superfamily are the most abundant in *C. orbiculare*, followed by elements from the *Gypsy* superfamily. Furthermore, it was observed that the largest fraction of TEs in the *C. orbiculare* genome consists of unknown elements, which

contrasts with the data observed in the present work for the species *C. lindemuthianum*, where it was possible to classify most of the identified elements.

The significant proportion of the genome of *C. lindemuthianum* consisting of repetitive elements, coupled with its status as the largest genome within the genus (Queiroz et al., 2017), presents an intriguing context for interpreting its observed RIP activity. To prevent potential damage by the movement of TEs into housekeeping genes, mechanisms such as DNA methylation (Goll and Bestor, 2005; Zemach et al., 2010), repeat-induced point mutation (RIP) (Cambareri et al., 1989), and RNA interference (RNAi) (Fulci and Macino, 2007) help to suppress excessive TE activity. The extensive presence of repetitive elements provides ample substrate for the RIP machinery to act upon, potentially driving the high RIP-affected genomic proportion observed in this species. *Colletotrichum lindemuthianum* has the characteristics described by van Wyk et al. (2021) for fungi that present extensive RIP: It performs sexual reproduction; has a genome rich in repetitive elements; has genes associated with RIP, DNA methylation and RIP direct heterochromatic silencing.

The sexual cycle in *C. lindemuthianum* is described as rare. This rarity would, in theory, limit the impact of the RIP silencing mechanism on its genome. However, this theoretical limitation contrasts with the significant RIP signatures presented by the *C. lindemuthianum* genome, which are the highest among the fungi evaluated. Gladyshev and Kleckner (2017) proposed a potential mechanism for generating RIP mutations that is independent of RID (C5-methyltransferase essential for the occurrence of one of the RIP pathways) activity, and thus not reliant on the occurrence of the sexual cycle in fungi. This mechanism, orchestrated by the DIM-2 protein in collaboration with the DIM-5 (histone methyltransferase) and HP1 (heterochromatin 1 protein) proteins, could be the primary factor responsible for generating RIP mutations in the *C. lindemuthianum* genome. This may explain the robust RIP signatures observed in its genome even in the absence of a frequent sexual cycle. Given that *C. lindemuthianum* exhibits the most pronounced RIP signatures among *Colletotrichum* fungi and has a rare sexual cycle, it emerges as an intriguing model for investigating the RIP generation mechanism proposed by Gladyshev and Kleckner (2017). Further investigations into the specific mechanisms governing RIP activity in this species could provide valuable insights into the evolutionary processes driving genomic diversity and adaptation in fungi.

One crucial step in comprehending the genetics of plant-pathogen interactions involves gaining access to the genomes of both entities. In this study, we present a high-quality genome of a significant pathogen in common bean cultivation, specifically, an intriguing *Colletotrichum* species. Interestingly, the expansion in the size of *C. lindemuthianum* genome did not result from an increase in the number of genes, nor in terms of families of genes that encode known virulence factors but was mainly associated with an increase in the content of repetitive elements. The data supplied in this study will bolster future investigations into this pathosystem, enabling a more in-depth understanding of the interaction between

the fungus and its host. This insight will facilitate the identification of important genes within the pathogen, serving as potential targets for novel strategies aimed at disease control.

## Data availability statement

The original contributions presented in the study are publicly available. This data can be found at the National Center for Biotechnology Information (NCBI) using accession number BioProject: PRJNA325493.

## Author contributions

LS: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. HC: Conceptualization, Methodology, Writing – original draft. OG: Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. PV: Conceptualization, Data curation, Formal analysis, Methodology, Software, Supervision, Writing – original draft. RR: Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. MS: Conceptualization, Data curation, Validation, Writing – review & editing. MQ: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. We are grateful to support by the following Brazilian agencies: the Minas Gerais Science Foundation (FAPEMIG– Fundação de Amparo à Pesquisa do Estado de Minas Gerais), the Brazilian Federal Agency of Support and Evaluation of Postgraduate Education – Finance code 001 (CAPES–Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Código de Financiamento 001) and the National Council of Scientific and Technological Development (CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico).

## 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.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/ffunb.2024.1459229/full#supplementary-material>

### SUPPLEMENTARY FIGURE 1

Analysis of the repertoires of *Colletotrichum lindemuthianum* effector candidates (CIECs). (A) Comparison of the number CIECs predicted by Queiroz et al. (2019) (pink circle) and the current assembly genome (blue circle). The intersection value between the two circles represents the number of proteins sharing 100% similarity. (B) Distribution of *Colletotrichum lindemuthianum* effector candidates across different categories. The bar chart shows the total number of identified effector candidates and their distribution across various functional and subcellular categories, including secreted (secreted), apoplastic (apoplastic), cytoplasmic (cytoplasmic), nuclear (localizer\_nucleus), chloroplastic (localizer\_chloro), and mitochondrial (localizer\_mito) effectors. "Phibase\_genes" represents candidates with genes linked to the PHI-base database, while "effector\_matches" indicates candidates matching known effectors.

### SUPPLEMENTARY FIGURE 2

Distribution of hybrid secondary metabolite (SM) gene clusters predicted in different fungi using antiSMASH (fungal version). The numbers inside the boxes represent the number of hybrid secondary metabolite clusters found of each type, and the color intensity is related to the number. NRPS: Non-ribosomal peptide synthetases; T1PKS: Polyketide synthases type I; T3PKS: Polyketide synthases type III; R-F: fungal-RiPP (ribosomally synthesized and post-translationally modified peptide).

### SUPPLEMENTARY TABLE 1

Fungal species used for analysis.

### SUPPLEMENTARY TABLE 2

Alignment-Hit table of effector comparison between candidate effectors from Queiroz et al., (2019) and current candidate effectors.

### SUPPLEMENTARY TABLE 3

Details of effector candidates predicted in the *Colletotrichum lindemuthianum* first genome assembly (Queiroz et al., 2019).

### SUPPLEMENTARY TABLE 4

Details of effector candidates predicted in the *Colletotrichum lindemuthianum* new genome assembly.

### SUPPLEMENTARY TABLE 5

Secondary metabolites clusters in *Colletotrichum lindemuthianum*.

### SUPPLEMENTARY TABLE 6

RIP index value of different fungal genomes.

## References

- Abang, M. M., Abraham, W. R., Asiedu, R., Hoffmann, P., Wolf, G., and Winter, S. (2009). Secondary metabolite profile and phytotoxic activity of genetically distinct forms of *Colletotrichum gloeosporioides* from yam (*Dioscorea* spp.). *Mycol Res.* 113, 130–140. doi: 10.1016/j.mycres.2008.09.004
- Albarouki, E., Schafferer, L., Ye, F., von Wirén, N., Haas, H., and Deising, H. B. (2014). Biotrophy-specific downregulation of siderophore biosynthesis in *Colletotrichum graminicola* is required for modulation of immune responses of maize. *Mol. Microbiol.* 92, 338–355. doi: 10.1111/mmi.12561
- Alkan, N., Friedlander, G., Ment, D., Prusky, D., and Fluhr, R. (2015). Simultaneous transcriptome analysis of *Colletotrichum gloeosporioides* and tomato fruit pathosystem reveals novel fungal pathogenicity and fruit defense strategies. *New Phytol.* 205, 801–815. doi: 10.1111/nph.13087
- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Available online at: <https://qubeshub.org/resources/fastqc> (Accessed April 9, 2023).
- Arivudainambi, U. S. E., Anand, T. D., Shanmugaiyah, V., Karunakaran, C., and Rajendran, A. (2011). Novel bioactive metabolites producing endophytic fungus *Colletotrichum gloeosporioides* against multidrug-resistant *Staphylococcus aureus*. *FEMS Immunol. Med. Microbiol.* 61, 340–345. doi: 10.1111/j.1574-695X.2011.00780.x
- Askani, L., Schumacher, S., and Fuchs, R. (2021). Sequence and gene expression analysis of recently identified NLP from *Plasmopara viticola*. *Microorganisms* 9, 1453. doi: 10.3390/microorganisms9071453
- Azmi, N. S. A., Singkaravanit-Ogawa, S., Ikeda, K., Kitakura, S., Inoue, Y., Narusaka, Y., et al. (2018). Inappropriate expression of an NLP effector in *Colletotrichum orbiculare* impairs infection on cucurbitaceae cultivars via plant recognition of the C-terminal region. *Mol. Plant Microbe Interact.* 31, 101–111. doi: 10.1094/MPMI-04-17-0085-FI
- Barcelos, Q. L., Pinto, J. M. A., Vaillancourt, L. J., and Souza, E. A. (2014). Characterization of *Glomerella* strains recovered from anthracnose lesions on common bean plants in Brazil. *PLoS One* 9, e90910. doi: 10.1371/journal.pone.0090910
- 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, 1–17. doi: 10.1186/s12864-016-2917-6
- Bhadauria, V., MacLachlan, R., Pozniak, C., Cohen-Skalie, A., Li, L., Halliday, J., et al. (2019). Genetic map-guided genome assembly reveals a virulence-governing minichromosome in the lentil anthracnose pathogen *Colletotrichum lentis*. *New Phytol.* 221, 431–445. doi: 10.1111/nph.15369
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Cai, L., Giraud, T., Zhang, N., Begerow, D., Cai, G., and Shivas, R. G. (2011). The evolution of species concepts and species recognition criteria in plant pathogenic fungi. *Fungal Divers.* 50, 121–133. doi: 10.1007/s13225-011-0127-8
- Cambareri, E. B., Jensen, B. C., Schabtach, E., and Selker, E. U. (1989). Repeat-induced G-C to A-T Mutations in *Neurospora*. *Science* 244, 1571–1575. doi: 10.1126/science.2544994
- Chagué, V., Maor, R., and Sharon, A. (2009). CgOpt1, a putative oligopeptide transporter from *Colletotrichum gloeosporioides* that is involved in responses to auxin and pathogenicity. *BMC Microbiol.* 9, 1–12. doi: 10.1186/1471-2180-9-173
- Chen, J., Inoue, Y., Kumakura, N., Mise, K., Shirasu, K., and Takano, Y. (2021). Comparative transient expression analyses on two conserved effectors of *Colletotrichum orbiculare* reveal their distinct cell death-inducing activities between *Nicotiana benthamiana* and melon. *Mol. Plant Pathol.* 22, 1006–1013. doi: 10.1111/mpp.13078
- Chen, L. H., Lin, C. H., and Chung, K. R. (2013). A nonribosomal peptide synthetase mediates siderophore production and virulence in the citrus fungal pathogen *Alternaria alternata*. *Mol. Plant Pathol.* 14, 497–505. doi: 10.1111/mpp.12021
- Cnossen-Fassoni, A., Bazzolli, D. M. S., Brommonschenkel, S. H., Fernandes de Araújo, E., and de Queiroz, M. V. (2013). The pectate lyase encoded by the *pecC1* gene is an important determinant for the aggressiveness of *Colletotrichum lindemuthianum*. *J. Microbiol.* 51, 461–470. doi: 10.1007/s12275-013-3078-9
- Collemare, J., Billard, A., Böhnert, H. U., and Lebrun, M. H. (2008). Biosynthesis of secondary metabolites in the rice blast fungus *Magnaporthe grisea*: the role of hybrid PKS-NRPS in pathogenicity. *Mycol Res.* 112, 207–215. doi: 10.1016/j.mycres.2007.08.003
- Črešnar, B., and Petrič, Š. (2011). Cytochrome P450 enzymes in the fungal kingdom. *Biochim. Biophys. Acta Proteins Proteom* 1814, 29–35. doi: 10.1016/j.jbbpap.2010.06.020
- Dean, R., Van Kan, J. A. L., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., et al. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* 13, 414–430. doi: 10.1111/j.1364-3703.2011.00783.x
- Duke, S. O., Gohbara, M., Paul, R. N., Duke, M. V., Duke, S. O., and Paul, R. N. (1992). Colletotrichin causes rapid membrane damage to plant cells. *J. Phytopathol.* 134, 289–305. doi: 10.1111/j.1439-0434.1992.tb01237.x



- Eisenhaber, B., Schneider, G., Wildpaner, M., and Eisenhaber, F. (2004). A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *J. Mol. Biol.* 337, 243–253. doi: 10.1016/j.jmb.2004.01.025
- Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2, 953–971. doi: 10.1038/nprot.2007.131
- Emanuelsson, O., Nielsen, H., Brunak, S., and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903
- English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., et al. (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7, e47768. doi: 10.1371/journal.pone.0047768
- FAOSTAT (2022). *Production quantities by country/Dry beans*. Available online at: <https://www.fao.org/faostat/en/#data/QCL/visualize> (Accessed January 17, 2023).
- Fiorini, G. L., Sánchez-Vallet, A., Thomazella, D. P., de T., do Prado, P. F. V., do Nascimento, L. C., et al. (2018). Suppression of plant immunity by fungal chitinase-like effectors. *Curr. Biol.* 28, 3023–3030.e5. doi: 10.1016/j.cub.2018.07.055
- 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.* 117, 9451–9457. doi: 10.1073/pnas.1921046117
- Fontenelle, M. R., Santana, M. F., Cnossen, A., Bazzoli, D. M. S., Bromonschenkel, S. H., de Araújo, E. F., et al. (2017). Differential expression of genes during the interaction between *Colletotrichum lindemuthianum* and *Phaseolus vulgaris*. *Eur. J. Plant Pathol.* 147, 653–670. doi: 10.1007/s10658-016-1033-4
- Forrest, L. R., Krämer, R., and Ziegler, C. (2011). The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta Bioenerg* 1807, 167–188. doi: 10.1016/j.bbabi.2010.10.014
- Fulci, V., and Macino, G. (2007). Quelling: post-transcriptional gene silencing guided by small RNAs in *Neurospora crassa*. *Curr. Opin. Microbiol.* 10, 199–203. doi: 10.1016/j.mib.2007.03.016
- Gan, P., Ikeda, K., Irieda, H., Narusaka, M., Connell, R. J. O., Narusaka, Y., et al. (2012). Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol.* 197, 1236–1249. doi: 10.1111/nph.12085
- Gan, P., Narusaka, M., Kumakura, N., Tsuchida, 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
- Gan, P., Tsuchida, A., Narusaka, M., Narusaka, Y., Takano, Y., Kubo, Y., et al. (2019). Genome sequence resources for four phytopathogenic fungi from the *colletotrichum orbiculare* species complex. *Mol. Plant Microbe Interact.* 32, 1088–1090. doi: 10.1094/mpmi-12-18-0352-a
- García-Serrano, M., Laguna, E. A., Simpson, J., and Rodríguez-Guerra, R. (2008). Analysis of the MAT1-2-1 gene of *Colletotrichum lindemuthianum*. *Mycoscience* 49, 312–317. doi: 10.1007/S10267-008-0424-6
- Gladyshev, E., and Kleckner, N. (2017). DNA sequence homology induces cytosine-to-thymine mutation by a heterochromatin-related pathway in *Neurospora*. *Nat. Genet.* 49, 887–894. doi: 10.1038/ng.3857
- Gohbara, M., Kosuge, Y. O., Yamasak, S., Kimura, Y., Suzuki, A., and Tamur, S. (1978). Isolation, structures and biological activities of colletotrichins, phytotoxic substances from *Colletotrichum nicotianae*. *Agric. Biol. Chem.* 42, 1037–1043. doi: 10.1080/00021369.1978.10863104
- Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721
- Guan, X., Song, M., Lu, J., Yang, H., Li, X., Liu, W., et al. (2022). The transcription factor CsAtf1 negatively regulates the cytochrome P450 Gene CsCyp51G1 to increase fludioxonil sensitivity in *Colletotrichum siamense*. *J. Fungi* 8, 1032. doi: 10.3390/jof8101032
- Herbert, C., O'Connell, R., Gaulin, E., Salesses, V., Esquerré-Tugayé, M. T., and Dumas, B. (2004). Production of a cell wall-associated endopolygalacturonase by *Colletotrichum lindemuthianum* and pectin degradation during bean infection. *Fungal Genet. Biol.* 41, 140–147. doi: 10.1016/j.fgb.2003.09.008
- Hoi, J. W. S., Herbert, C., Bacha, N., O'Connell, R., Lafitte, C., Borderies, G., et al. (2007). Regulation and role of a STE12-like transcription factor from the plant pathogen *Colletotrichum lindemuthianum*. *Mol. Microbiol.* 64, 68–82. doi: 10.1111/j.1365-2958.2007.05639.x
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259
- Ipcho, S. V. S., Hane, J. K., Antoni, E. A., Ahren, D., Henrissat, B., Friesen, T. L., et al. (2012). Transcriptome analysis of *Stagonospora nodorum*: gene models, effectors, metabolism and pantothenate dispensability. *Mol. Plant Pathol.* 13, 531–545. doi: 10.1111/j.1364-3703.2011.00770.x
- Ishikawa, F. H., Souza, E. A., Read, N. D., and Roca, M. G. (2010). Live-cell imaging of conidial fusion in the bean pathogen *Colletotrichum lindemuthianum*. *Fungal Biol.* 114, 2–9. doi: 10.1016/j.funbio.2009.11.006
- Ishikawa, F. H., Souza, E. A., Read, N. D., and Roca, G. M. (2013). *Colletotrichum lindemuthianum* exhibits different patterns of nuclear division at different stages in its vegetative life cycle. *Mycologia* 105, 795–801. doi: 10.3852/12-298
- Jones, D. A. B., Rozano, L., Debler, J. W., Mancera, R. L., Moolhuijzen, P. M., and Hane, J. K. (2021). An automated and combinative method for the predictive ranking of candidate effector proteins of fungal plant pathogens. *Sci. Rep.* 11, 19731. doi: 10.1038/s41598-021-99363-0
- Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* 35, W429–W432. doi: 10.1093/nar/gkm256
- Kelkar, Y. D., and Ochman, H. (2012). Causes and consequences of genome expansion in fungi. *Genome Biol. Evol.* 4, 13–23. doi: 10.1093/gbe/evr124
- Kleemann, J., Rincon-Rivera, L. J., Takahara, H., Neumann, U., van Themaat, E. V. L., van der Does, H. C., et al. (2012). Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog.* 8, e1002643. doi: 10.1371/journal.ppat.1002643
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8
- Krueger, F., James, F., Ewels, P., Afyounian, E., Weinstein, M., Schuster-Boeckler, B., et al. (2015). doi: 10.5281/zenodo.7598955
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, 1–10. doi: 10.1186/gb-2009-10-3-r25
- Lee, B. N., Kroken, S., Chou, D. Y. T., Robbertse, B., Yoder, O. C., and Turgeon, B. G. (2005). Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, NPS6, involved in virulence and resistance to oxidative stress. *Eukaryot Cell* 4, 545–555. doi: 10.1128/EC.4.3.545-555.2005
- 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
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Hu, J., Wei, H., Solomon, P. S., Vuong, D., Lacey, E., et al. (2018). Chemical ecogenomics-guided discovery of phytotoxic  $\alpha$ -pyrones from the fungal wheat pathogen *parastagonospora nodorum*. *Org. Lett.* 20, 6148–6152. doi: 10.1021/acs.orglett.8b02617
- Liang, X., Wang, B., Dong, Q., Li, L., Rollins, J. A., Zhang, R., et al. (2018). Pathogenic adaptations of *Colletotrichum* fungi revealed by genome wide gene family evolutionary analyses. *PLoS One* 13, e0196303. doi: 10.1371/journal.pone.0196303
- 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. doi: 10.3389/fmicb.2020.01484
- Liu, L., Yan, Y., Huang, J., Hsiang, T., Wei, Y., Li, Y., et al. (2017). A novel MFS transporter gene ChMfs1 is important for hyphal morphology, conidiation, and pathogenicity in *Colletotrichum higginsianum*. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.01953
- Lü, F., Li, X., Chi, L., Meng, L., and Wang, B. (2020). A new acyclic peroxide from *Aspergillus nidulans* SD-531, a fungus obtained from deep-sea sediment of cold spring in the South China Sea. *J. Oceanol. Limnol.* 38, 1225–1232. doi: 10.1007/s00343-020-0052-3
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaSci* 1, 18. doi: 10.1186/2047-217X-1-18
- Mahuku, G. S., and Riascos, J. J. (2004). Virulence and molecular diversity within *Colletotrichum lindemuthianum* isolates from Andean and Mesoamerican bean varieties and regions. *Eur. J. Plant Pathol.* 110, 253–263. doi: 10.1023/B:EJPP.0000019795.18984.74
- Mattinen, L., Tshuikina, M., Mäe, A., and Pirhonen, M. (2004). Identification and Characterization of Nip, Necrosis-Inducing Virulence Protein of *Erwinia carotovora* subsp. *carotovora*. *Mol. Plant Microbe Interact.* 17, 1366–1375. doi: 10.1094/MPMI.2004.17.12.1366
- Medema, M. H., Blin, K., Cimermanic, P., De Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39, W339–W346. doi: 10.1093/nar/gkr466
- Moktali, V., Park, J., Fedorova-Abrams, N. D., Park, B., Choi, J., Lee, Y. H., et al. (2012). Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. *BMC Genomics* 13, 525. doi: 10.1186/1471-2164-13-525
- Moraga, J., Gomes, W., Pinedo, C., Cantoral, J. M., Hanson, J. R., Carbú, M., et al. (2019). The current status on secondary metabolites produced by plant pathogenic *Colletotrichum* species. *Phytochem. Rev.* 18, 215–239. doi: 10.1007/s11101-018-9590-0
- Nelson, D. R. (2009). The cytochrome P450 homepage. *Hum. Genomics* 4, 59. doi: 10.1186/1479-7364-4-1-59
- Nogueira, G. B., dos Santos, L. V., de Queiroz, C. B., Ribeiro Corrêa, T. L., Pedrozo Menicucci, R., Soares Bazzoli, D. M., et al. (2019). The histidine kinase slnCl1 of

*Colletotrichum lindemuthianum* as a pathogenicity factor against *Phaseolus vulgaris* L. *Microbiol. Res.* 219, 110–122. doi: 10.1016/j.micres.2018.10.005

O'Connell, R. J. (1987). Absence of aspecialized interface between intracellular hyphae of *Colletotrichum lindemuthianum* and cells of *Phaseolus vulgaris*. *New Phytol.* 107, 725–734. doi: 10.1111/j.1469-8137.1987.tb00910.x

O'Connell, R. J., Bailey, J. A., and Richmond, D. V. (1985). Cytology and physiology of infection of *Phaseolus vulgaris* by *Colletotrichum lindemuthianum*. *Physiol. Plant Pathol.* 27, 75–98. doi: 10.1016/0048-4059(85)90058-X

O'Connell, R. J., Pain, N. A., Hutchison, K. A., Jones, G. L., and Green, J. R. (1996). Ultrastructure and composition of the cell surfaces of infection structures formed by the fungal plant pathogen *Colletotrichum lindemuthianum*. *J. Microsc.* 181, 204–212. doi: 10.1046/j.1365-2818.1996.104384.x

O'Connell, R. J., and Ride, J. P. (1990). Chemical detection and ultrastructural localization of chitin in cell walls of *Colletotrichum lindemuthianum*. *Physiol. Mol. Plant Pathol.* 37, 39–53. doi: 10.1016/0885-5765(90)90084-B

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

Oide, S., Moeder, W., Krasnoff, S., Gibson, D., Haas, H., Yoshioka, K., et al. (2006). NPS6, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes. *Plant Cell* 18, 2836–2853. doi: 10.1105/tpc.106.045633

Oliveira, M. C., dos Santos, G. Q., Teixeira, J. A., Correia, H. L. N., da Silva, L. L., de Araújo, E. F., et al. (2022). The AbcCl1 transporter of *Colletotrichum lindemuthianum* acts as a virulence factor involved in fungal detoxification during common bean (*Phaseolus vulgaris*) infection. *Braz. J. Microbiol.* 53, 1121–1132. doi: 10.1007/s42770-022-00787-1

Oome, S., and Van Den Ackerveken, G. (2014). Comparative and functional analysis of the widely occurring family of Nep1-like proteins. *Mol. Plant-Microbe Interact.* 27, 1081–1094. doi: 10.1094/MPMI-04-14-0118-R

Ottmann, C., Luberacki, B., Küfner, I., Koch, W., Brunner, F., Weyand, M., et al. (2009). A common toxin fold mediates microbial attack and plant defense. *Proc. Natl. Acad. Sci.* 106, 10359–10364. doi: 10.1073/pnas.0902362106

Padder, B. A., Sharma, P. N., Awale, H. E., and Kelly, J. D. (2017). Offered review *Colletotrichum lindemuthianum*, the causal agent of bean anthracnose. *J. Plant Pathol.* 99, 317–330. doi: 10.4454/jpp.v99i2.3867

Palmer, J. M., and Stajich, J. (2020). Funannotate v1.8.1: Eukaryotic genome annotation. *Zenodo* 1. doi: 10.5281/zenodo.4054262

Parisot, D., Dufresne, M., Veneault, C., Laugé, R., and Langin, T. (2002). clap1, a gene encoding a copper-transporting ATPase involved in the process of infection by the phytopathogenic fungus *Colletotrichum lindemuthianum*. *Mol. Genet. Genomics* 268, 139–151. doi: 10.1007/s00438-002-0744-8

Pellier, A. L., Laugé, R., Veneault-Fourrey, C., and Langin, T. (2003). CLNR1, the AREA/NIT2-like global nitrogen regulator of the plant fungal pathogen *Colletotrichum lindemuthianum* is required for the infection cycle. *Mol. Microbiol.* 48, 639–655. doi: 10.1046/j.1365-2958.2003.03451.x

Pereira, M. F., de Araújo dos Santos, C. M., de Araújo, E. F., de Queiroz, M. V., and Bazzolli, D. M. S. (2013). Beginning to understand the role of sugar carriers in *Colletotrichum lindemuthianum*: The function of the gene mfs1. *J. Microbiol.* 51, 70–81. doi: 10.1007/s12275-013-2393-5

Perfect, S. E., Hughes, H. B., O'Connell, R. J., and Green, J. R. (1999). *Colletotrichum*: A model genus for studies on pathology and fungal-plant interactions. *Fungal Genet. Biol.* 27, 186–198. doi: 10.1006/fgbi.1999.1143

Persegui, J. M. K. C., Oblessuc, P. R., Rosa, J. R. B. F., Gomes, K. A., Chiorato, A. F., Carbonell, S. A. M., et al. (2016). Genome-wide association studies of anthracnose and angular leaf spot resistance in common bean (*Phaseolus vulgaris* L.). *PLoS One* 11, e0150506. doi: 10.1371/journal.pone.0150506

Petersen, T. N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Pinto, J. M. A., Pereira, R., Mota, S. F., Ishikawa, F. H., and Souza, E. A. (2012). Investigating phenotypic variability in *Colletotrichum lindemuthianum* populations. *Phytopathology* 102, 490–497. doi: 10.1094/PHYTO-06-11-0179

Pirc, K., Hodnik, V., Snoj, T., Lenarčič, T., Caserman, S., Podobnik, M., et al. (2021). Nep1-like proteins as a target for plant pathogen control. *PLoS Pathog.* 17, e1009477. doi: 10.1371/journal.ppat.1009477

Pusztahelyi, T., Holb, I. J., and Pócsi, I. (2015). Secondary metabolites in fungus-plant interactions. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00573

Queiroz, C. B., Correia, H. L. N., Menicucci, R. P., Vidigal, P. M. P., and de Queiroz, M. V. (2017). Draft genome sequences of two isolates of *colletotrichum lindemuthianum*, the causal agent of anthracnose in common beans. *Genome Announc* 5, 17–18. doi: 10.1128/genomeA.00214-17

Queiroz, C. B., Correia, H. L. N., Santana, M. F., Batista, D. S., Vidigal, P. M. P., Brommonschenkel, S. H., et al. (2019). The repertoire of effector candidates in *Colletotrichum lindemuthianum* reveals important information about *Colletotrichum* genus lifestyle. *Appl. Microbiol. Biotechnol.* 103, 2295–2309. doi: 10.1007/s00253-019-09639-9

Queiroz, C. B., Santana, M. F., Pereira Vidigal, P. M., and de Queiroz, M. V. (2018). Comparative analysis of the mitochondrial genome of the fungus *Colletotrichum lindemuthianum*, the causal agent of anthracnose in common beans. *Appl. Microbiol. Biotechnol.* 102, 2763–2778. doi: 10.1007/s00253-018-8812-0

Qutob, D., Kamoun, S., and Gijzen, M. (2002). Expression of a *Phytophthora sojae* necrosis-inducing protein occurs during transition from biotrophy to necrotrophy. *Plant J.* 32, 361–373. doi: 10.1046/j.1365-313X.2002.01439.x

Raffaele, S., and Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* 10, 417–430. doi: 10.1038/nrmicro2790

Rao, S., and Nandinini, M. R. (2017). Genome sequencing and comparative genomics reveal a repertoire of putative pathogenicity genes in chilli anthracnose fungus *Colletotrichum truncatum*. *PLoS One* 12, e0183567. doi: 10.1371/journal.pone.0183567

Rao, S., Sharda, S., Oddi, V., and Nandinini, M. R. (2018). The landscape of repetitive elements in the refined genome of chilli anthracnose fungus *Colletotrichum truncatum*. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.02367

Rawlings, N. D., and Bateman, A. (2021). How to use the MEROPS database and website to help understand peptidase specificity. *Protein Sci.* 30, 83–92. doi: 10.1002/pro.3948

Redman, R. S., and Rodriguez, R. J. (2002). Characterization and isolation of an extracellular serine protease from the tomato pathogen *Colletotrichum coccodes*, and its role in pathogenicity. *Mycol. Res.* 106, 1427–1434. doi: 10.1017/S0953756202006883

Reveglia, P., Agudo-Jurado, F. J., Barilli, E., Masi, M., Evidente, A., and Rubiales, D. (2023). Uncovering phytotoxic compounds produced by *colletotrichum* spp. Involved in legume diseases using an OSMAC–metabolomics approach. *J. Fungi* 9, 610. doi: 10.3390/jof9060610

Roca, M. G., Davide, L. C., Mendes-Costa, M. C., and Wheals, A. (2003). Conidial anastomosis tubes in *Colletotrichum*. *Fungal Genet. Biol.* 40, 138–145. doi: 10.1016/S1087-1845(03)00088-4

Romero, G., González, S., Royero, W., and González, A. (2024). Morphological and transcriptional analysis of *Colletotrichum lindemuthianum* race 7 during early stages of infection in common bean. *Genet. Mol. Biol.* 47, e20220263. doi: 10.1590/1678-4685-gmb-2022-0263

Rosada, L. J., Franco, C. C. S., Sant'Anna, J. R., Kaneshima, E. N., Goncalves-Vidigal, M. C., and Castro-Prado, M. A. A. (2010). Parasexuality in Race 65 *Colletotrichum lindemuthianum* isolates. *J. Eukaryotic Microbiol.* 57, 383–384. doi: 10.1111/j.1550-7408.2010.00486.x

Saier, M. H., Reddy, V. S., Moreno-Hagelsieb, G., Hendargo, K. J., Zhang, Y., Iddamsetty, V., et al. (2021). The transporter classification database (TCDB): 2021 update. *Nucleic Acids Res.* 49, D461–D467. doi: 10.1093/nar/gkaa1004

Schmitz, K., Protzko, R., Zhang, L., and Benz, J. P. (2019). Spotlight on fungal pectin utilization—from phytopathogenicity to molecular recognition and industrial applications. *Appl. Microbiol. Biotechnol.* 103, 2507–2524. doi: 10.1007/s00253-019-09622-4

Schumacher, S., Grosser, K., Voegelé, R. T., Kassemeyer, H. H., and Fuchs, R. (2020). Identification and characterization of nep1-like proteins from the grapevine downy mildew pathogen *plasmopara viticola*. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00065

Seidl, M. F., and Van den Ackerveken, G. (2019). Activity and phylogenetics of the broadly occurring family of microbial nep1-like proteins. *Annu. Rev. Phytopathol.* 57, 367–386. doi: 10.1146/annurev-phyto-082718-100054

Sharma, P. N., Padder, B. A., Sharma, O. P., Pathania, A., and Sharma, P. (2007). Pathological and molecular diversity in *Colletotrichum lindemuthianum* (bean anthracnose) across Himachal Pradesh, a north-western Himalayan state of India. *Australas. Plant Pathol.* 36, 191–197. doi: 10.1071/AP07013

Sharma, P. N., Sharma, O. P., Padder, B. A., and Kapil, R. (2008). Yield loss assessment in common bean due to anthracnose (*Colletotrichum lindemuthianum*) under sub temperate conditions of North-Western Himalayas. *Indian Phytopath* 61, 323–330.

Shnaiderman, C., Miyara, I., Kobiler, I., Sherman, A., and Prusky, D. (2013). Differential activation of ammonium transporters during the accumulation of ammonia by *Colletotrichum gloeosporioides* and its effect on appressoria formation and pathogenicity. *Mol. Plant-Microbe Interact.* 26, 345–355. doi: 10.1094/MPMI-07-12-0170-R

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi: 10.1038/msb.2011.75

Silva, L. L., Morgan, T., Garcia, E. A., Rosa, R. O., Mendes, T. A., de, O., et al. (2022). Pectinolytic arsenal of *Colletotrichum lindemuthianum* and other fungi with different lifestyles. *J. Appl. Microbiol.* 133, 1857–1871. doi: 10.1111/jam.15692

Silva, L. L., Rosa, R. O., Garcia, E. A., Teixeira, J. A., Mendes, T. A., de, O., et al. (2021). Use of *Agrobacterium tumefaciens*-mediated transformation for the inactivation of pectinase genes in *Colletotrichum lindemuthianum*. *Eur. J. Plant Pathol.* 162, 595–608. doi: 10.1007/s10658-021-02423-4

Simão, F. A., Waterhouse, R. M., Loannidis Kriventseva, P. 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

Soares, M. A., Nogueira, G. B., Bazzolli, D. M. S., de Araújo, E. F., Langin, T., and de Queiroz, M. V. (2014). PacCl, a pH-responsive transcriptional regulator, is essential in the pathogenicity of *Colletotrichum lindemuthianum*, a causal agent of anthracnose in bean plants. *Eur. J. Plant Pathol.* 140, 769–785. doi: 10.1007/s10658-014-0508-4

- Souza, E. A., Camargo, O. A., and Pinto, J. M. (2010). Sexual recombination in *Colletotrichum lindemuthianum* occurs on a fine scale. *Genet. Mol. Res.* 9, 1759–1769. doi: 10.4238/vol9-3gmr863
- Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120
- Tanford, C. (1983). Mechanism of free energy coupling in active transport. *Annu. Rev. Biochem.* 52, 379–409. doi: 10.1146/annurev.bi.52.070183.002115
- Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* 5, 4–10. doi: 10.1002/0471250953.bi0410s25
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621.Transcript
- Upadhyay, S. K., Gautam, P., Pandit, H., Singh, Y., Basir, S. F., and Madan, T. (2012). Identification of fibrinogen-binding proteins of *Aspergillus fumigatus* using proteomic approach. *Mycopathologia* 173, 73–82. doi: 10.1007/s11046-011-9465-z
- Urban, M., Pant, R., Raghunath, A., Irvine, A. G., Pedro, H., and Hammond-Kosack, K. E. (2015). The Pathogen-Host Interactions database (PHI-base): Additions and future developments. *Nucleic Acids Res.* 43, D645–D655. doi: 10.1093/nar/gku1165
- Van Wyk, S., Harrison, C. H., Wingfield, B. D., De Vos, L., van der Merwe, N. A., and Steenkamp, E. T. (2019). The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. *PeerJ* 7, e7447. doi: 10.7717/peerj.7447
- Van Wyk, S., Wingfield, B. D., De Vos, L., van der Merwe, N. A., and Steenkamp, E. T. (2021). Genome-wide analyses of repeat-induced point mutations in the ascomycota. *Front. Microbiol.* 11. doi: 10.3389/fmicb.2020.622368
- Viggiano, A., Salo, O., Ali, H., Szymanski, W., Lankhorst, P. P., Nygård, Y., et al. (2018). Pathway for the biosynthesis of the pigment chrysogine by *penicillium chrysogenum*. *Appl. Environ. Microbiol.* 84, e02246–e02217. doi: 10.1128/AEM.02246-17
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9, e112963. doi: 10.1371/journal.pone.0112963
- Wang, L., Sun, Y., Sun, X., Yu, L., Xue, L., He, Z., et al. (2020). Repeat-induced point mutation in *Neurospora crassa* causes the highest known mutation rate and mutational burden of any cellular life. *Genome Biol.* 21, 1–23. doi: 10.1186/s13059-020-02060-w
- Wijesundera, R. L. C., Bailey, J. A., Byrde, R. J. W., and Fielding, A. H. (1989). Cell wall degrading enzymes of *Colletotrichum lindemuthianum*: their role in the development of bean anthracnose. *Physiol. Mol. Plant Pathol.* 34, 403–413. doi: 10.1016/0885-5765(89)90067-2
- Winnenburg, R. (2006). PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res.* 34, D459–D464. doi: 10.1093/nar/gkj047
- Wisecaver, J. H., and Rokas, A. (2015). Fungal metabolic gene clusters—caravans traveling across genomes and environments. *Front. Microbiol.* 6. doi: 10.3389/fmicb.2015.00161
- Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, 445–451. doi: 10.1093/nar/gks479
- Zemach, A., McDaniel, I. E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919. doi: 10.1126/science.1186366
- Zhang, L., Huang, X., He, C., Zhang, Q. Y., Zou, X., Duan, K., et al. (2018). Novel fungal pathogenicity and leaf defense strategies are revealed by simultaneous transcriptome analysis of *Colletotrichum fructicola* and strawberry infected by this fungus. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00434
- Zhou, B.-J., Jia, P.-S., Gao, F., and Guo, H.-S. (2012). Molecular characterization and functional analysis of a necrosis-and ethylene-inducing, protein-encoding gene family from *verticillium dahliae*. *Mol. Plant Microbe Interact.* 25, 964–975. doi: 10.1094/MPMI-12-11-0319
- Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., and Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29, 2669–2677. doi: 10.1093/bioinformatics/btt476