ORIGINAL PAPER

Genomic Analysis of the Mycoparasite *Pestalotiopsis* sp. PG52

DENGYUN ZHANG[1] ⬤, JINDE YU[1], CHANGLE MA[2], LEI KONG[1], CHENGZHONG HE[1*] and JING LI[1*]

[1] College of Life Science, Southwest Forestry University, Kunming, China
[2] School of Landscape Architecture and Horticulture Science, Southwest Forestry University, Kunming, China

A b s t r a c t

Pestalotiopsis sp. is a mycoparasite of the plant pathogen *Aecidium wenshanense*. To further understand the mycoparasitism mechanism of *Pestalotiopsis* sp., we assembled and analyzed its genome. The genome of *Pestalotiopsis* sp. strain PG52 was assembled into 335 scaffolds and had a size of 58.01 Mb. A total of 20,023 predicted genes and proteins were annotated. This study compared PG52 with the mycoparasites *Trichoderma harzianum*, *Trichoderma atroviride,* and *Trichoderma virens*. This study reveals the entirely different mycoparasitism mechanism of *Pestalotiopsis* compared to *Trichoderma* and reveals this mycoparasite's strong ability to produce secondary metabolites.

K e y w o r d s: genome, DNA sequencing, *Pestalotiopsis* sp., mycoparasite

Introduction

Pestalotiopsis sp. is a mitosporic fungus with spore-forming conidia. This fungus has a wide distribution and a variety of life habits, including pathogenicity (Wang et al. 2019), saprophytic (Zi 2015), and endophytic characteristics (Tanapichatsakul et al. 2019; Liao et al. 2020). It is an important plant pathogen and an asexual fungus with specific economic value (Taylor et al. 2001). The types of compounds isolated from *Pestalotiopsis* in recent years include alkaloids, polyols, cyclic peptides, terpenes, isocoumarins, coumarins, quinones, semiquinones, chromones, simple phenols, phenolic acids, esters, and other novel compounds (Yang et al. 2012; Xie et al. 2015). Many of these compounds have important application prospects. Among the reported *Pestalotiopsis* species, most are pathogens, saprophytes, or endophytes, but there has been little research on the mycoparasitism of these species.

Mycoparasitism is the most critical form of antagonism involving direct physical contact with the host mycelium (Pal et al. 2006). It involves typical growth of biocontrol fungal mycelia toward the target pathogen followed by extensive coiling and secretion of various hydrolytic enzymes, leading to the dissolution of the pathogen's cell wall or membrane. This mycoparasitism is common among *Trichoderma*. However, the mycoparasitism of *Pestalotiopsis* species is utterly different from that of *Trichoderma* based on microscopic observation. The aeciospores' outer walls appear deformed and completely broken after treatment with *Trichoderma* (Li et al. 2014). *Pestalotiopsis* concentrates the contents of rust spores by producing toxins, and the cell walls sag inward. The contents of the affected rust spores are concentrated, and most of the spores are empty shells (Li et al. 2017).

Mycoparasitic *Pestalotiopsis* species produce secondary metabolites different from those of endophytic or pathogenic *Pestalotiopsis* species (Xie et al. 2015), and these species have yet to be developed and used as important fungal resources. Both the lifestyle and secondary metabolite richness of mycoparasitic fungi are not comprehensively understood. In this study, the genome of the mycoparasite *Pestalotiopsis* sp. PG52 isolated from *Aecidium wenshanense* was sequenced and annotated. A large set of genes involved in secondary metabolism was identified. The purpose of this research was to investigate the possible mechanisms of mycoparasitism, potential active secondary substances (antifungal or antibacterial substances), and gene resources for resistance breeding against fungal diseases using genomic sequencing.

* Corresponding authors: Ch. He, College of Life Science, Southwest Forestry University, Kunming, China; **e-mail:hcz70@163.com**
  J. Li, College of Life Science, Southwest Forestry University, Kunming, Yunnan, China; **e-mail: lijingcas@163.com**

Experimental

Materials and Methods

Microbial material. The aeciospores of *A. wenshanense* were collected in Kunming, Yunnan Province, People's Republic of China, in September 2012. The species was mistakenly identified as *Aecidium pourthiaea* Syd. (Cai and Wu 2008) and has been corrected to *A. wenshanense* (Zhuang and Wei 2016; Zhu et al. 2020). The aeciospores were incubated on distilled filter paper at 25°C and cultured until mycelium or colony formation was observed. After being cultured for approximately one week, strain PG52 was isolated from the aeciospores, identified as *Pestalotiopsis kenyana* (Sui et al. 2020), and preserved at Southwest Forest University, Kunming, China.

Mycelial sample preparation. The conidia of *Pestalotiopsis* sp. PG52 were cultured on modified Fries culture agar. After incubation at room temperature for three days, the mycelium was carefully scraped off and stored in liquid nitrogen for later use.

DNA extraction and WGS library construction. *Pestalotiopsis* sp. PG52 DNA was extracted using a TIANGEN (Tiangen, Beijing, China) Bacterial Genomic DNA Extraction Kit and sheared into fragments between 100 and 800 bp in size by a Covaris E220 ultrasonicator (Covaris, Brighton, UK). High-quality DNA was selected using AMPure XP beads (Agencourt, Beverly, MA, USA). After repair using T4 DNA polymerase (Enzymatics, Beverly, MA, USA), the selected fragments were ligated at both ends to T-tailed adapters and amplified using KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, NC, USA). Then, amplification products were subjected to a single-strand circularization process using T4 DNA ligase (Enzymatics) to generate a single-stranded circular DNA library.

Genome sequencing and assembly. The NGS library was loaded and sequenced on the BGISEQ-500 platform. Raw data are available in the GenBank. The raw reads with a high proportion of Ns (ambiguous bases) and low-quality bases were filtered out using SOAPnuke (v1.5.6) (Chen et al. 2018) with the parameters "-l 15 -q 0.2 -n 0.05 -Q 2 -c 0". Then, the clean NGS ("Next-generation" sequencing technology) data were assembled using Canu (Koren et al. 2017) with the parameters "-useGrid=false maxThreads=30 maxMemory=60 g -nanopore-raw *.fastq -p -d". BUSCO (v3.0.1) was used to assess the confidence of the assembly with *Pestalotiopsis* sp. PG52.

Identification of Repetitive Elements and Non-Coding RNA Genes. Repetitive sequences were identified using multiple tools. TEs were identified by aligning against the Repbase (Bao et al. 2015) database using RepeatMasker (v4.0.5) (Tarailo-Graovac and Chen 2009) with parameters "-nolow -no_is -norna -engine wublast" and RepeatProteinMasker (v4.0.5) with parameters "-noLowSimple -pvalue 0.0001" at DNA and protein levels respectively. Meanwhile, the de novo repeat library was detected using RepeatModeler (v1.0.8) and LTR-FINDER (v1.0.6) (Xu and Wang 2007) with default parameters. Based on the de novo identified repeats, repeat elements were classified using RepeatMasker (v4.0.5) (Tarailo-Graovac and Chen 2009) with the same parameters. Furthermore, the tandem repeats were identified using Tandem Repeat Finder (v4.07) (Benson 1999) with parameters "-Match 2 -Mismatch 7 -Delta 7 -PM 80 -PI 10 -Minscore 50 -MaxPeriod 2000".

For non-coding RNA (ncRNA), the tRNA genes were predicted using tRNAscan-SE (v1.3.1) (Lowe and Eddy 1997) with default parameters. The rRNA fragments were identified using RNAmmer (v1.2). The snRNA and miRNA genes were predicted using CMsearch (v1.1.1) (Cui et al. 2016) with default parameters after aligning against the Rfam database (Kalvari et al. 2018) with a blast (v2.2.30).

Gene prediction and genome annotation. The predicted genes were aligned to the KEGG (Kanehisa 1997; Kanehisa et al. 2004; Kanehisa et al. 2006), Swiss-Prot (Magrane and UniProt Consortium 2011), COG (Tatusov et al. 1997; 2003), CAZy (Cantarel et al. 2009), NR and GO (Ashburner et al. 2000) databases using blastall (v2.2.26) (Altschul et al. 1990) with the parameters "-p blastp -e 1e-5 -F F -a 4 -m 8". The *Pestalotiopsis* sp. PG52 assembly was uploaded to the antiSMASH (v5.0) (Medema et al. 2011) website to identify the secondary metabolite gene cluster.

Transcriptome analysis. In order to define secondary metabolite clusters using transcriptional data, *Pestalotiopsis* sp. PG52 was inoculated on modified Fries medium for experiment. Abundant secondary metabolites were detected in the study. Total RNA was extracted from tissue samples. The mRNA was purified and then reverse transcribed into cDNA, and the library was constructed according to the large-scale parallel signature scheme. They were then sequenced using Illumina's technology. The genomic annotation results were compared with transcriptome data, and if mRNA of a gene was detected, the gene was considered to be expressed.

Results

Pestalotiopsis sp. PG52 genome extraction and quality inspection. The quality and concentration of the extracted *Pestalotiopsis* sp. PG52 genomic DNA were measured using a Qubit fluorometer, and then the DNA was subjected to 1% agarose gel electrophoresis. The sample volume was 1 μl. The test results are shown in Fig. 1 and indicate that the extracted genomic DNA had
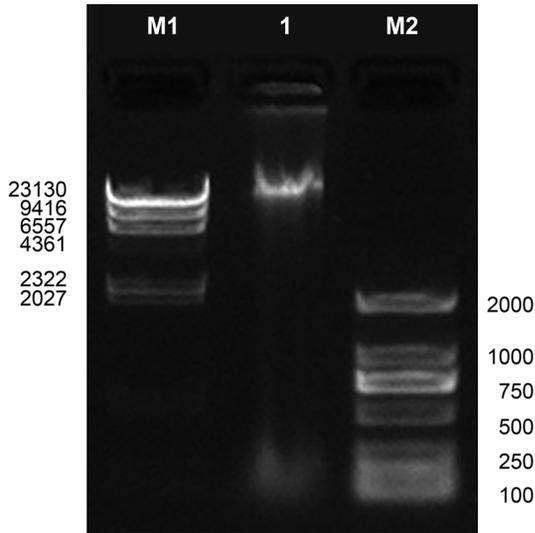
Fig. 1. Electrophoresis pattern of *Pestalotiopsis kenyana* PG52 genome. Agarose concentration (%): 1; voltage: 180 V; time: 35 min.; molecular weight standard name: M1: λ-Hind III digest (Takara), M2: D2000 (Tiangen); sample volume: M1: 3 μl, M2: 6 μl.

good integrity. BD Image Lab software was used to calculate the amounts of DNA in the electrophoresis image. The total amount of DNA in the samples was 3.78 μg, which meets the requirements for library construction and sequencing; this amount could meet the requirements for two or more samples for library construction.

Genomic sequencing quality analysis. Fqcheck software was used to evaluate the quality of the data. Fig. 2 and 3 show the base composition and quality of PG52.

The slight fluctuation at the beginning of the curve is typical of the BGI-seq 500 sequencing platform and does not affect the data. Normally, the distribution curves of the A and T and the C and G bases should coincide with each other. If an abnormality occurs in the sequencing process, it may cause abnormal fluctuations in the middle of the curve. If a particular library construction method or library is used, the base distribution may also be changed (Fig. 2).

The base quality distribution reflects the accuracy of the sequencing reads. The sequencer, sequencing reagents, and sample quality can all affect base quality. Overall, the low-quality (< 20) base proportion was low, indicating that the sequencing quality of the lane was relatively good (Fig. 3).

Genome assembly and gene prediction. The long fragment of *Pestalotiopsis* sp. PG52 was sequenced on the Nanopore platform, and a total of 12.18 Gb of data was generated. Before assembly, k-mer was selected as 15, and k-mer analysis was performed based on the second-generation data to estimate the genome's size (assembly results indicate the true genome size), degree of heterozygosity, and repeatability. Using Jellyfish software to process the filtered data, the results showed that the genome size of the PG52 strain was 50.7 Mb. We used Canu to assemble the Nanopore data and then with Pilon used the second-generation data for base error correction to obtain the final assembly result. BUSCO integrity assessment was conducted using the genome database (SordariomycetA_ODB9). More than 97.0% of core genes could be annotated in the genome, reflecting the high integrity of assembly results. A total of 335 scaffolds were assembled by genome stitching. The genome size was 58.01 Mbp, and the values of N50 and N90 were 6,598,051 bp and 55,791 bp, respectively. The entire genome's size was larger than those of the *Pestalotiopsis fici* (51.91 Mbp), *Pestalotiopsis* sp. JCM 9685 (48.23 Mbp) and *Pestalotiopsis* sp. NC0098 (46.41 Mbp) genomes, which have been sequenced.
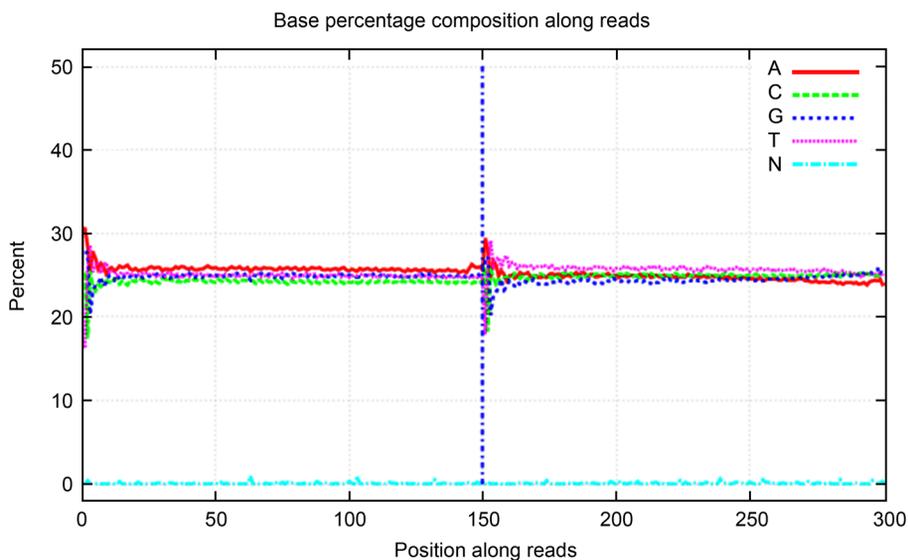

Fig. 2. *Pestalotiopsis kenyana* PG52 base composition distribution map. The X axis represents the position on reads, and the Y axis represents the percentage of bases.
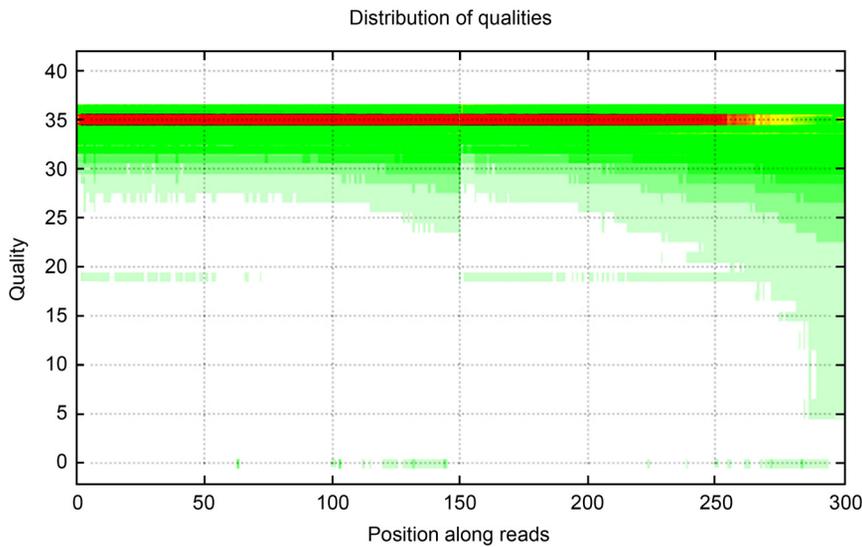
Fig. 3. *Pestalotiopsis kenyana* PG52 base mass distribution map. The X axis is the position of the base in reads, and the Y axis is the base quality value. Each point in the figure represents the total number of bases at this position that reach a certain.

A total of 20,023 genes were predicted in the *Pestalotiopsis* sp. PG52 genome, with an average length of 1,714.03 bp, an average CDS length of 1,478.29 bp, an average of 3.13 exons per gene, an average exon length of 472.00 bp, and an average intron length of 110.57 bp. The reported average length of the predicted genes of *P. fici* (Wang et al. 2015) is 1,683.88 bp, and the average number of exons contained in each gene is 3. Another reported average length of the predicted genes of *Pestalotiopsis* sp. NC0098 is 1864 bp, and the average number of exons contained in each gene is 2.83. The above comparison results indicate the reliability of the sequencing data for the *Pestalotiopsis* sp. PG52 genome and the similarity to the other two *Pestalotiopsis* strain genomes (Table I).

Gene prediction and functional annotation. The NCBI NR database was used to annotate the predicted genes, with a total of 17,500 genes annotated (accounting for 87.40% of the total predicted genes), and the KEGG database was used to annotate the predicted genes, with a total of 11,847 genes annotated (account-

ing for 59.17% of the total predicted genes). Using the GO database to annotate the predicted genes, a total of 10,454 genes were annotated (accounting for 52.21% of the total predicted genes).

KEGG (Kyoto Encyclopedia of Genes and Genomes). KEGG enrichment analysis showed that 11,847 genes that corresponded to KEGG pathways were enriched in 129 metabolic pathways, and most of these genes were involved in metabolic pathways (ko01100) (4,306 genes), biosynthesis of secondary metabolites (ko01110) (1,677 genes), biosynthesis of antibiotics (ko01130) (1,267 genes) and biosynthesis of amino acids (ko01230) (495 genes) (Fig. 4).

GO (Gene Ontology). A total of 10,454 genes can be used to extract GO annotation information with Blast2GO. Based on function, the genes can be divided into three subcategories, namely, biological process (25 branches), cellular component (14 branches) and molecular function (13 branches), with a total of 52 branches (Fig. 5: 1 – metabolic process, 2 – cellular process, 3 – localization, 4 – biological regulation, 5 – cellular component organization or biogenesis, 6 – regulation of biological process, 7 – response to stimulus, 8 – signaling, 9 – negative regulation of biological process, 10 – positive regulation of biological process, 11 – reproduction, 12 – reproductive process, 13 – developmental process, 14 – multi-organism process, 15 – growth, 16 – biological adhesion, 17 – detoxification, 18 – nitrogen utilization, 19 – cell aggregation, 20 – carbon utilization, 21 – biological phase, 22 – cell proliferation, 23 – immune system process, 24 – pigmentation, 25 – rhythmic process, 26 – membrane, 27 – membrane part, 28 – cell, 29 – cell part, 30 – organelle, 31 – macromolecular complex,

Table I
The comparison of *Pestalotiopsis* genome sequences.

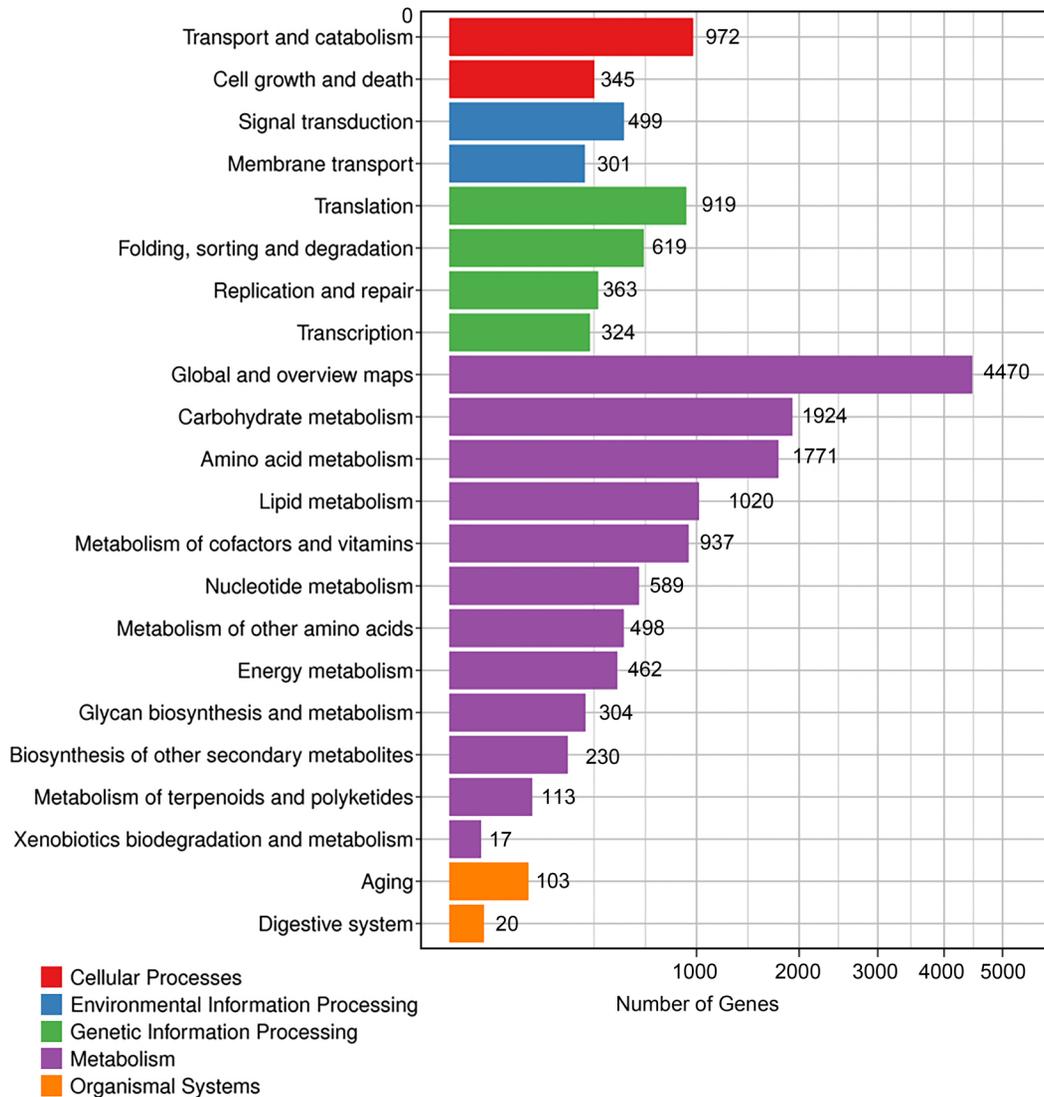|	PG52	FICI	NC0098
Assembly size (Mb)	55	52	46.61
Scaffold N50 (Mb)	6.6	4.0	5
Coverage (fold)	335.0	24.5	24
GC content (%)	53.30	48.73	51.28
Protein-coding genes	20,023	15,413	15,180
Gene density (genes per Mb)	345.22	296.90	327.08
Exons per gene	3.13	2.76	2.83

Fig. 4. KEGG function analysis.

32 – organelle part, 33 – membrane-enclosed lumen, 34 – extracellular region, 35 – supramolecular complex, 36 – virion, 37 – virion part, 38 – nucleoid, 39 – extracellular region part, 40 – catalytic activity, 41 – binding, 42 – transporter activity, 43 – transcription regulator activity, 44 – structural molecule activity, 45 – molecular function regulator, 46 – signal transducer activity, 47 – antioxidant activity, 48 – molecular transducer activity, 49 – molecular carrier activity, 50 – nutrient reservoir activity, 51 – protein tag, 52 – translation regulator activity). Most of the genes in the biological process category are involved in the metabolic processes and cellular processes, most of the genes in the cellular component category are involved in membrane and membrane part, and highest number of the genes in the molecular function category are involved in catalytic activity and binding.

COG (Cluster of Orthologous Groups of proteins). In the COG classification prediction of *Pestalotiopsis* sp. PG52 obtained by sequencing, a total of 8,975 genes were divided into 24 categories. In addition to the general function prediction category, the five categories with the greatest number of genes were amino acid transport and metabolism (905 genes, accounting for 10.08%), energy production and conversion (737 genes, accounting for 8.21%), carbohydrate transport and metabolism (709 genes, accounting for 7.90%), lipid transport and metabolism (657 genes, accounting for 7.32%), and secondary metabolite biosynthesis, transport and catabolism (516 genes, 5.75%) (Fig. 6: A – Cell cycle control, cell division, chromosome partitioning, B – Cell motility, C – Cell wall/membrane/envelope biogenesis, D – Defense mechanisms, E – Extracellular structures, F – Intracellular trafficking, secretion, and vesicular transport, G – Posttranslational modification, protein turnover, chaperones, H – Signal transduction mechanisms, I – Chromatin structure and dynamics, J – Replication, recombination and repair, K – RNA processing and modification, L – Transcription, M – Translation, ribosomal structure
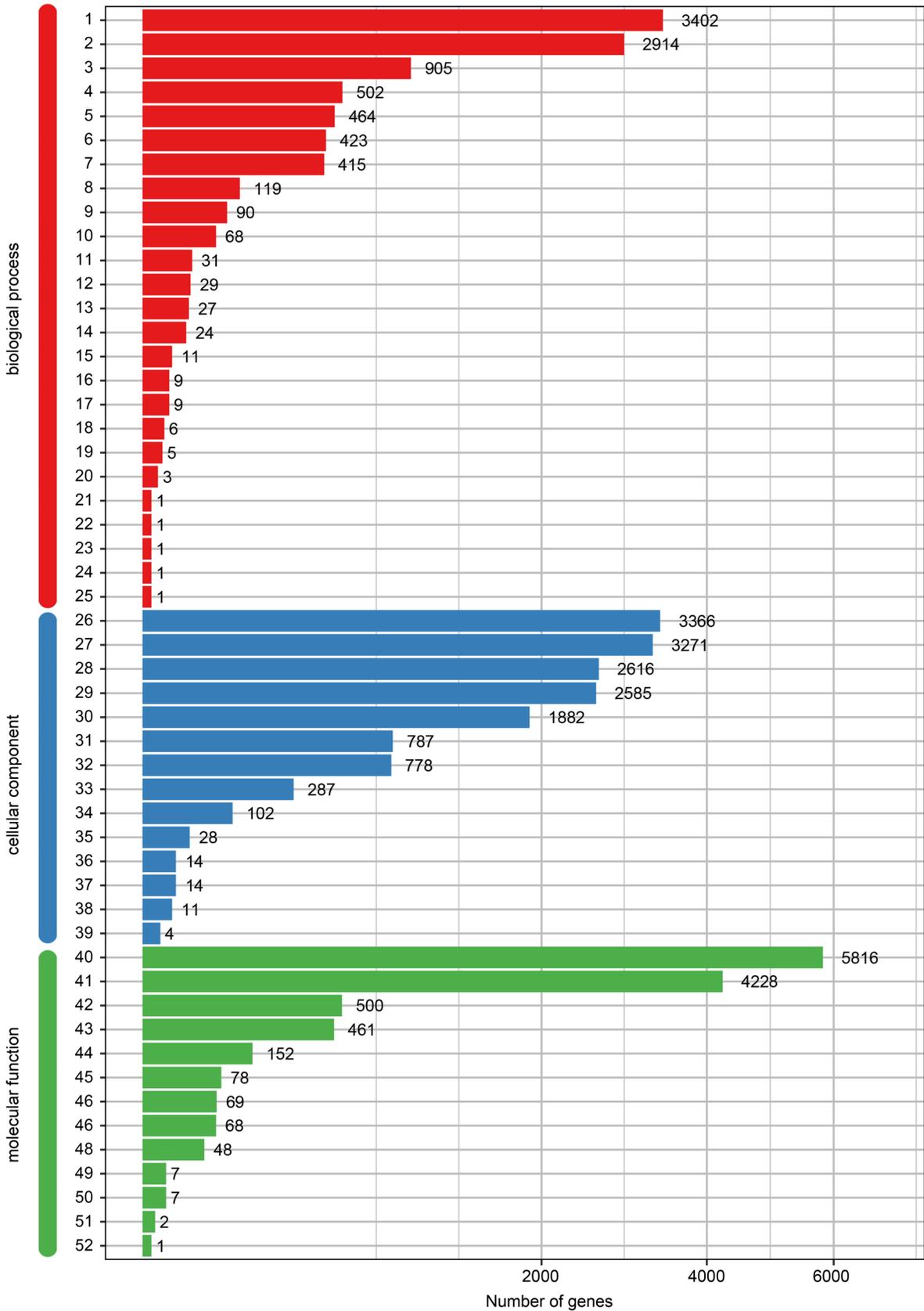
Fig. 5. GO functional classification map of all unigenes with GO annotation.

and biogenesis, N – Amino acid transport and metabolism, O – Carbohydrate transport and metabolism, P – Coenzyme transport and metabolism, Q – Energy production and conversion, R – Inorganic ion transport and metabolism, S – Lipid transport and metabolism, T – Mobilome – rophages, transposons, U – Nucleotide
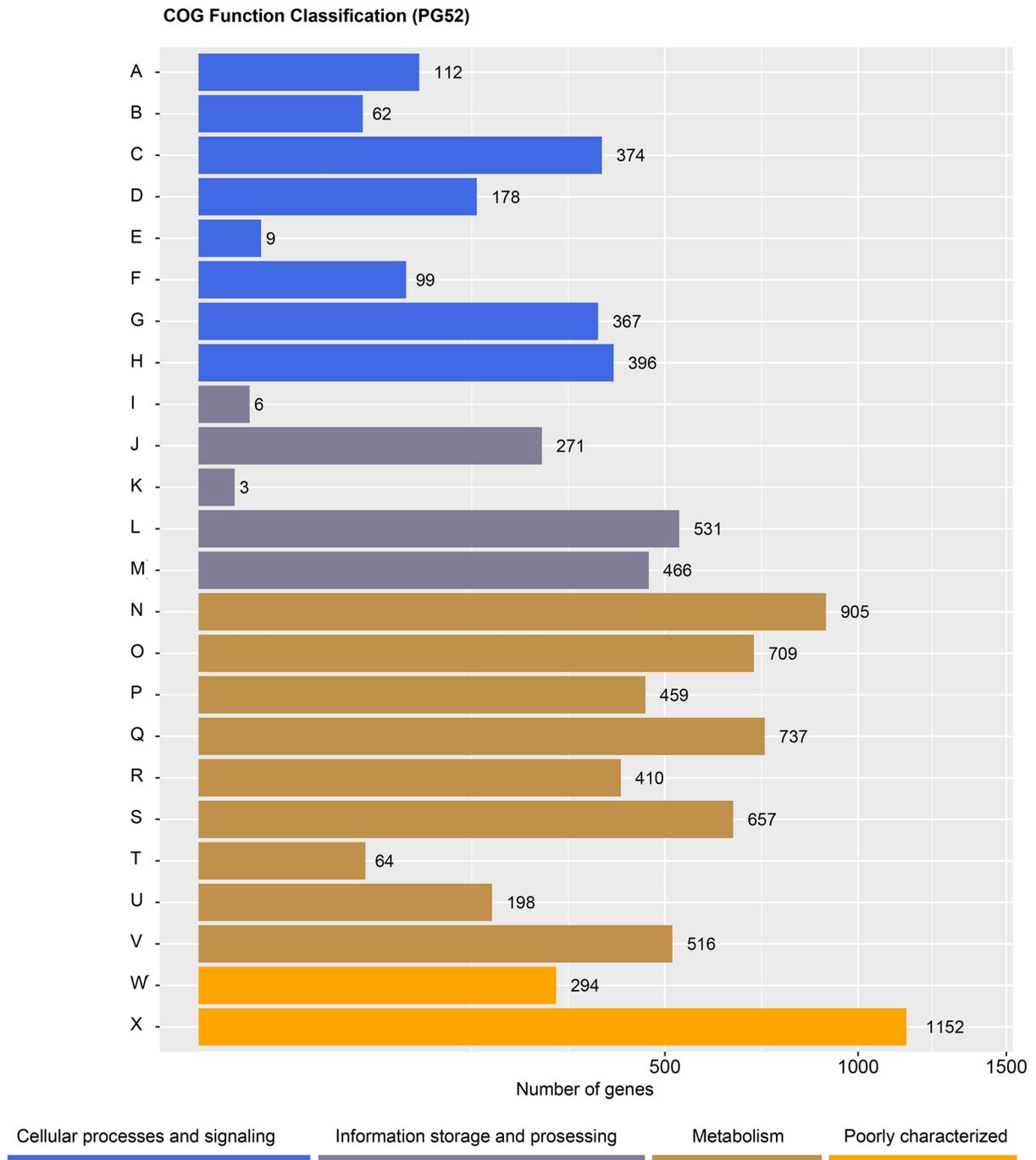
COG Function Classification (PG52)

Fig. 6. COG function analysis.

transport and metabolism, V – Secondary metabolites biosynthesis, transport and catabolism, W – Function unknown, X – General function prediction only).

CAZy (Carbohydrate-Active enZYmes) database. Carbohydrate-active enzymes participate in many important biological processes, including cell wall synthesis and signal and energy production, which are related to the fungal nutritional mode and infection mechanism (Zhao et al. 2014). The *Pestalotiopsis* sp. PG52 genome contains 345 hydrolase family genes (GHs), 150 glycosyltransferase family-like genes (GTs), 17 polysaccharide lyase family genes (PLs), 61 carbohy-

drate esterase family genes (CEs) and 196 carbohydrate-binding domain family genes (CBMs).

Further analysis of the GHs of mycoparasites is shown in Table II. *Pestalotiopsis* sp. PG52 has three GH18 and seven GH19 families (mainly chitinase) genes, significantly fewer than the number in the other three mycoparasites. This species contains 31 β-1,3-glucanase genes (GH17, GH55, GH64, and GH81 families), of which the GH55 gene is significantly redundant with those of *Trichoderma harzianum* (Antal et al. 2002; Steindorff et al. 2014; Baroncelli et al. 2015), *Trichoderma atroviride* (Kubicek et al. 2011; Shi-Kunne et al.

Table II

Glycosyl hydrolase families related to mycoparasitic in mycoparasites.

Species	GH18	GH75	GH17	GH55	GH64	GH81
Pestalotiopsis sp. PG52	3	6	7	19	3	2
Trichoderma harzianum	20	4	4	5	3	2
Trichoderma atroviride	29	5	5	8	3	2
Trichoderma virens	36	5	4	10	3	1

Table III

The number of polyketide synthases and nonribosomal peptide synthetases
of *P. fici*, *Pestalotiopsis* sp. NC0098 and mycoparasites.

Secondary metabolites	*Pestalotiopsis* sp. PG52	*P. fici*	*Pestalotiopsis* sp. NC0098	*T. harzianum*	*T. atroviride*	*T. virens*
NRPS	13	12	12	17	16	28
PKS	102	27	21	27	18	18
Total	115	39	33	44	34	46

2015) and *Trichoderma virens* (Kubicek et al. 2011); GH75 family chitosanase has been reported to degrade the host cell wall, and this process was also greatly enhanced (Cuomo et al. 2007). There are six GH75 genes in *Pestalotiopsis* sp. PG52, which is more than the number in the other three mycoparasites. The above results showed that the number of carbohydrate enzymes in *Pestalotiopsis* sp. PG52 is comparable to that of other mycoparasites, but the number of hydrolytic enzymes associated with mycoparasitism is similar to that in *T. harzianum* and lower than in *T. atroviride* and *T. virens*.

Others. The numbers of polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) in *Pestalotiopsis* sp. PG52, *Pestalotiopsis* sp. NC0098, *P. fici* and *Trichoderma* were compared. The results showed that the total numbers of PKSs and NRPSs from *Pestalotiopsis* sp. NC0098, *P. fici*, *T. virens*, *T. harzianum* and *T. atroviride* were similar but significantly lower than those from *Pestalotiopsis* sp. PG52 (Table III). We compared the number of NRPSs and PKSs in PG52 with that in *Pestalotiopsis* sp. NC0098, *P. fici* and 3 other mycoparasites. The results showed that the total number of the above enzymes in PG52 was much larger than that in *Pestalotiopsis* sp. NC0098, *P. fici* and *Trichoderma*, indicating that there may be more secondary metabolites in mycoparasitic *Pestalotiopsis* species. In

our previous study, four novel PKs were isolated from PG52. Their configurations were identified, and their toxic activities against human tumor cells were tested (Xie et al. 2015). Cytochrome P450 is a kind of multifunctional oxidase that is closely related to secondary metabolism (Črešnar and Petrič 2011). There are 317 cytochrome P450-coding genes in the *Pestalotiopsis* sp. PG52 genome, which is higher than the number reported in *T. harzianum*, *T. atroviride* and *T. virens*. A total of 175 proteases were found in the genome of *Pestalotiopsis* sp. PG52, which is significantly higher than the numbers in the genomes of *T. harzianum*, *T. atroviride* and *T. virens* (Table IV). There are more cytochromes and proteases in PG52 than in the other three *Trichoderma* mycoparasites. In addition, transcription factors (TFs) play a vital role in the fungal regulatory network. A total of 202 transcription factors were found in the genome sequencing results, including 19 genes encoding C2H2-type transcription factors and Zn2/Cys6-type transcription factors. There are 4 Zn2/Cys6-type transcription factor genes, which is significantly less than the number of such genes in *T. atroviride* and *T. virens* (Table IV).

Transcriptome analysis. The whole genome results of the *Pestalotiopsis* sp. PG52 were compared with the transcription group data (Table V), and 82 of the

Table IV

Numbers of P450, protease and Zn2/Cys6 transcription factor genes
of mycoparasites.

	Pestalotiopsis sp. PG52	*T. harzianum*	*T. atroviride*	*T. virens*
Cytochrome P450	317	50	15	40
Zn2/Cys6 transcription factor	4	7	69	95
Protease	175	53	23	28

Table V
The genome is compared with the transcription group.

Genes	Transcription groups	Genome	Expression rate
PKS	82	102	80.39%
NRPS	10	13	76.92%
Protease	137	175	78.29%
Cytochrome P450	245	317	77.29%
Zn2Cys6 transcription factor	1	4	25.00%

102 PKs genes found in the genome were detected in the transcription group with an expression rate of 80.39%. Ten NRPS genes were detected in the transcription group with an expression rate of 76.92%. Protease, Cytochrome P450 and Zn2Cys6 transcription factor have expression rates of 78.29 percent, 77.29 and 25.00 percent, respectively.

Discussion

Long fragments of *Pestalotiopsis* sp. PG52 were sequenced and assembled to obtain the complete genome sequence using BGISEQ-500 and Oxford Nanopore NGS technology, and this sequence was compared with relevant genome-wide information for *P. fici* and *Pestalotiopsis* sp. NC0098. The results showed that the genomes of similar *Pestalotiopsis* species are similar. Genes related to mycoparasitism and secondary metabolism were analyzed and compared with *T. harzianum*, *T. atroviride,* and *T. virens*. The results showed differences in the characteristics of mycoparasites in terms of parasitic ability and secondary metabolism.

In *Pestalotiopsis* sp. PG52, the number of mycoparasitism-related hydrolases, including chitinase (GH18), is less than that in *Trichoderma*, but the total number of β-1,3-glucanases (GH17, GH55, GH64, GH81) is greater than that in *Trichoderma*. All the chitinase genes in *Trichoderma* belong to the GH18 family (Seidl-Seiboth et al. 2014); however, a new chitinase family, GH19, was found in *Pestalotiopsis*; this family is always found in bacteria and higher plants (Suginta et al. 2016). The expression of this gene was also detected in transcriptome data analysis. The number of chitinase genes in *Pestalotiopsis* is far less than that in *Trichoderma* in general, which is consistent with the mycoparasitism characteristics of *Pestalotiopsis* and *Trichoderma*. *Pestalotiopsis* may produce toxins to concentrate the pathogenic bacterial content and generate dents in the cell wall, while *Trichoderma* produces enzymes (mainly chitinase) to destroy the cell wall of the pathogenic bacteria and cause pathogen lysis (Gruber et al. 2011).

A large number of protease genes were detected in the gene annotation results of *Pestalotiopsis* sp. PG52.

There are many proteins containing polysaccharides in the outermost layer of the cell wall of host fungi, and the expression of a large number of proteases in PG52 may enhance its parasitic ability to the host fungi. It has been reported that aspartic acid proteases may be involved in mycoparasitism, and some subtilisin-like serine proteases are homology of *Metarhizium anisopliae* PR1c and are involved in corneous degradation (Hu and Leger 2004, Herrera-Estrella 2014). These findings may be important in the involvement of proteases in the initial stages of mycoparasitism.

Mycoparasites produce secondary metabolites, proteases, and gene transcription regulation factors that are all closely related to mycoparasitism. Polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) are large multimodular enzymes involved in polyketide and peptide biosynthesis toxins produced by fungi. PKS is a key enzyme that regulates the synthesis of polyketides, mainly catalyzing the synthesis of secondary metabolites and pigments; NRPS can catalyze the synthesis of antimicrobial peptides (Gallo et al. 2013). Cytochrome P450 can catalyze some endogenous substances' biosynthesis with important physiological functions, such as hormones, fatty acids, and terpenoids, and play an important role in the modification of secondary metabolites (Črešnar and Petrič 2011). The higher amount of cytochrome P450 indicates that there may be more types of secondary metabolites in PG52. Some proteins secreted by fungi can play an important role in the process of infecting plant pathogenic fungi, reduce the defense capacity of plant pathogenic fungi and destroy pathogenic fungal cells, but their role in the process of mycoparasitism is still unclear (Mueller et al. 2008; Doehlemann et al. 2009). There are a high number of secreted proteins in the PG52 genome, and these proteins may play an important role in the process of mycoparasitism. Transcription factors can regulate gene expression and participate in fungi's secondary metabolic process (Schoberle et al. 2014). A Zn2/Cys6-type transcription factor found in PG52 can upregulate the β-glucosidase gene expression (Nitta et al. 2012). The number of Zn2/Cys6-type transcription factors in different mycoparasites varies greatly, and further research on this aspect is needed.

In this article, we report for the first time the complete genome information for the mycoparasite *Pestalotiopsis* sp. PG52, identifying a large number of genes related to mycoparasitism. We also show a preliminary comparison and analysis of four mycoparasite genomes, laying the foundation for studying the systematic evolution and revealing the mechanism of mycoparasitism of *Pestalotiopsis*. Additionally, this study provides reference information for genomic research on other filamentous fungi.

ORCID

Dengyun Zhang https://orcid.org/0000-0002-8077-4721

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

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct;215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Antal Z, Manczinger L, Kredics L, Kevei F, Nagy E. Complete DNA sequence and analysis of a mitochondrial plasmid in the mycoparasitic *Trichoderma harzianum* strain T95. Plasmid. 2002 Mar; 47(2):148–152. https://doi.org/10.1006/plas.2001.1559

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

Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015 Dec; 6(1):11. https://doi.org/10.1186/s13100-015-0041-9

Baroncelli R, Piaggeschi G, Fiorini L, Bertolini E, Zapparata A, Pè ME, Sarrocco S, Vannacci G. Draft Whole-Genome Sequence of the Biocontrol Agent *Trichoderma harzianum* T6776. Genome Announc. 2015 Jun 25;3(3):e00647–15. https://doi.org/10.1128/genomeA.00647-15

Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999 Jan 01;27(2):573–580. https://doi.org/10.1093/nar/27.2.573

Cai C, Wu JR. Preliminary study on *Aecidium pourthiaea* Syd. of heather rust. North Horticult. 2008;1:208–210.

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 01;37 Database: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 01;7(1):1–6. https://doi.org/10.1093/gigascience/gix120

Črešnar B, Petrič Š. Cytochrome P450 enzymes in the fungal kingdom. Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics. 2011 Jan;1814(1):29–35. https://doi.org/10.1016/j.bbapap.2010.06.020

Cui X, Lu Z, Wang S, Jing-Yan Wang J, Gao X. CMsearch: simultaneous exploration of protein sequence space and structure space improves not only protein homology detection but also protein structure prediction. Bioinformatics. 2016 Jun 15;32(12):i332–i340. https://doi.org/10.1093/bioinformatics/btw271

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 07;317(5843):1400–1402. https://doi.org/10.1126/science.1143708

Doehlemann G, van der Linde K, Aßmann D, Schwammbach D, Hof A, Mohanty A, Jackson D, Kahmann R. Pep1, a secreted effector protein of *Ustilago maydis*, is required for successful invasion of plant cells. PLoS Pathog. 2009 Feb 6;5(2):e1000290–e1000290. https://doi.org/10.1371/journal.ppat.1000290

Gallo A, Ferrara M, Perrone G. Phylogenetic study of polyketide synthases and nonribosomal peptide synthetases involved in the biosynthesis of mycotoxins. Toxins (Basel). 2013 Apr 19;5(4):717–742. https://doi.org/10.3390/toxins5040717

Gruber S, Kubicek CP, Seidl-Seiboth V. Differential regulation of orthologous chitinase genes in mycoparasitic *Trichoderma* species. Appl Environ Microbiol. 2011 Oct;77(20):7217–7226. https://doi.org/10.1128/AEM.06027-11

Herrera-Estrella A. Chapter 33 – Genome-wide approaches toward understanding mycotrophic *Trichoderma* species. In: Gupta VK, Schmoll M, Herrera-Estrella A, Upadhyay RS, Druzhinina I, Tuohy MG, editors. Biotechnology and Biology of *Trichoderma*. Amsterdam (The Netherlands): Elsevier; 2014. p. 455–464. https://doi.org/10.1016/B978-0-444-59576-8.00033-3

Hu G, Leger RJS. A phylogenomic approach to reconstructing the diversification of serine proteases in fungi. J Evol Biol. 2004 Nov;17(6):1204–1214. https://doi.org/10.1111/j.1420-9101.2004.00786.x

Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, Petrov AI. Non-coding rna analysis using the rfam database. Curr Protoc Bioinformatics. 2018 Jun;62(1):e51. https://doi.org/10.1002/cpbi.51

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 2006 Jan 01;34(90001):D354–D357. https://doi.org/10.1093/nar/gkj102

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004 Jan 01;32(90001):277D–280. https://doi.org/10.1093/nar/gkh063

Kanehisa M. A database for post-genome analysis. Trends Genet. 1997 Sep;13(9):375–376. https://doi.org/10.1016/S0168-9525(97)01223-7

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 Apr;12(4):R40. https://doi.org/10.1186/gb-2011-12-4-r40

Li J, Xie J, Li XN, Zhou ZF, Liu FL, Chen YH. [Isolation, identification and antimicrobial activity of mycoparasites (*Pestalotiopsis*) from *Aecid-*

ium pourthiaea] (in Chinese). Biotechnol Bull. 2017;33(3):122–127. https://doi.org/10.13560/j.cnki.biotech.bull.1985.2017.03.018

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 Jun;162(6):396–401. https://doi.org/10.1111/jph.12202

Liao G, Wu P, Liu Z, Xue J, Li H, Wei X. 2*H*-pyranone and isocoumarin derivatives from the endophytic fungus *Pestalotiopsis microspora* SC3082 derived from *Scaevola taccada* (Gaertn.) Roxb. Nat Prod Res. 2020 Jan 30:1–8. https://doi.org/10.1080/14786419.2020.1719488

Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997 Mar 01;25(5):955–964. https://doi.org/10.1093/nar/25.5.955

Magrane M, UniProt Consortium. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011 Mar 29; 2011(0): bar009. https://doi.org/10.1093/database/bar009

Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011 Jul 01;39(Web Server issue) suppl_2:W339–W346. https://doi.org/10.1093/nar/gkr466

Mueller O, Kahmann R, Aguilar G, Trejo-Aguilar B, Wu A, de Vries RP. The secretome of the maize pathogen *Ustilago maydis*. Fungal Genet Biol. 2008 Aug;45 Suppl 1:S63–S70. https://doi.org/10.1016/j.fgb.2008.03.012

Nitta M, Furukawa T, Shida Y, Mori K, Kuhara S, Morikawa Y, Ogasawara W. A new Zn(II)2Cys6-type transcription factor BglR regulates *β-glucosidase* expression in *Trichoderma reesei*. Fungal Genet Biol. 2012 May;49(5):388–397. https://doi.org/10.1016/j.fgb.2012.02.009

Pal KK, Scholar V, Gardener BBM. Biological Control of Plant Pathogens. Plant Health Instructor. 2006. https://doi.org/10.1094/phi-a-2006-1117-02

Schoberle TJ, Nguyen-Coleman CK, Herold J, Yang A, Weirauch M, Hughes TR, McMurray JS, May GS. A novel C_2H_2 transcription factor that regulates *gliA* expression interdependently with GliZ in *Aspergillus fumigatus*. PLoS Genet. 2014 May 1;10(5): e1004336. https://doi.org/10.1371/journal.pgen.1004336

Seidl-Seiboth V, Ihrmark K, Druzhinina I, Karlsson M. Chapter 5 – Molecular Evolution of *Trichoderma* Chitinases. In: Gupta VK, Schmoll M, Herrera-Estrella A, Upadhyay RS, Druzhinina I, Tuohy MG, editors. Biotechnology and Biology of *Trichoderma*. Amsterdam (The Netherlands): Elsevier; 2014. p. 67–78. https://doi.org/10.1016/B978-0-444-59576-8.00005-9

Shi-Kunne X, Seidl MF, Faino L, Thomma BPHJ. Draft genome sequence of a strain of cosmopolitan fungus *Trichoderma atroviride*. Genome Announc. 2015 Jun 25;3(3):e00287–15. https://doi.org/10.1128/genomeA.00287-15

Steindorff AS, Ramada MHS, Coelho ASG, Miller RNG, Pappas GJ Jr, Ulhoa CJ, Noronha EF. Identification of mycoparasitism-related genes against the phytopathogen *Sclerotinia sclerotiorum* through transcriptome and expression profile analysis in *Trichoderma harzianum*. BMC Genomics. 2014 Dec;15(1):204. https://doi.org/10.1186/1471-2164-15-204

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

Sui GQ, Zhang DY, Kong L, Chen YH, Li J. [Identification of the mycoparasites from *Aecidium Pourthiaea* and screening] (in Chinese). Biotechnol Bull. 2020;36:25–32. https://doi.org/10.13560/j.cnki.biotech.bull.1985.2019-1154

Tanapichatsakul C, Khruengsai S, Monggoot S, Pripdeevech P. Production of eugenol from fungal endophytes *NeoPestalotiopsis* sp. and *Diaporthe* sp. isolated from *Cinnamomum loureiroi* leaves. Peer J. 2019 Feb 12;7:e6427. https://doi.org/10.7717/peerj.6427

Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009 Mar;Chapter 4:Unit 4.10. 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;4(1):41. https://doi.org/10.1186/1471-2105-4-41

Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science. 1997 Oct 24;278(5338):631–637. https://doi.org/10.1126/science.278.5338.631

Taylor JE, Crous PW, Palm ME. Foliar and stem fungal pathogens of *Proteaceae* in Hawaii. Mycotaxon. 2001;78:449–490.

Wang S, Mi X, Wu Z, Zhang L, Wei C. Characterization and pathogenicity of *Pestalotiopsis* -like species associated with gray blight disease on *Camellia sinensis* in Anhui Province, China. Plant Dis. 2019 Nov;103(11):2786–2797. https://doi.org/10.1094/PDIS-02-19-0412-RE

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; 16(1):28. https://doi.org/10.1186/s12864-014-1190-9

Xie J, Li J, Yang YH, Li XN, Chen YH, Zhao PJ. Four pestalpolyols from a mycoparasite *Pestalotipsis* sp. PG52. Arch Pharm Res. 2015 Oct 20. https://doi.org/10.1007/s12272-015-0675-7

Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007 May 08;35 Web Server:W265–W268. https://doi.org/10.1093/nar/gkm286

Yang XL, Zhang JZ, Luo DQ. The taxonomy, biology and chemistry of the fungal *Pestalotiopsis* genus. Nat Prod Rep. 2012;29(6):622–641. https://doi.org/10.1039/c2np00073c

Zhao Z, Liu H, Wang C, Xu JR. Correction: comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics. 2014;15(1):6. https://doi.org/10.1186/1471-2164-15-6

Zhu L, Tao K, Zuo YB, Zhao ZY, Chen YH. [The biological characteristics of aeciospore germination of *Aecidium wenshanense*] (in Chinese). Plant Protect. 2020;46:203–207.

Zhuang JY, Wei SX. Additional notes on anamorphic rust fungi of China I. Some aecial form species. J Mycosystema. 2016;35: 1468–1474.

Zi W. Enzyme activities on decomposing needle litter of *Pinus armandii* by five dominant saprophytic fungi. Microbiol China. 2015;42:654–664.