ORIGINAL PAPER
Polish Journal of Microbiology
2023, Vol. 72, No 4, 433–442
https://doi.org/10.33073/pjm-2023-041

Genomic Characterization of the Mycoparasite *Pestalotiopsis* sp. Strain cr013 from *Cronartium ribicola*

JINDE YU[#], LEI KONG[#], SHICHANG FAN, MINGJIAO LI and JING LI*

College of Life Science, Southwest Forestry University, Kunming, People's Republic of China

Submitted 17 July 2023, accepted 24 October 2023, published online 15 December 2023

Abstract

The *Pestalotiopsis* sp. strain cr013 is a mycoparasite of *Cronartium ribicola*, a potential biocontrol fungus for Armand pine (*Pinus armandii*) blister rust. A previous study showed that the strain cr013 has great potential to produce new compounds. However, there has been no report of the whole-genome sequence of the mycoparasite *Pestalotiopsis* sp. In this study, the BGISEQ-500 and Oxford Nanopore GridION X5 sequencing platforms were used to sequence the strain cr013 isolates and assemble the reads to obtain the complete genome. We first report the whole-genome information of the mycoparasite *Pestalotiopsis* sp. strain cr013 (GenBank accession number: JACFXT010000000, BioProject ID: PRJNA647543, BioSample ID: SAMN15589943), and the genomic components and gene functions related to the mycoparasitism process were analyzed. This study provides a theoretical basis for understanding the lifestyle strategy of the mycoparasite *Pestalotiopsis* sp. and reveals the mechanisms underlying secondary metabolite diversity in the strain cr013.

K e y w o r d s: mycoparasitism, *Pestalotiopsis* sp., genome, secondary metabolites, lifestyle strategy

Introduction

Pestalotiopsis spp. are important plant pathogens with economic value, and they exhibit a wide distribution and diverse lifestyles, including pathogens, endophytes, and mycoparasites. The genus *Pestalotiopsis* produces a large number of different secondary metabolites. It is a prominent resource for synthesizing new compounds (Xu et al. 2010; Maharachchikumbura et al. 2011). There are currently more reports on endophytic and pathogenic *Pestalotiopsis* sp. strains (Wei et al. 2007; Ding et al. 2008), but few reports on mycoparasitic *Pestalotiopsis* sp. strains.

Mycoparasitism is a natural phenomenon in which the host of a parasitic fungus is another fungus, and mycoparasites are often used as biocontrol agents in agriculture (Alfiky and Weisskopf 2021). The known mechanisms of mycoparasitism include the production of cell wall-degrading enzymes and toxins (Li et al. 2017). In the interaction between a fungus and a host, the fungus needs to hydrolyze the fungal cell wall of the host. Fungal cell walls are mainly composed of chitin, glycoproteins, and glycan. Therefore, chitinase and 1, 3-β-glucanase are important for cell wall degradation and fungal disease resistance (Gruber and Seidl-Seiboth 2012; Tzelepis et al. 2012). Chitinases can be divided into the GH18 and GH19 families. GH18 mainly exists in prokaryotes and eukaryotes, and GH19 mainly exists in higher plants and bacteria (Suginta et al. 2016).

At present, the most significant number of studies involving whole-genome sequencing and the mechanism of mycoparasitism are available for *Trichoderma* spp. Whole-genome sequencing is critical to elucidate the mycoparasitism mechanism of *Trichoderma*. For example, by analyzing the whole genome of *Trichoderma harzianum*, Sun et al. (2016) found that there are 38 hydrolases in *T. harzianum*, and these enzymes are involved in the degradation of the cell wall. Martinez et al. (2008) analyzed the whole genome of *Trichoderma reesei*, and showed that approximately 70% of the carbohydrate-active enzyme (CAZyme) genes encode GHs in the identified clusters. This is consistent with the finding that the products of many genes located in CAZyme-rich regions are involved in plant cell wall degradation.

In a study of the mycoparasitism mechanism of *Pestalotiopsis* spp., Li et al. (2017) found that the wall of the rust spore was not broken and only caused severe deformation of the rust spore and abnormal concen-

[#] Jinde Yu and Lei Kong contribute equally to this work and are co-first authors.

* Corresponding author: J. Li, College of Life Science, Southwest Forestry University, Kunming, People's Republic of China; **e-mail: lijingcas@163.com**

© 2023 Jinde Yu et al.
This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

trations of the contents. Therefore, the mycoparasitism mechanism of *Pestalotiopsis* spp. is different from that of *Trichoderma* spp., and it may be related to the production of toxins. Previous studies have shown that the strain cr013 can produce novel compounds with good antibacterial (Xie et al. 2014) and anticancer (Li et al. 2015) activities that differ from the secondary metabolites of endophytic fungi (Bai et al. 2013). At present, the only whole genomes that have been reported for *Pestalotiopsis* spp. are for the endophytic strains *Pestalotiopsis fici* W106-1 (GCA_000516985.1) and *Pestalotiopsis* sp. JCM 9685 (GCA_001599175.1), but these two fungi are not mycoparasites. Moreover, there are no reports on mycoparasitic *Pestalotiopsis* whole genomes, which could provide a good foundation for understanding the specific lifestyle strategy of mycoparasites.

The strain cr013 of *Pestalotiopsis* spp. has been isolated from the aeciospores of *Cronartium ribicola* and subjected to phytochemical analysis according to its specific niche (Li et al. 2015). In this study, the BGISEQ-500 (BGI-Qingdao, China) and GridION X5 (Oxford Nanopore Technologies, UK) sequencing platforms were used to sequence and assemble the whole genome of the strain cr013. Its genomic components and gene functions were explored. The CAZyme of *Trichoderma atroviride*, *Trichoderma virens*, *T. reesei* (Kubicek et al. 2011), and *T. harzianum* (Sun et al. 2016) were compared to the strain cr013. Additionally, the lifestyle strategies of *Pestalotiopsis* and *Trichoderma* were compared. The results of this study will provide a better understanding of the mycoparasitism lifestyle and the underlying mechanisms.

Experimental

Materials and Methods

Microorganisms and culture conditions. The aeciospores of a fungus infecting *Pinus armandii* were collected in Kunming, Yunnan Province, China, in August 2013. The aeciospores were incubated on filter paper soaked with distilled water until colonies or mycelia appeared at 25°C. After seven days of cultivation, a strain was isolated from the aeciospores and identified as *Pestalotiopsis* sp. (strain cr013). The conidia of the strain cr013 were inoculated into an improved Fries liquid medium (1 g KH_2PO_4, 0.5 g $MgSO_4 \cdot 7H_2O$, 0.1 g NaCl, 0.13 g $CaCl_2 \cdot 2H_2O$, 20.0 g sucrose, 5.0 g ammonium tartrate, 1.0 g yeast extract, 1.0 g NH_4NO_3, 1 l distilled water; natural pH). After 60 hours of culture at room temperature at 150 r/min, the strain cr013 mycelia were filtered and collected, washed with sterile water three times, stored in liquid nitrogen, and sent to BGI Technology Co., Ltd. (China) for follow-up work.

DNA extraction and WGS library construction. *Pestalotiopsis* sp. strain cr013 DNA was isolated using the bacterial genomic DNA Extraction Kit (Tiangen Biotech(Beijing)Co., Ltd., China) and cut into 50–800 bp fragments with a Covaris® E220 Focused-ultrasonicator (Covaris, LLC., UK). High-quality 150–250 bp DNA fragments were screened using AMPure XP beads (Beckman Coulter, Inc., USA) and repaired with T4 DNA polymerase (Enzymatics, Inc., USA). The ends of selected fragments were ligated with T-tailed adapters and magnified over eight cycles by KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Inc., USA). Then, the amplification products were subjected to single-strand circularization using T4 DNA ligase to generate a single-stranded circular DNA library. Subsequently, T4 DNA ligase was used to cyclize the magnified products to create a single-stranded circular DNA library.

Genome sequencing and assembly. The WGS library was sequenced using the BGISEQ-500 platform (BGI-Qingdao, China), and the original 100 bp double-terminal sequencing data were obtained. The raw reads with a high ratio of N (ambiguous) bases and low-mass bases were removed using SOAPnuke (v1.6.5) (Chen et al. 2018) with the following parameters: "-l 15 -q 0.2 -n 0.05 -Q 2 -c 0", to obtain clean data. Genomic DNA was quantified by dsDNA BR analysis, and DNA integrity was evaluated by pulsed-field gel electrophoresis (PFGE). To obtain a long Oxford Nanopore fragment, BluePippin (SAGE Science, Inc., USA) was used to select fragments of approximately five μg of genomic DNA. Library preparation and sequencing were performed using Ligation Sequencing Kits SQK-LSK109 (Oxford Nanopore Technologies, UK). DNA repair, end repair, and A-tailing were carried out with NEBNext® FFPE DNA Repair Mix (M6630; New England Biolabs® Inc., USA) and NEBNext® Ultra™ II End Repair/dA-tailing Module (E7546; New England Biolabs® Inc., USA). The DNA was then purified with AMPure XP (A63882; Beckman Coulter Inc., USA). DNA was linked to the adaptor by NEBNext® Quick T4 DNA Ligase (E6056; New England Biolabs® Inc., USA). According to the instructions, samples were loaded on FLO-MIN 106D R9.4.1 and sequenced on GridION for 48 hours. Prior to assembly, the size of the genome, heterozygosity, and repeatability were estimated by *k*-mer analysis according to the second-generation data. The nanopore assembly of the clean BGI-seq 500 WGS reads was conducted with Canu (Koren et al. 2017), which is larger than two kb with the following parameters: "useGrid = false maxThreads = 30 maxMemory = 60 g -nanopore-raw *.fastq -p -d". Then, we used Pilon (Walker et al. 2014) to correct base errors with the second-generation data and obtain the final assembly results. The assembly's confidence with sordariomyceta_odb9 was assessed using BUSCO (v3.0.1) (Simão et al. 2015).

Identification of repetitive elements. Multiple tools were used to identify the repetitive elements. Transposable elements (TE) were recognized by alignment against the Repbase (Bao et al. 2015) database using RepeatMasker (v 4.0.5) (Tarailo-Graovac and Chen 2009) with the following parameters: "-nolow -no_is -norna -engine wublast", and RepeatProteinMasker (v 4.0.5) with the following parameters: "-noLowSimple -pvalue 0.0001", at the DNA and protein levels, respectively. The *de novo* repeat library was evaluated using RepeatModeler (v1.0.8) and LTR-FINDER (v1.0.6) (Xu and Wang 2007) with the default parameters. For the *de novo*-identified repeats, repeat sequences were classified using RepeatMasker (v4.0.5) with the same parameters. In addition, the tandem repeats were identified using Tandem Repeat Finder (v4.04) (Benson 1999) with the following parameters: "2 7 7 80 10 50 000 -d -h".

Gene prediction and functional annotation. For the prediction of homologous genes, the protein sequences of *P. fici* W106-1 (GCA_000516985.1) and *Pestalotiopsis* sp. JCM 9685 (GCA_001599175.1) were downloaded from NCBI by homology-based prediction. The genomes were annotated by homologue using GeneWise (http://www.ebi.ac.uk/Tools/psa/genewise) with default parameters. For *de novo* prediction, Augustus (http://bioinf.uni-greifswald.de/augustus) and GeneMark (http://exon.gatech.edu/GeneMark) softwares were used with the parameters "--ES – fungus". Finally, EVM (http://evidencemodeler.sourceforge.net) integration software was used to integrate different annotation results. Among them, the weight of homologous annotation results was set as 10, and the weight of prediction results of two *de novo* software programs was set as 1. To evaluate the integrity of the predicted gene set, BUSCO software and the sordariomyceta_Odb9 database were used to evaluate the integrity of our assembled protein-coding genes.

The predicted genes were aligned to the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2004), SwissProt (Magrane and UniProt Consortium 2011), Cluster of Orthologous Groups of proteins (COG) (Tatusov et al. 2003), CAZyme (Cantarel et al. 2009), NR database, Pathogen Host Interactions (PHI) and NR using BLAST (v2.2.26) with the parameters "-p BLASTP -e 1e-5 -F F -a 4 -m 8", E-value $< 1 \times 10^{-5}$. Additionally, the motifs and domains of genes were identified by protein databases, including Pfam, SMART, PANTHER, PRINTS, ProSite, and ProDom using InterProScan (5.16–55.0) (http://www.ebi.ac.uk/Tools/InterProScan), and Gene Ontology (GO) (Ashburner et al. 2000) databases using BLAST (v2.2.26) with the parameters "-p BLASTP -e 1e-5 -F F -a 4 -m 8", E-value $< 1 \times 10^{-5}$. The cr013 strain assembly was uploaded to the antiSMASH (v5.0) (Blin et al. 2019) website to identify the secondary metabolite gene clusters.

Identification of candidate gene families. The mechanism of mycoparasites mainly includes the production of cell wall degrading enzymes and small molecular toxins, and these processes involve CAZyme, cytochrome P450, protease, and Zn2Cys6 transcription factor. Therefore, to comparatively analyze these genes in strain cr013, *T. atroviride*, *T. virens*, *T. reesei* and *T. harzianum*, BLASTP and Hmm (http://hmmer.janelia.org) software were used to identify candidate gene families. First, the protein sequences of four other *Trichoderma* species were downloaded from the NCBI database and aligned against the protein sequences of five species using BLASTP v2.2.31 with an E-value of 1×10^{-5} to search for significant hits. Additionally, Hmmsearch was used to predict gene families in conjunction with the Hmm model from the Pfam database. The protein sequences of each gene family were aligned using MUSCLE v3.8.1551 (Edgar 2004) if the gene family lacked the Hmm model, and hmm models were built using Hmmbuild. The significant hits (E-value of 1×10^{-5}) from BLASTP and Hmmsearch were dereplicated, filtered, and merged to generate final hits. The sequences without domains of the resulting hits were removed using the Conserved Domain Database (CDD) (Lu et al. 2020) and Simple Modular Architecture Research Tool (SMART) (Letunic and Bork 2018; Letunic et al. 2020) databases. Furthermore, the sequences coding fewer than 80 amino acids were discarded, and the longest transcript was reserved because multiple transcripts were identified.

Results

Genome extraction and quality detection. The concentration and purity of the strain cr013 were determined using a Qubit™ 4 Fluorometer (Invitrogen™, USA) and a NanoDrop™ Spectrophotometer (Thermo Fisher Scientific Inc., USA) and then assessed by 1% agarose gel electrophoresis. DNA integrity was assessed by gel electrophoresis (Fig. 1). According to the Quality Standard of Nanopore Genome Sequencing Samples, the quality of the samples met the requirements for database building, and subsequent genome sequencing could be carried out.

Genome sequencing and assembly. The Nanopore platform was used to sequence the long fragment of the strain cr013, and a total of 11.10 G of raw data was generated. Prior to assembly, the K-mer was selected as 15, and *k*-mer analysis was performed according to the second-generation data to evaluate the genome size, heterozygosity, and repeatability. The filtered data were processed using Jellyfish software. The results indicate that the genome size of strain cr013 was 47.4 Mb, the genome repeatability was 18.5%, and the heterozygosity was 0.06%. Nanopore data were assembled using Canu,

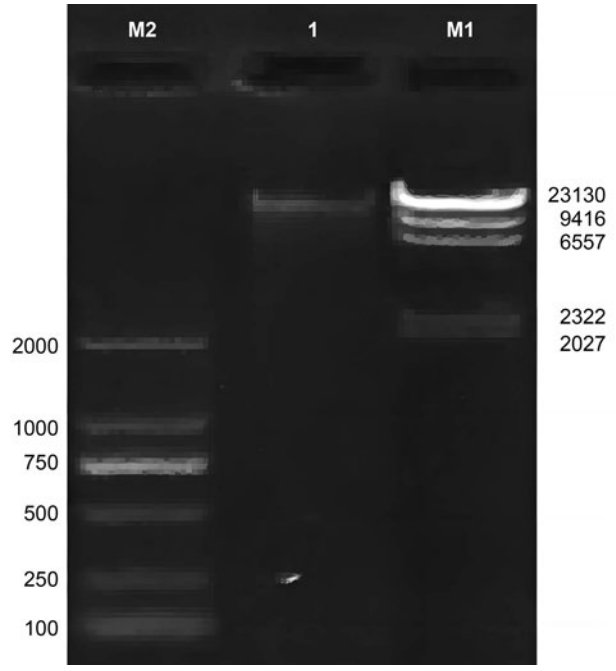


Fig. 1. Detection of the strain cr013 genome by electrophoresis.
The concentration of agarose: 1%; voltage: 180 V; time: 36 min.; molecular weight standard name: M1: λ-Hind III digest (TaKaRa Bio, Inc., Japan), M2: D2000 (Tiangen Biotech(Beijing)Co., Ltd., China); sample volume: M1: 2.5 μl, M2: 5 μl.

and the second-generation data were used for basic error correction using Pilon to obtain a fine genome map. A total of 13 scaffolds were assembled into the genome with a size of 46.69 Mb. BUSCO was used to evaluate the integrity of the genome assembly. The results showed that 97.2% of the core genes could be annotated in the genome. Complete and single-copy BUSCOs were 96.80%, complete and duplicated BUSCOs accounted for 0.4%, fragmented BUSCOs accounted for 1%, missing BUSCOs accounted for 1.8%, and 3,725 BUSCO groups, which reflected the high integrity of the assembly results, were searched in total.

Features of the strain cr013 genome. The *Pestalotiopsis* sp. strain cr013 mycoparasite genome (GenBank accession number: JACFXT010000000, BioProject ID: PRJNA647543, BioSample ID: SAMN15589943) was assembled into 13 scaffolds with an N50 of 6.39 Mb, total size of 46.69 Mb, and a 52.1% GC content. A total of 14,893 genes were predicted. Through BUSCO and sordariomyceta_Odb9 database analysis, 96.5% of the genes were predicted to be complete, reflecting the high annotation integrity of the annotated gene set. Regarding repetitive sequences, the cr013 strain showed a 2.07% repeat content and 1.68% transposable element (TE) content (Table I).

Gene prediction and functional annotation. After obtaining the genome of the cr013 strain, we compared and annotated the gene database to determine the functions and related descriptions of the genes. The results reflect the functional classification of the gene set of the strain as a whole and facilitate subsequent research to identify target functional genes. In this study, the predicted gene set was aligned to NR, KEGG, COG, SwissProt, GO, ARDB, PHI, and CAZyme using BLASTP (E-value $\leq 1 \times 10^{-5}$). The results are shown in Table II. Finally, approximately 94.10% of the protein sequences were aligned to eight databases.

KEGG pathway analysis. In KEGG annotations, genes are usually classified into five branches based on the following KEGG metabolic pathway categories:

Table I
Main features of genome of the strain cr013 and four isolates of *Trichoderma* species.

Feature	Strain cr013	*Trichoderma atroviride*	*Trichoderma virens*	*Trichoderma reesei*	*Trichoderma harzianum*
Genome size (Mb)	46.69	36.1	38.8	34.1	38.8
Number of scaffolds	13	50	135	89	498
Scaffold N50 (Mb)	6.39	2	1.8	1.2	6.2
GC content (%)	52.1	49.5	49	52.5	49.5
Protein-coding genes	14,893	11,865	12,518	9,143	11,264
Exons per gene	2.95	2.93	2.98	3.06	/
Exon length (bp)	498.14	528.17	506.13	507.81	/
Intron length (bp)	113.76	104.20	104.95	119.64	/

Table II
Statistics of the annotation results of different database functions.

Strain cr013	Total	NR	Swissport	KEGG	COG	GO	CAZy	PHI	ARDB	Overall
Number	14893	14003	8869	9768	2812	8924	710	20	1735	14,014
Percentage	100%	94.02%	59.55%	65.59%	18.88%	59.92%	4.77%	0.13%	11.65%	94.10%

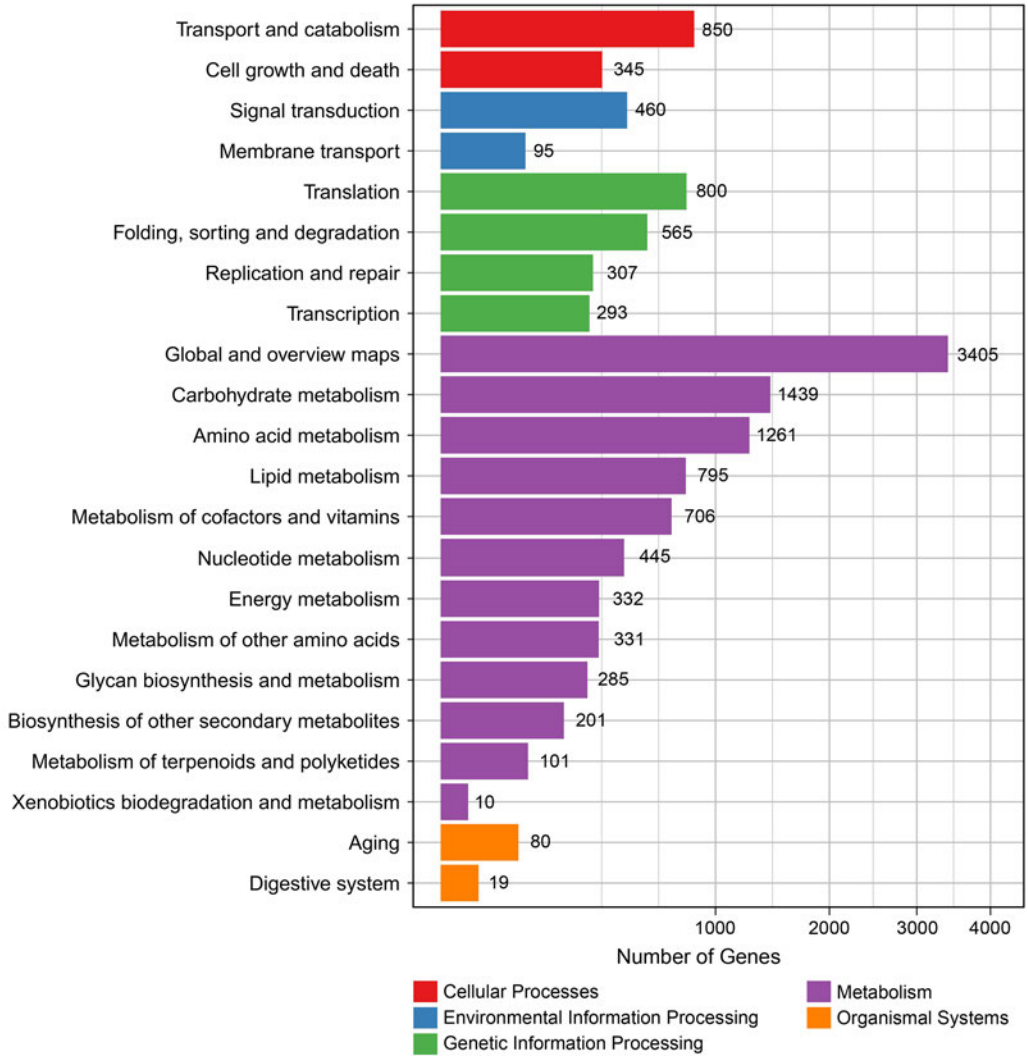


Fig. 2. Classification of the KEGG pathways of the cr013 strain.

cell processes, environmental information processing, genetic information processing, metabolism, and organic systems. Among the five pathways annotated through the KEGG analysis of the strain cr013, metabolic pathways were the most enriched category, mainly related to the metabolism of carbohydrates, amino acids, lipids, coenzymes, and vitamins (Fig. 2). The analysis of KEGG enrichment showed that 9,768 genes corresponding to KEGG pathways were enriched in 129 metabolic pathways. Five branches of metabolic pathways involved the most genes (Table III).

COG functional classification. According to the COG classification of the strain cr013 genes obtained by sequencing, 2,812 genes were classified into 24 categories (Fig. 3), except for general function prediction. The most involved genes were posttranslational modification, protein turnover, chaperones (195 genes, 6.93%), signal transduction mechanisms (171 genes, 6.08%), translation, ribosomal structure and biogenesis (228 genes, 8.1%), amino acid transport and metabolism (271 genes, 9.63%), carbohydrate transport and metabolism (306 genes, 10.9%), coenzyme transport and metabolism (206 genes, 7.32%), energy production and conversion (327 genes, 11.62%), lipid transport and metabolism (376 genes, 13.37%), and the biosynthesis, transport and catabolism of secondary metabolites (319 genes, 11.34%).

GO functional annotation. A total of 8,924 genes were annotated in the GO database, which can be divided into three categories with 52 components (Fig. 4): biological processes (25 branches), cellular components (14 branches), and molecular functions (13 branches). The most significant number of genes in the biological process category were associated with metallic processes and cellular processes; the greatest number of genes in the cellular component category were associated with membranes, membrane parts, cells, cell parts, and organelles; and the greatest number of genes in the molecular function category were related to quantitative activity and binding.

CAZyme of the strain cr013. CAZyme is a classic database relevant to carbohydrate-active enzymes,

Table III
Gene annotation of the KEGG pathways of the strain cr013.

Pathway	Genes with pathway annotation	Pathway ID
Metabolic pathways	3,279	ko01100
Biosynthesis of secondary metabolites	1,153	ko01110
Biosynthesis of antibiotics	816	ko01130
MAPK signaling pathway – yeast	404	ko04011
Glycine, serine and threonine metabolism	356	ko00260
RNA transport	304	ko03013
Biosynthesis of amino acids	288	ko01230
Carbon metabolism	280	ko01200
Glycerophospholipid metabolism	231	ko00564
Meiosis – yeast	230	ko04113

which include many enzymes that can catalyze the degradation, modification, and biosynthesis of carbohydrates. Among the 709 genes annotated in strain cr013, 255 were glycoside hydrolases (GHs), 92 were glycosyl transferases (GTs), 17 were polysaccharide lyases (PLs), 41 were carbohydrate esterases (CEs), 170 were

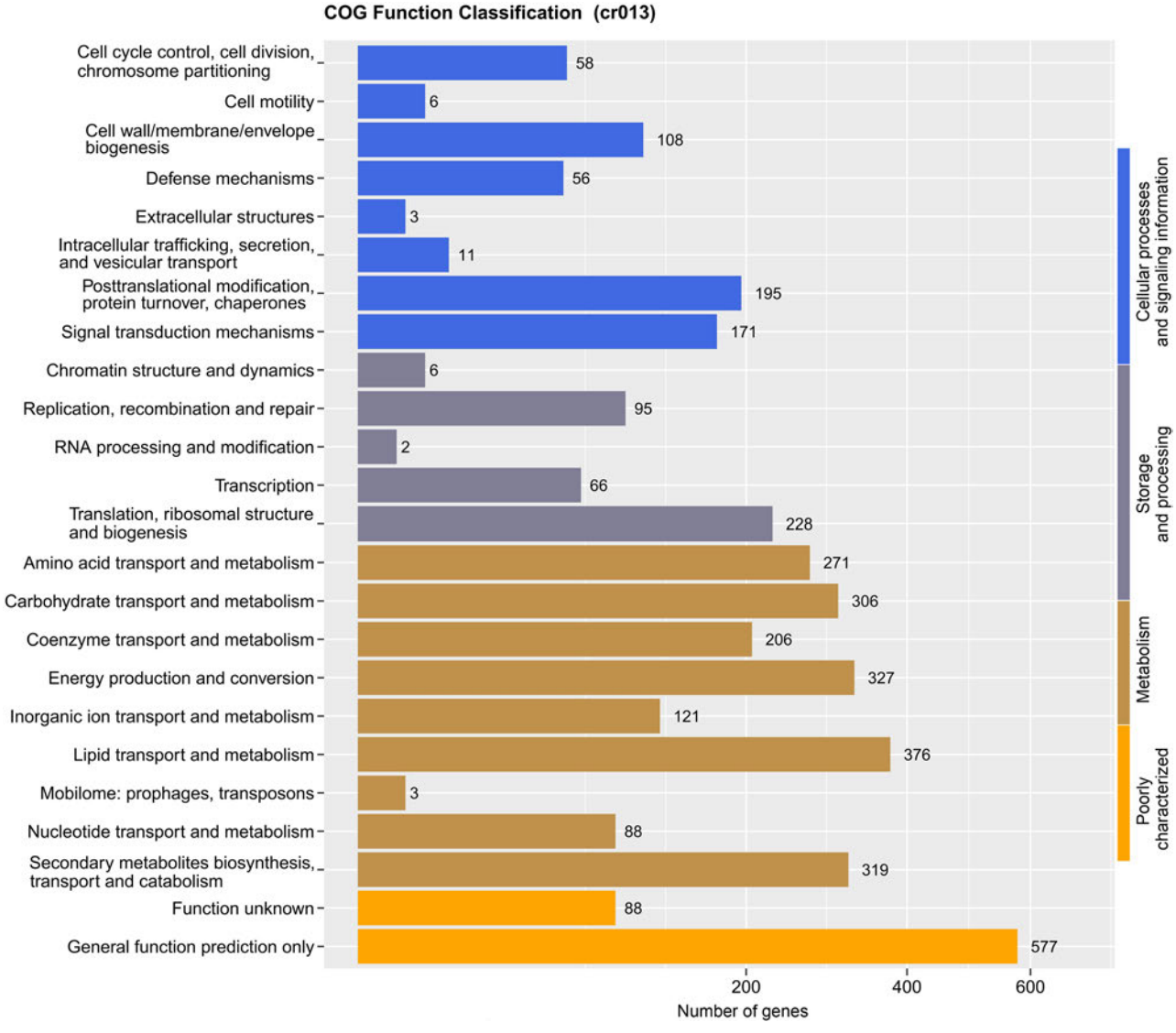


Fig. 3. COG functional classification of the cr013 strain.

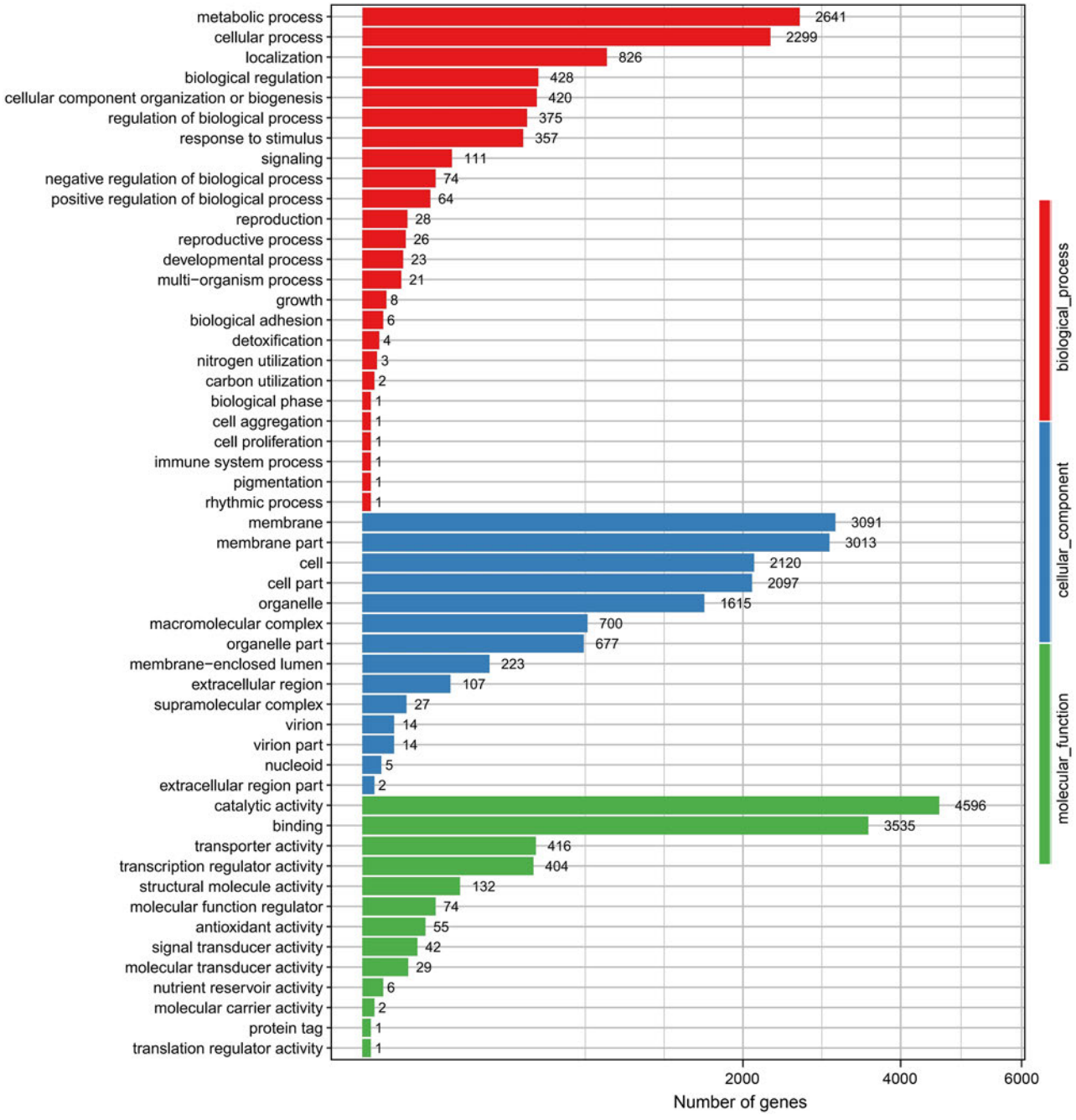


Fig. 4. GO functional annotation of the cr013 strain.

carbohydrate-binding modules (CBMs), and 134 were auxiliary activities (AASs).

Fungi can produce many carbohydrate-active enzymes, among which chitinase and 1, 3-β-glucan enzymes are important hydrolase enzymes (Zhao et al. 2013). Therefore, the GH families of *Trichoderma* spp. and the strain cr013 were further analyzed. There were three GH18 and four GH19 genes in the chitinase family of the cr013 strain. The number of chitinase genes in the cr013 strain was significantly lower than that in *Trichoderma* sp. mycoparasite. The number of 1, 3-β-glucans (GH17, GH55, GH64 and GH81 families) in the strain cr013 was 20, more significant than the number in the selected *Trichoderma* sp. mycoparasite. Cuomo et al. (2007) found that the number of GH75 family genes (chitosanase) increased significantly during cell wall degradation. The number of GH75 family genes in the strain cr013was similar to that in the *Trichoderma* sp. mycoparasite (Table IV). The results showed that there were significantly fewer hydrolytic enzymes related to cell wall degradation in the strain cr013 than in the selected *Trichoderma* sp. mycoparasite, which could explain the difference in mycoparasitism mechanism between *Trichoderma* and *Pestalotiopsis*.

Secondary metabolite biosynthesis ability of the strain cr013. The secondary metabolite gene clusters of

Table IV
Numbers of glycoside hydrolases related to mycoparasitism between the strain cr013 and four *Trichoderma* species.

Species	GH18	GH17	GH55	GH64	GH81	GH75	Total
Strain cr013	3	7	9	3	1	5	28
Trichoderma atroviride	29	5	8	3	2	5	52
Trichoderma virens	36	4	10	3	1	5	59
Trichoderma reesei	20	4	10	3	1	5	43
Trichoderma harzianum	20	4	5	3	2	4	38

Table V
Numbers of BGCs in the strain cr013 and selected fungi.

Species	Terpene	NRPS	NRPS-like	T1 PKS	T3 PKS	Indole	NRPS-PKS	Batalactone	NRPS-terpene	RiPP	Indole-PKS	Total
Strain cr013	6	9	15	19	1	2	6	1	2	1	2	64
Pestalotiopsis fici	10	11	16	24	1	4	4	2		2	1	75
Pestalotiopsis sp. JCM	10	12	14	18	1	3	5	1		1	2	67
Trichoderma atroviride	8	10	10	12			4					44
Trichoderma virens	10	16	11	14			6		1		1	59
Trichoderma reesei	8	6	5	9			4					32
Trichoderma harzianum	9	9	9	20			8			1		56

strain cr013 and other selected fungi were analyzed using antiSMASH with default parameters, and the results are shown in Table V. Two ambuic acids and four reducing polyketides with novel structures that had strong antitumor activity were isolated from the mycoparasitic strain cr013. Polyketide compounds were found for the first time in the mycoparasitic *Pestalotiopsis* sp. Therefore, the biological synthesis mechanism of these compounds was explored by genome sequencing, and more secondary metabolites were expected to be found. Further analysis results by antiSMASH showed that the genes of the strain cr013 were distributed in 64 gene clusters, including six terpenes, nine NRPSs, 15 similar NRPSs, 19 T1 PKSs, one T3 PKS, two indoles, six hybrid NRPS-PKSs, one betalactone, two hybrid NRPS-terpenes, one RiPP, and two hybrid indole-PKSs. Compared to the selected fungi, the number of gene clusters of secondary metabolites of the strain cr013 was much greater than *Trichoderma* sp. mycoparasite. The number of gene clusters of secondary metabolites was less than endophytic *Pestalotiopsis* sp. However, the strain cr013 contained hybrid NRPS-terpenes, which was not predicted in the *Pestalotiopsis* sp. reference genome. According to the results analysed by antiSMASH, the *Pestalotiopsis* strain cr013 had the potential to produce abundant secondary metabolites.

Cytochrome P450 participates in many important cellular processes (van den Brink et al. 1998) and complex biotransformation processes in fungi, and the enzymes of this group catalyze the conversion of hydrophobic intermediates of primary and secondary metabolic pathways (Crešnar and Petrič 2011). There are 234 cytochrome P450-coding bases and 74 protease-coding bases in the strain cr013 genome, which is a significantly higher number than in the selected *Trichoderma* sp. genome. A total of to 152 transcription factors were found in the genome sequence analysis, among which the 13 genes encoding C2H2 transcription factors, and three genes encoding Zn2cys6 transcription factors represented a significantly smaller number than in the *Trichoderma* sp. genome (Table VI).

Table VI
Numbers of P450 enzymes, proteases and Zn2Cys6 transcription factors in the strain cr013 and four *Trichoderma* species.

Other genes	Strain cr013	*Trichoderma atroviride*	*Trichoderma virens*	*Trichoderma reesei*	*Trichoderma harzianum*
Cytochrome P450	234	69	120	73	120
Protease	74	24	28	14	53
Zn2Cys6 transcription factor	3	69	95	9	7

Discussion

Mycoparasitism is an important mechanism by which fungi antagonize pathogens (Wang et al. 2023). Thus far, the mechanism of *Trichoderma* sp. mycoparasitism is mainly related to the degradation of the pathogen cell wall, and the outer wall of the rust aeciospores was observed to be distorted or even completely broken after treatment (Li et al. 2014). In this research, the whole genome of a *Pestalotiopsis* sp. mycoparasite was analyzed. However, the number of hydrolytic enzymes related to cell wall degradation in the strain cr013 was significantly lower than in the *Trichoderma* sp. mycoparasite. Additionally, a previous study had shown that the cell walls of rust spores were deformed but not broken when the strain cr013 was the mycoparasite of rust aeciospores (Li et al. 2017). In view of this, we speculate that cell wall-degrading enzymes may not be the primary mycoparasitism mechanism of the strain. However, the number of 1,3-β-glucans (GH17, GH55, GH64, and GH81 families) is greater than that in the selected fungi, which requires further study.

The genes of transporters and regulators, and some other genes that usually cluster with the core genes, are necessary for the biosynthesis of secondary metabolites in fungi (Wang et al. 2015). Polyketide synthase (PKS) is a complex multienzyme system that regulates the synthesis of secondary metabolite polyketide compounds (Dutta et al. 2014). Nonribosomal peptide synthase (NRPS) can synthesize secondary metabolites, such as antibiotics and siderophores, which are pharmacologically important (Fischbach and Walsh 2006). We compared the number of PKSs and NRPSs between the strain cr013 and the selected fungi, and the results showed that the strain cr013 harbors significantly more PKSs and NRPSs than the selected fungi (Table V). This might be relevant to the higher potential of the strain cr013 to produce secondary metabolites. Cytochrome P450 enzymes play an important role in fungal biology and ecology. Cytochrome P450s are involved in producing diverse secondary metabolites and are essential in an organism's adaptation to specific environments (van den Brink et al., 1998). The strain cr013 harbors significantly more P450 enzymes and proteases than the selected fungi (Table VI), which indicated that the strain cr013 might produce more secondary metabolites. Transcription factors participate in the synthesis of secondary metabolites and regulate gene expression (Seidl et al. 2009). Zn2Cys6 transcription factors may regulate the expression of gene clusters. The number of Zn2Cys6 transcription factors in the strain cr013 was less than that in the selected fungi (Table VI), and there were apparent differences between them, which require further study.

The mycoparasite *Pestalotiopsis* sp. strain cr013 is lethal to rust spores and safe for host plants, and it can be used as a potential biocontrol fungus. Furthermore, it shows a high potential to produce secondary metabolites. Therefore, it has good research value. In this study, we revealed the first whole genome information for the *Pestalotiopsis* mycoparasite, and we obtained many genes related to mycoparasitism. The genome sequence data offer a better understanding of the lifestyle strategy of mycoparasitic *Pestalotiopsis* spp. The genome sequence will be helpful for future studies on the mining of novel bioactive secondary metabolites and metabolic pathways of *Pestalotiopsis* mycoparasites.

Acknowledgments

This work was supported by the Ycr013unnan Agricultural Basic Research Joint Special Area Project (202101BD070001-056) and the Biology Quality Engineering Project (503190106).

Conflict of interest

The authors do not report any financial or personal connections with other persons or organizations, which might negatively affect the contents of this publication and/or claim authorship rights to this publication.

Literature

Alfiky A, Weisskopf L. Deciphering *Trichoderma*-plant-pathogen interactions for better development of biocontrol applications. J Fungi (Basel). 2021 Jan;7(1):61. https://doi.org/10.3390/jof7010061

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000 May;25(1):25–29. https://doi.org/10.1038/75556

Bai ZQ, Lin XP, Liu YH. [Research progress on the chemical constituents from the endophytic fungus *Pestalotiopsis* spp.] (in Chinese) Nat Prod Res Dev. 2013 May;25:706–715. https://doi.org/10.16333/j.1001-6880.2013.05.028

Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015 Jun 2;6:11. https://doi.org/10.1186/s13100-015-0041-9

Benson G. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 1999 Jan;27(2):573–580. https://doi.org/10.1093/nar/27.2.573

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019 Jul; 47(W1):W81–W87. https://doi.org/10.1093/nar/gkz310

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. Nucleic Acids Res. 2009 Jan; 37(suppl_1):D233–D238. https://doi.org/10.1093/nar/gkn663

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, et al. SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 2018 Jan;7(1):gix120. https://doi.org/10.1093/gigascience/gix120

Crešnar B, Petrič S. Cytochrome P450 enzymes in the fungal kingdom. Biochim Biophys Acta. 2011 Jan;1814(1):29–35. https://doi.org/10.1016/j.bbapap.2010.06.020

Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M, et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science. 2007 Sep;317(5843):1400–1402. https://doi.org/10.1126/science.1143708

Ding G, Jiang L, Guo L, Chen X, Zhang H, Che Y. Pestalazines and pestalamides, bioactive metabolites from the plant pathogenic fungus *Pestalotiopsis theae*. J Nat Prod. 2008 Nov;71(11):1861–1865. https://doi.org/10.1021/np800357g

Dutta S, Whicher JR, Hansen DA, Hale WA, Chemler JA, Congdon GR, Narayan AR, Håkansson K, Sherman DH, Smith JL, et al. Structure of a modular polyketide synthase. Nature. 2014 Jun; 510(7506):560–564. https://doi.org/10.1038/nature13423

Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004 Mar;32(5): 1792–1797. https://doi.org/10.1093/nar/gkh340

Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: Logic, machinery, and mechanisms. Chem Rev. 2006 Aug;106(8):3468–3496. https://doi.org/10.1021/cr0503097

Gruber S, Seidl-Seiboth V. Self versus non-self: Fungal cell wall degradation in *Trichoderma*. Microbiology. 2012 Jan;158(1):26–34. https://doi.org/10.1099/mic.0.052613-0

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004 Jan;32(Suppl_1):D277–D280. https://doi.org/10.1093/nar/gkh063

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. 2017 May; 27(5):722–736. https://doi.org/10.1101/gr.215087.116

Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, et al. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. Genome Biol. 2011;12(4):R40. https://doi.org/10.1186/gb-2011-12-4-r40

Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018 Jan;46(D1):D493–D496. https://doi.org/10.1093/nar/gkx922

Letunic I, Khedkar S, Bork P. SMART: Recent updates, new developments and status in 2020. Nucleic Acids Res. 2021 Jan;49(D1): D458–D460 https://doi.org/10.1093/nar/gkaa937

Li J, Liu FL, Cheng YH. Culture condition screening for toxin production of mycopa-rasites (*Pestalotiopsis* spp.) from *Cronartium ribicola*. Acta Agric Univ Jiangxiensis. 2017;39:395–401. https://doi.org/10.13836/j.jjau.2017051

Li J, Xie J, Yang YH, Li XL, Zeng Y, Zhao PJ. Pestalpolyols A–D, cytotoxic polyketides from *Pestalotiopsis* sp. cr013. Planta Med. 2015 Sep;81(14):1285–1289. https://doi.org/10.1055/s-0035-1546257

Li J, Yang YH, Zhou L, Cheng LJ, Chen YH. Destructive effects of a mycoparasite *Trichoderma atroviride* SS003 on aeciospores of *Cronartium ribicola*. J Phytopathol. 2014; 162(6):396–401. https://doi.org/10.1111/jph.12202

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, et al. CDD/SPARCLE: The conserved domain database in 2020. Nucleic Acids Res. 2020 Jan;48(D1):D265–D268. https://doi.org/10.1093/nar/gkz991

Magrane M; UniProt Consortium. UniProt Knowledgebase: A hub of integrated protein data. Database. 2011 Mar;2011:bar009. https://doi.org/10.1093/database/bar009

Maharachchikumbura SSN, Guo LD, Chukeatirote E, Bahkali A, Hyde KD. *Pestalotiopsis* – morphology, phylogeny, biochemistry and diversity. Fungal Diversity. 2011;50:167–187. https://doi.org/10.1007/s13225-011-0125-x

Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, et al. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat Biotechnol. 2008 May;26(5):553–560. https://doi.org/10.1038/nbt1403

Seidl V, Song L, Lindquist E, Gruber S, Koptchinskiy A, Zeilinger S, Schmoll M, Martínez P, Sun J, Grigoriev I, et al. Transcriptomic response of the mycoparasitic fungus *Trichoderma atroviride* to the presence of a fungal prey. BMC Genomics. 2009 Nov;10:567. https://doi.org/10.1186/1471-2164-10-567

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

Suginta W, Sirimontree P, Sritho N, Ohnuma T, Fukamizo T. The chitin-binding domain of a GH-18 chitinase from *Vibrio harveyi* is crucial for chitin-chitinase interactions. Int J Biol Macromol. 2016 Dec;93(Pt A):1111–1117. https://doi.org/10.1016/j.ijbiomac.2016.09.066

Sun Q, Jiang XL, Pang L, Wang LR, Li M. The genome sequence of *Trichoderma harzianum* Th-33. Chinese J Biol Control. 2016;32: 205–214. https://doi.org/10.16409/j.cnki.2095-039x.2016.02.011

Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009 Mar;25(1):4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al. The COG database: An updated version includes eukaryotes. BMC Bioinformatics. 2003 Sep;4:41. https://doi.org/10.1186/1471-2105-4-41

Tzelepis GD, Melin P, Jensen DF, Stenlid J, Karlsson M. Functional analysis of glycoside hydrolase family 18 and 20 genes in *Neurospora crassa*. Fungal Genet Biol. 2012 Sep;49(9):717–730. https://doi.org/10.1016/j.fgb.2012.06.013

van den Brink HM, van Gorcom RF, van den Hondel CA, Punt PJ. Cytochrome P450 enzyme systems in fungi. Fungal Genet Biol. 1998 Feb;23(1):1–17. https://doi.org/10.1006/fgbi.1997.1021

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014 Nov;9(11):e112963. https://doi.org/10.1371/journal.pone.0112963

Wang X, Zhang X, Liu L, Xiang M, Wang W, Sun X, Che Y, Guo L, Liu G, Guo L, et al. Genomic and transcriptomic analysis of the endophytic fungus *Pestalotiopsis fici* reveals its lifestyle and high potential for synthesis of natural products. BMC Genomics. 2015 Jan;16(1):28. https://doi.org/10.1186/s12864-014-1190-9

Wang Y, Zhu X, Wang J, Shen C, Wang W. Identification of mycoparasitism-related genes against the phytopathogen *Botrytis cinerea* via transcriptome analysis of *Trichoderma harzianum* T4. J Fungi. 2023 Mar;9(3):324. https://doi.org/10.3390/jof9030324

Wei JG, Xu T, Guo LD, Liu AR, Zhang Y, Pan XH. Endophytic *Pestalotiopsis* species associated with plants of *Podocarpaceae*, *Theaceae* and *Taxaceae* in southern China. Fungal Diversity. 2007;24: 55–74.

Xie J, Li J, Yang YH, Chen YH, Zhao PJ. Two new ambuic acid analogs from *Pestalotiopsis* sp. cr013. Phytochem Lett. 2014 Oct;10: 291–294. https://doi.org/10.1016/j.phytol.2014.10.002

Xu J, Ebada S, Chaidir C. *Pestalotiopsis* a highly creative genus: Chemistry and bioactivity of secondary metabolites. Fungal Divers. 2010 Aug;44:15–31. https://doi.org/10.1007/s13225-010-0055-z

Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007 Jul; 35(suppl_2):W265–W268. https://doi.org/10.1093/nar/gkm286

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